A macro-finance term structure model with stochastic volatility

Linlin Niu* Università Bocconi

[Job Market Paper]

November 2007

Abstract

This paper proposes a term structure model with macro VAR in a stochastic volatility setting. The specific feature of this model is that the risk premium of yields is directly driven by the time-varying variance-covariance of the VAR innovation, which is modeled by a Wishart Autoregressive process. Extending the essentially affine term structure model, this framework not only incorporates the stochastic variance-covariance in the VAR innovation, but also preserves the tractability and interpretability from a macro-finance perspective. Hence it provides a modeling tool to bridge the two strands of macroeconomic research: the DSGE-VAR with stochastic volatility and the macro-finance model of term structure. The baseline model implies that: 1) the stochastic variance-covariance of the VAR innovation has sizable effect on medium to long maturity yields; 2) volatility is a curvature factor of the yield curve, and the net effect of the time-varying variance-covariance matrix is also a curvature factor; 3) simulation study shows that it can well explain the bond yield "conundrum", where differences in volatility can result in different shapes of the yield curve with the underlying macro variables remaining at the same level.

Keyword: Term structure, Stochastic volatility, Wishart Autoregressive process, Bond yield "conundrum"

JEL Classification: G12, E43

*Linlin Niu: PhD student in economics at Università Bocconi, Via Sarfatti 25, 20136, Milan, Italy. E-mail: niu@unibocconi.it. I am grateful to the invaluable support from the members of my dissertation committee: Carlo A. Favero (Chair), Andrew Ang, Francesco Giavazzi, and Luca Sala. I have benefited from helpful discussion with Paolo Colla, Claudio Tebaldi, Anna Battauz, Paolo Bianchi, and Junye Li. I thank research fundings from the PhD School and the Department of Economics in University Bocconi.

1 Introduction

This paper proposes a convenient framework for studying the effects of the volatility and covolatility of macro variables on the yield curve. It also provides a useful tool for utilitzing the yield curve information for the inference on macroeconomic volatility.

The stochastic behavior of the variance-covariance of macro variables and of financial asset prices are of great importance in our understanding of their joint dynamics.

In the macroeconomic research agenda, substantial effort has been devoted to the investigation of the "Great Moderation" of volatility in recent economic history (Stock and Watson (2003) for an overview). Important studies using DSGE-VAR models with stochastic volatility are due to Primiceri (2005), Justiniano and Primiceri (2007). However, these studies only rely on limited macro data. The rich information contained in bond yield data is not utilized for the estimation and inference.

In the macro-finance field where the term structure and the macro economy are jointly studied, commonly used models assume constant volatility (Ang and Piazzesi (2003), Ang, Piazzesi and Wei (2006), Ang, Dong and Piazzesi (2007), Diebold, Rudebusch and Aruoba (2006), Rudebusch and Wu (2005), to name just a few). These models usually perceive the yield curve as driven by a VAR state dynamics of macro variables and yield factors. In particular, models featuring no-arbitrage restrictions provide a powerful tool in understanding the joint dynamics in a parsimonious and coherent manner. But these no-arbitrage macro-finance models are usually confined to the class of essentially affine term structure models (Dai and Singleton(2002), Duffee(2002)) with constant variance-covariance of the VAR innovations. Though these works have contributed to our understanding of the relationship between bond yield dynamics, monetary policy transmission and the macro economy, the assumption of constant volatility is more likely to be violated in yield data than in macro data, and the likely effects of changing macro volatility on yields cannot be explored. Some papers have examined the implication of regime switching (and possibly change in volatility) on the yield curve and the macro economy (Ang and Bekaert (2002)), however, within each regime, variance-covariance matrix of the underlying state residuals is assumed to be constant. The omitted stochastic volatility might be crucial in driving the bond yield dynamics in some specific periods when the market volatility strongly deviates from its mean level, even within one regime. In examining the recent bond yield "conundrum", for example, volatility is found to be an important factor correlated with the unusually low level of long-term interest rate. The commonly used essentially affine term structure models cannot capture that behavior (Rudebusch, Swanson and Wu (2007)).

The reason why few macro-finance models incorporate stochastic volatility-covolatility might be due to the complexity of modeling such features. As noted by many studies, in the affine Gaussian class of term structure models, there seems to be a trade-off between matching properties of the conditional mean and the conditional volatilities of yields (Singleton (2006) for a detailed discussion). The choice of constant volatility affine model by macroeconomists might be due to the concern of matching the first moment of macroeconomic dynamics. On the other hand, some restrictive assumptions underlying yield models with stochastic volatility are hard to be justified from macroeconomic theory. For example, some quadratic term structure models dealing with stochastic volatility assume that the short rate is determined by the variance-covariance matrix of the state variables (Ahn, Dittmar, and Gallant (2002)); while macro economists usually regard the short rate as a monetary policy instrument which targets on the level of inflation and output gap.

Extending the essentially affine term structure model, this paper proposes a simple framework to not only incorporate a stochastic variance-covariance in the VAR innovations, but also preserve the tractability and interpretability from a macro-finance perspective. Hence the model provides a modeling tool to bridge two strands of macroeconomic research: the DSGE-VAR model with stochastic volatility and the macro-finance model of term structure. Using this framework, macroeconomists will be able to study the role of stochastic volatility in macro VAR by using information from the financial market; on the other hand, the effect of stochastic volatility underlying the macro economy on the term structure can be examined explicitly.

The proposed equilibrium no-arbitrage model of the yield curve can be understood as a generalized framework extending Vasicek (1977) model with (matrix) affine form of stochastic variance-covariance dynamics. In this setting, the time-varying risk premia come from uncertainty in the variance-covariance of innovations to the state risk factors that drive the short rate. This uncertainty then maps into longer maturity yields through no-arbitrage restrictions. The dynamics of volatility-covolatility, though evolving independently from the state risk factors, also drives the yield curve at medium-to-long maturities. Hence the volatility matrix can be deemed as "auxiliary" factor in yields. The model is denoted as Auxiliary Stochastic Volatility-covolatility Affine Term Structure Model (ASV-ATSM). If the innovations to the variance-covariance process are Gaussian, the model reduces to the Dai-Singleton(2002) Affine Term Structure Models (ATSM), but with a set of structural restrictions imposed on the parameter space. In the extreme case where the distribution of variance-covariance of VAR innovations collapses into a constant, the model converges to the essentially affine $A_0(m)$ model with constant risk price.

In the baseline model, both the VAR dynamics and the variance-covariance of VAR innovations affect the yield curve, without correlations between the two. This flexibility helps to capture both the feature of linear projection of the yield curve level, and the behaviors of stochastic volatility. In an extended model, leverage effect (i.e. correlation between the VAR variables and their contemporaneous variance-covariance factors) can be integrated, which enriches the dynamics of yields with respect to the volatility factors.

The paper is structured as follows: Section 2 explains the basic building blocks of the baseline model . Section 3 derives the model and the no-arbitrage restrictions. Section 4 discusses the general state-space form, model classification and extension with leverage effects. Section 5 is devoted to simulation study in which I examine the basic features of the model. Section 6 discusses estimation strategy. Section 7 concludes.

2 Model building blocks

The model is cast in discrete time. The basic building blocks are similar to the discrete-time essentially affine term structure model in Ang and Piazzesi (2003) with two exceptions: (i) risk prices are assumed to be directly driven by the stochastic variance-covariance matrix of the state VAR innovations; (ii) this variance-covariance matrix follows a Wishart Autoregressive process.

2.1 Short rate

The short rate r_t is affine in a state vector X_t, which includes some macro factors and possibly latent factors from the yields

$$r_t = \delta_0 + \delta_1' X_t \tag{1}$$

δ_0: a scalar.

δ_1: a $K \times 1$ vector.

2.2 State variable dynamics

The transition equation for X_t follows a VAR(1):

$$X_t = \mu + \Phi X_{t-1} + v_t, \qquad v_t \sim N(0, \Omega_t) \tag{2}$$

X_t: a $K \times 1$ vector.

The variance-covariance matrix Ω_t of the VAR innovation v_t follows a Wishart Autoregressive (WAR) process,

$$\Omega_t = M\Omega_{t-1}M' + \Sigma^* + \eta_t, \tag{3}$$

where $\Sigma^* = J\Sigma$, J denotes the degree of freedom ($J \geq K$ to ensure nondegeneracy of the distribution of Ω_t), and η_t is a matrix of stochastic errors with zero conditional mean. In particular, $\Omega_t \equiv \sum_{j=1}^{J} z_{j,t}z_{j,t}'$, $z_{j,t} = Mz_{j,t-1} + \xi_{j,t}$, $\xi_{j,t} \sim N(0, \Sigma)$. M is the latent autoregressive coefficient and Σ the latent variance of the innovations. Gourieroux, Jasiak, and Sufana(2004) study the property of this process in details.

Interesting features of this process is that at any time, the conditional distribution of Ω_t is a well-defined non-central Wishart with

$$(\Sigma^* + \eta_t) \sim W(J, \Sigma)$$

and

$$Cov\left[vec\left(\Omega_t\right)\right] = Cov\left[vec\left(\eta_t\right)\right] = J\left(I_{k^2} + H\right)\left(\Sigma \otimes \Sigma\right)$$

, where K is the number of VAR state factors that determine the short rate, and H is the commutation matrix $H = \sum H_{ij} \otimes H_{ij}'$, where H_{ij} denotes the $K \times K$ matrix with $h_{ij} = 1$ and all other elements zero (Muirhead, R.J.(1982)). The distribution of $vec\left(\eta_t\right)$ is highly skewed when J is low, and it slowly approaches to a Normal as J increases.

2.3 Prices of risk

The prices of risk, denoted by a vector Λ_t, are determined by the square root of the variance-covariance matrix Ω_t, adjusted by a constant $K \times 1$ vector Λ

$$\Lambda_t = \Omega_t^{1/2} \Lambda.$$

The prices of risk are associated with the sources of uncertainty in v_{t+1}. In the essentially affine term structure model with constant Ω, in order to capture time-varying risk premia, Λ_t is assumed to be affine in the VAR state X_t, $\Lambda_t = \lambda_0 + \lambda_1 X_t$. In the current model, the time-varying risk premia can be captured by the stochastic variance-covariance Ω_t naturally, since market risk is fundamentally linked to volatility - the second moment, instead of the first moment in the level of state X_t.

2.4 Pricing kernel

No arbitrage opportunity between bonds with different maturities implies that there is a discount factor m linking the price of bond with maturity n at time t with the price of bond with maturity $n-1$ at time $t+1$.

$$P_t^{(n)} = E_t \left[m_{t+1} P_{t+1}^{(n-1)} \right] \tag{4}$$

The stochastic discount factor is related to the short rate and risk perceived by the market, which is defined as

$$\begin{aligned} m_{t+1} &= \exp\left(-r_t - \frac{1}{2} \Lambda_{t+1}' \Lambda_{t+1} - \Lambda_{t+1}' \varepsilon_{t+1} \right) && (5) \\ &= \exp\left(-r_t - \frac{1}{2} \Lambda' \Omega_{t+1} \Lambda - \Lambda' v_{t+1} \right) && (6) \end{aligned}$$

with $v_{t+1} = \Omega_{t+1}^{1/2} \varepsilon_{t+1}$, $\varepsilon_{t+1} \sim N(0,1)$.

Notice that in essentially affine models as in Ang and Piazzesi (2003),

$$m_{t+1} \equiv \exp\left(-r_t \; - \frac{1}{2} \Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1} \right),$$

where $\Lambda_t = \lambda_0 + \lambda_1 X_t$, hence the time variation there in the risk premium is due to dynamics in the first moment X_t. That discount factor can be represented with a transformation of Λ_t so that

$$m_{t+1} \equiv \exp\left(-r_t \; - \frac{1}{2} \tilde{\Lambda}_t' \Omega \tilde{\Lambda}_t - \tilde{\Lambda}_t' v_{t+1} \right),$$

where $\Lambda_t = \Omega^{1/2}\tilde{\Lambda}_t$. The similarity between these discount factor and those in equations (5) and (6) implies that they can be observationally equivalent, though the driving forces to time-varying risk primium are different.

A no-arbitrage recursive relation can then be derived from the above equations as:

$$\begin{aligned}
P_t^{(n)} &= E_t\left[m_{t+1}P_{t+1}^{(n-1)}\right] = E_t\left[m_{t+1}m_{t+2}P_{t+2}^{(n-2)}\right] = \cdots \\
&= E_t\left[m_{t+1}m_{t+2}...m_{t+n}P_{t+n}^{(0)}\right] = E_t\left[m_{t+1}m_{t+2}...m_{t+n}\cdot 1\right] \\
&= E_t\left[\exp\left(-\sum_{i=0}^{n-1}\left(r_{t+i}+\tfrac{1}{2}\Lambda'\Omega_{t+1+i}\Lambda+\Lambda' v_{t+1+i}\right)\right)\right] \\
&= E_t\left[\exp\left(A_n + B_n' X_t + C_n\left(\Omega_t\right)\right)\right] = E_t\left[\exp\left(-n y_{t,n}\right)\right] \\
&= E_t^Q\left[\exp\left(-\sum_{i=0}^{n-1} r_{t+i}\right)\right]
\end{aligned}$$

E_t^Q denotes the expectation operator under the risk-neutral probability measure.

The relationships between bond yields and the bond prices are:

$$y_{t,t+n} = \frac{-1}{n} p_{t,t+n} \tag{7}$$

$$p_{t,t+n} \equiv \ln P_t^{(n)} = A_n + B_n' X_t + C_n\left(\Omega_t\right)$$

$$y_{t,t+n} = a_n + b_n' X_t + c_n\left(\Omega_t\right) = \frac{-1}{n}\left(A_n + B_n' X_t + C_n\left(\Omega_t\right)\right) \tag{8}$$

Note that the short rate equation imposes $C_1(\Omega_t) = 0$ as a boundary condition.

3 Econometric model representation and no-arbitrage restrictions

The above assumptions on the model building blocks imply first, that yields with different maturities are driven by both the level of the state risk factors X_t and the variance-covariance of the VAR innovations Ω_t; second, the factor loadings are tightly related by the no-arbitrage condition.

This amounts to an econometric representation of a state-space model augmented with a stochastic process of the variance-covariance matrix. That is, there are three blocks of equations: the first block defines the measurement equations of yields with different maturities

n, where $c_n(\Omega_t)$ is an affine function of elements in Ω_t.; the second block is the VAR state dynamics of X_t with time-varying variance-covariance Ω_t of the VAR innovations ; and the third block gives the autoregressive dynamics of Ω_t.

$$\begin{array}{lll} y_{t,t+n} & = & a_n + b_n' X_t + c_n(\Omega_t) + \varepsilon_{t,t+n}, \quad \varepsilon_{t,t+n} \sim N(0,\sigma^2) \\ X_t & = & \mu + \Phi X_{t-1} + v_t, \quad v_t \sim N(0,\Omega_t) \\ \Omega_t & = & M\Omega_{t-1}M' + J\Sigma + \eta_t \quad J\Sigma + \eta_t \sim W(J,\Sigma) \end{array} \tag{M-0}$$

3.1 Unrestricted model

The econometric model can be estimated in an unrestricted manner, where no restrictions are imposed on the yield equations to assure no-arbitrage. In this case, the model is the stochastic counterpart of Engle's (G)ARCH-in-mean model of asset returns, in which the contemporaneous volatility affects returns. In particular, the WAR process shares a similar spirit of the BEKK-GARCH model (Engle and Kroner (1995)), where the variance-covariance has matrix autoregressive dynamics. The WAR process automatically ensures positive definitiveness of Ω_t with a well-defined dynamics and distribution.

In the ARCH-in-mean model, the levels of VAR states are often ignored and the role of variations in variance-covariance is emphasized. This seems to be a reasonable simplification when applied to high frequency data. At monthly or quarterly frequency, on the contrary, VAR state dynamics seems to be helpful to understanding the relatively slow movement and effects of macro state variables like inflation and output on the yield curve.

3.2 Restricted (No-arbitrage) model

The no-arbitrage restrictions on the bond price equations in such a framework can be derived as follows (Appendix A.1) :

With parameter space $\left(\delta_0, \delta_1, \sigma^2, \mu, \Phi, M, \Sigma, J, \Lambda\right)$:

$$\begin{array}{lll} A_{n+1} & = & A_1 + A_n + B_n'\mu + D_n \\ B_{n+1}' & = & (B_n'\Phi + B_1') \\ C_{n+1}(\Omega_t) & = & Tr\left[G_{n+1} \cdot \Omega_t\right] \end{array} \tag{9}$$

in which

$$\begin{aligned} A_1 &= -\delta_0 \\ B_1 &= -\delta_1 \\ G_1 &= 0 \end{aligned}$$

and

$$\begin{aligned} G_{n+1} &\equiv M'\Gamma_n \left(I - 2\Sigma\Gamma_n\right)^{-1} M \text{ for } n > 0 \\ D_n &\equiv -\tfrac{J}{2} \ln\left[\det\left(I - 2\Sigma\Gamma_n\right)\right] \end{aligned}$$

with Γ_n defined as follows:

$$\Gamma_n = -\tfrac{1}{2}\Lambda B_n' - \tfrac{1}{2}B_n\Lambda' + \tfrac{1}{2}B_n B_n' + G_n$$

Restrictions on the yield equations are accordingly:

$$\begin{aligned} b_{n+1} &= -\tfrac{1}{(n+1)} B_{n+1} \\ &= \tfrac{1}{(n+1)} \left[\sum_{i=0}^{n} \left(\Phi'\right)^i\right] b_1 \\ a_{n+1} &= -\tfrac{1}{(n+1)} A_{n+1} \\ &= = a_1 + \tfrac{1}{(n+1)} b_1' \left[\sum_{i=0}^{n-1} \Phi^i\right] \mu + \tfrac{J}{2(n+1)} \sum_{i=1}^{n} \ln\left[\det\left(I - 2\Sigma\Gamma_i\right)\right] \\ c_{n+1}\left(\Omega_t\right) &= -\tfrac{1}{(n+1)} C_{n+1}\left(\Omega_t\right) \\ &= -\tfrac{1}{(n+1)} vec\left(G\left(n+1\right)\right)' \cdot vec(\Omega_t) \end{aligned}$$

3.3 Stochastic volatility and the curvature factor

What are the effects of stochastic volatility on yields under this model setting? What are the shapes of its factor loadings?

Suppose there is one state factor in X_t and its innovation follows a one-dimension Wishart Autoregressive process – a Chi-square Autoregressive process. Approximate the short rate by the 1-month yeilds, and calibrate its dynamics with an AR(1) and innovations an ARCH(1) with monthly data from 1974:2 to 2001:12, I can set the VAR and WAR parameters as: $\mu = 2.32 \times 10^{-3}$, $\Phi = 0.95$, $M = 0.924$, $\Sigma^* \equiv J\Sigma = 1.1 \times 10^{-7}$. Then set $\delta_0 = 0$, $\delta_1 = 1$, use the discount factor definition of $m_{t+1} \equiv \exp\left(-r_t - \frac{1}{2}\Lambda'\Omega\Lambda - \Lambda' v_{t+1}\right)$ and calibrate a Vasicek model to obtain $\Lambda = -400$. The resulting factor loadings of the volatility Ω_t on yields has a hump shape, which is quite similar to the familiar curvature factor from the Nelson-Siegel representation.

Take the above calibrated parameter values, Figure 1 shows the implied intercepts a_n of yields, factor loadings b_n on the states X_t and loadings $-\frac{1}{(n+1)}G(n)$ of volatility matrix Ω_t, respectively. It clearly depicts the state X_t as a slope factor, and Ω_t as a curvature factor.

[**Figure 1. Factor loadings of yields with one state in X and one volatility factor**]

Figure 2 compares how the model parameters affect the volatility loadings on yields. The first row shows that the persistence parameters Φ of the state dynamics and M of the stochastic volatility process both have positive effects on the factor loadings. The higher is the persistence, the bigger is the volatility effect on medium-to-long yields, and the peak of the curvature factor also depends on the persistence. The first panel in the second row shows that of risk price λ governs the sign as well as magnitude of the factor loadings. When risk price is negative(positive), the factor loadings are positive(negative), hence higher volatility results in lower(higher) price of medium-to-long term bonds. The degree of freedom parameter, J, instead, has little effect on the factor loadings.

[**Figure 2. Parameters affecting volatility factor loading**]

The above figures presents some basic features of the volatility factors when there is only one state factor and hence one volatility factor. More general cases with multiple state factors and volatility-covolatility factors will be discussed in the simulation studies in section 5.

3.4 Forward rate and excess returns

This model implies that forward rate is a function of both the state X_t and the volatility-covolatility of state innovation Ω_t, but excess returns are only driven by Ω_t. (Appendix B).

Forward rate

Let $f_{t,n}^{(1)}$ denote the log forward rate at time t for loans between time $t+n-1$ and $t+n$. It has the following expression:

$$f_{t,n}^{(1)} = (A_{n-1} - A_n) + \left(B_{n-1}' - B_n'\right) X_t + Tr\left[(G_{n-1} - G_n)\,\Omega_t\right] \tag{10}$$

Excess returns

Define rx_{t+1}^n as the log holding period return from buying an n-period bond at time t and selling it as an $n-1$ period bond at time $t+1$, the excess return rx_{t+1}^n is driven by current volatility-covolatility of state innovations together with all innovations v_{t+1} and η_{t+1}, to X_{t+1} and Ω_{t+1}, respectively.

$$rx_{t+1}^n \quad = \quad const. + Tr\left[(M'G_{n-1}M - G_n)\,\Omega_t\right] + g(v_{t+1}, \eta_{t+1}) \ , \tag{11}$$

where $g(v_{t+1}, \eta_{t+1})$ is a linear combination of the innovations.

The expected excess return $E_t\left[rx_{t+1}^n\right]$ is only a function of current volatility-covolatility of state innovations:

$$E_t\left[rx_{t+1}^n\right] = const. + Tr\left[\left(M'G_{n-1}M - G_n\right)\Omega_t\right] \tag{12}$$

In general the expected excess return between n-period bonds at time t and $n-s$ period bonds at time $t+s$ can be expressed as:

$$E_t\left[rx_{t+s}^n\right] = const. + Tr\left\{\left[(M^s)'G_{n-s}M^s - G_n + G_s\right]\Omega_t\right\}. \tag{13}$$

4 Compact State-space form, model classification and extension with leverage effects

In the econometric model representation (M-0), it is easy to understand that the augmented process of the variance-covariance matrix Ω_t is also state dynamics in addition to the VAR process of X_t. Hence, the three equations can be written in a more compact state-space form. I classify this type of model as Auxiliary Stochastic Volatility-covolatility (ASV) affine term structure models (ATSM). The original representation (M-0) is helpful in understanding intuitively the restriction derivation and distribution property of the state elements. A more compact form is useful in understanding the relative classification of this model with respect to other affine term structure models.

In macro VAR with time-varying variance-covariance, the exogeneity of Ω_t to VAR states X_t is usually assumed. I maintain this assumption in the baseline model derived above. However, in financial data, leverage effects are often observed, i.e., there is significant correlation between the variance-covariance and the level of the returns. This phenomenon is also relevant in the dynamics of inflation, which is an important factor determining the yield movement. At the end of this section, I shall discuss the possibility of extension to allow leverage effect in the model.

4.1 Compact State Space form of the ASV-ATSM model

Define the entire state vector as

$$Z_t = \begin{bmatrix} X_t \\ vech\,(\Omega_t) \end{bmatrix},$$

and coefficient vector of state Z_t on the measurement equation of yield with maturity n as

$$\beta_n = \begin{bmatrix} b_n \\ -\frac{1}{n} vec\,(G_n) \cdot S \end{bmatrix},$$

where S is the operator for transformation $vec(X) = S \cdot vech(X)$.

Further, with the following reparameterization of the state dynamics,

$$U = \begin{bmatrix} \mu \\ J \cdot vech\,(\Sigma) \end{bmatrix}, \; F = \begin{bmatrix} \Phi & 0 \\ 0 & (M \otimes M) \cdot S \end{bmatrix}, \text{ and } Q_t = \begin{bmatrix} \Omega_t & 0 \\ 0 & S \cdot V_\eta \cdot S' \end{bmatrix},$$

the compact state space model representation can be written as

$$\begin{array}{lll} y_{t,t+n} & = & a_n + \beta_n' Z_t + \varepsilon_{t,t+n}, \quad \varepsilon_{t,t+n} \sim N(0, \sigma^2) \\ Z_t & = & U + F Z_{t-1} + v_t \,, \qquad v_t \sim (0, Q_t) \end{array} . \tag{M-I}$$

The necessary intermediate results of matrix transformation on vectorization is listed in Appendix C.

4.2 Classification of the ASV-ATSM model

The ASV-ATSM model can be classified according to the number of state variables X_t driving short rate, K, and the number of stochastic volatility-covolatility elements that govern the innovations v_t to the state variables X_t, m. I denote the by $A^{+m}(K)$., where $0 \leq m \leq \frac{K(K+1)}{2}$. In this class of models, X_t is conditionally Gaussian with variance-covariance Ω_t, but none of the K factors drive stochastic volatility; instead, it is those m additional stochastic volatility-covolatility factors at work, which do not enter the short rate equation. In addition, these m factors jointly follow Wishart Autoregressive process and have non-central Wishart distribution. Hence, I put $^{+m}$ into the notation to distinguish them from the Dai-Singleton classification of Affine Term Structure Model $A_m(n)$, where the m stochastic volatility-covolatility factors belong to the n state factors that usually drive the short rate, and the state VAR have conditional Gaussian distribution. Some special cases are described below:

- When $m = \frac{K(K+1)}{2}$, the innovations v_t to X_t is subject fully to stochastic volatility-covolatility.

- When $m = 0$, the model collapses to the essentially affine term structure model: $A^{+0}(K) \overset{m=0}{=} A_0(K)$.

- In between, there are intermediate cases, where some volatility-covolatility elements can be restricted to constant. For example, in one case which assumes no correlation risk, all off-diagonal elements for covolatility are restricted to be 0, then $m = K$, and each diagonal elements of Ω_t follows a Chi-square autoregressive process. (The no-arbitrage restrictions of such case is derived in Appendix A 1.4).

- Usually volatility distribution of yields presents high skewness, which means that the degree of freedom in the WAR process is likely to be rather low. In this model, it is restricted that $J \geq K$, where K is the dimension of the stochastic volatility-covolatility.

- When the stochastic volatility-covolatility is characterized by a Gaussian matrix autoregressive process (GMAR), then it becomes a structurally restricted $A_m(K+m)$ model. The restrictions are such that the short rate has zero loadings on those m factors, there is no interaction between the autoregressive dynamics of the two blocks of states, i.e., F is block-diagonal; the variance-covariance Q_t is also block-diagonal, in which the second group of factors $vech\left(\Omega_t\right)$ transforms into the variance-covariance Ω_t for the innovation of the other K state factors X_t. However, one should notice that although the limiting case of a Wishart Autoregressive process is Gaussian when $J \to \infty$, it is unlikely that this limiting process serves to study significant fluctuation in the variance-covariance matrix Ω_t. Because Given a mean of this process $\bar{\Omega}$, when $J \to \infty$, $\eta_t \to 0$, and $\Omega_t \to \bar{\Omega}$, which is constant again. If the Gaussian matrix autoregressive is not a limiting case of WAR, then it is challenging to restrict the parameters such that at any point of time, Ω_t is positive definite. The model restrictions with GMAR process are derived in Appendix A.2.

An interesting feature of $A^{+m}\left(K\right)$ in comparison with $A_0\left(K\right)$ model is that, with the same number of VAR states, there are m more factor dynamics in the $A^{+m}\left(K\right)$ model, but still comparable number of parameters with respect to an $A_0\left(K\right)$ model with time-varying risk prices. Because the number of parameters in the WAR process (M: $K \times K$) is the same as the number of parameters in the time-varying risk price coefficient matrix (λ_1: $K \times K$), just with an additional degree of freedom parameter J. This might help to capture richer dynamics in the yield curve while maintaining the same level of parsimony in parameterization.

4.3 Model extension with leverage effects

One way to incorporate leverage effects is to allow "volatility-in-mean" in the VAR state equation. In this case, although Ω_t is still exogenous, the level of Ω_t also affects X_t, hence there is correlation between Ω_t and X_t. Under this setting, the short rate is indirectly affected by Ω_t through the state variables X_t. However, unlike the square-root or quadratic term structure models, this treatment does not allow the causal effect to run the other way from X_t to the variance-covariance.

Model restrictions with this extension is derived in Appendix A.3. With the macro VAR in mind, this extension may be useful to study the state dynamics of inflation, in which the leverage effect is often observed, i.e., high inflation corresponds to high inflation fluctuation. When X_t is affected by contemporaneous Ω_t, the VAR equation in X_t needs to be transformed with rotation to derive the compact state-space model (M-I)

5 Simulation study

This model has rich implications for the yield curve with respect to stochastic variance-covariance in the state VAR innovations. Its characteristics can be studied by some simulation exercises with simple $A^{+m}(K)$ models. In this section, I first show the results of a simulation for $A^{+1}(1)$ model in comparison with $A_0(1)$ model with constant and time-varying risk price. Then I study the variance-covariance effects jointly from an $A^{+3}(2)$ model. In the end, I compare the time-varying effect of volatility-covolatility from these models with the empirical curvature components in yield data and discuss the link between them.

5.1 $A^{+1}(1)$ model

In an $A^{+1}(1)$ model, there is one state factor that drives short rate, and its innovations are subject to a one-dimension stochastic volatility process. Since the volatility factor loading is not sensitive to the degree of freedom J, and volatility distribution of yields is usually highly skewed, I choose a low degree of freedom to capture this property.

In the simulation study, I choose the following parameter values: $J = 1$, . $\delta_0 = 2.32 \times 10^{-3}/(1-0.95)$, $\delta_1 = 1$, $\mu = 0$, $\Phi = 0.95$, $M = 0.924$, $\Sigma^* \equiv J\Sigma = 1.1 \times 10^{-7}$, $\sigma_\varepsilon = 3 \times 10^{-5}$ and $\Lambda = -400$ I use the steady state values of X and Ω as the initial state values at time $t = 0$.

Figure 3 shows one possible path simulated with $T = 300$. The state X is highly persistent, and Ω presents significant heteroskedasticity.

Figure 4 shows simulated yields of different maturities with one possible path in which $T = 300$ These yields comove with common dynamics.

Figure 5.shows selected yield curves along the simulation path. The $A^{+1}(1)$ can display all kinds of shapes of the term structure: upward sloping, downward sloping, hump shape, inverted hump, etc.

Figure 6. collects the selected yield curves in one graph with the same scale of index. The dashed line represents the average yield curve.

[Figure 3] [Figure 4] [Figure 5] [Figure 6]

Figure 7. is a study on the effects of different levels of stochastic volatility on yield curve given the same level of state X_t. Each graph depicts the yield curve with a certain level of

X_t with its average volatility (dashed line in the middle), high volatility (upper line), and low volatility (lower line), where the high and low volatility is taken from the maximum and minimum realization of a simulated path with $T = 300$. Since the volatility factor is exogenous to the VAR state, for the selected level of X_t, the scenario of high or low level of volatility is with positive probability. The main message here is that the volatility factor makes sizable difference on yield curve, especially in the medium range around 2 years. However, with the parameter values, even at the 10 year maturity, the difference between high and low volatility can be still significant as much as 100 basis points. The last graph shows an interesting scenario where when the average curve flattens out, the volatility dominates the eventual shape of the yield curve, not only in the slope, but also in the direction of the hump. The potential implication for the bond yield conundrum is that the inverted yield curve is a result of low volatility in the short rate at the time compared to previous periods when X_t were at similar level.

[**Figure 7**]

Figure 8 compares the factor loadings and average yield curve of $A^{+1}(1)$ and $A_0(1)$ models assuming that the underlying VAR state X_t has the same mean and autoregressive coefficient, and the underlying variance-covariance matrix has the same mean. For each type of models, there are two specifications as follows.

$A^{+1}(1)$	Baseline model without leverage effect (Blue dotted line)	With leverage effect (Black solid line)
$A_0(1)$	Constant risk price (Red dash-dotted line)	Time-varying risk price (Greet dashed line)

[**Figure 8**]

As can be seen from the graph, the loadings of X_t for all models display almost identical pattern, even the $A_0(1)$ model with time-varying risk price, as the effect of λ_1 is relatively small with respect to the autoregressive coefficient, it is not distinguishable from other models. The loadings for volatility significantly differ between the two types of models. Essentially, since Ω is constant in $A_0(1)$ model, the graph captures only the conceptual coefficient, i.e.

use the components including Ω in the $A_0(1)$ model, $-\frac{1}{n}\left(B_n'\Omega\lambda_0+\frac{1}{2}B_n'\Omega B_n-B_n'\Omega\lambda_1\bar{X}\right)$, to calculate the coefficient for Ω. It turns out that this is also a curvature factor, but with much smaller "loadings" with respect to the $A^{+1}(1)$ models. By examining the loadings on Ω or Ω_t, the "curvature" effect is mainly driven by the Jensen's inequality term $B_n'\Omega B_n$ or $B_n'\Omega_t B_n$. What's particular striking is the high volatility loadings once leverage effect is allowed in the $A^{+1}(1)$ model. Suppose that the underlying VAR has the same parameters and agents have the same risk price, then the $A_0(1)$ model understate significantly the volatility effect. A regime-switching $A_0(1)$ model might account for volatility shift, but only captures a very small proportion of the effect.

5.2 $A^{+3}(2)$ model

After visualizing the volatility effect form a one-factor model, I now calibrate a $A^{+3}(2)$ model with two parameterizations. Then one can see from the factor loadings of the elements in Ω_t, that the general effect of the whole variance-covariance matrix is still a curvature factor. And the covariance coefficient now comes into effect either to mitigate or to propogate the effects of variances, meaning that changing correlation in the shocks to VAR has important implication on the yield dynamics.

I use the Diebold and Li (2005) data on US yield curve from the period of 1984:1-2000:12 together with growth rate of indrustrial production and CPI inflation to calibrate the $A_0(2)$ models. Then I parameterize the WAR process with different coefficient matrix M, to see how the variance covariance factors affect the yield curve. I make two specifications of the $A_0(2)$ model as follows:

	VAR states	δ_1
1	$[\pi_{cpi}, m1]$	$\begin{bmatrix} 0 & 1 \end{bmatrix}'$
2	$[g_{ip}, \pi_{cpi}]$	$\begin{bmatrix} 0.5 & 1.5 \end{bmatrix}'$

In the first specification, m1 denotes the one month rate, which is taken as the proxy of the short rate r. It has the following parameters:

$$\mu = \begin{bmatrix} 0.4 \\ 1.3 \end{bmatrix} \cdot 10^{-4}; \qquad \Phi = \begin{bmatrix} .953 & .024 \\ .015 & .971 \end{bmatrix}; \qquad \delta_0 = 0;$$
$$\bar{\Omega} = \begin{bmatrix} 4.49 & .76 \\ .76 & 10.11 \end{bmatrix} \cdot 10^{-8}; \quad \Lambda = \begin{bmatrix} 5500 \\ -1600 \end{bmatrix}; \qquad M = \begin{bmatrix} .95 & 0 \\ 0 & .95 \end{bmatrix};$$
$$\Sigma^* = \bar{\Omega} - M\bar{\Omega}M'.$$

The second specification has the following parameters:

$$\mu = \begin{bmatrix} 6.2 \\ 0.89 \end{bmatrix} \cdot 10^{-4}; \qquad \Phi = \begin{bmatrix} .928 & -.162 \\ -.002 & .967 \end{bmatrix}; \qquad \delta_0 = \bar{r} - \bar{X}'\delta_1;$$
$$\bar{\Omega} = \begin{bmatrix} 30 & 1.95 \\ 1.95 & 4.54 \end{bmatrix} \cdot 10^{-8}; \quad \Lambda = \begin{bmatrix} 300 \\ -1400 \end{bmatrix}; \qquad M = \begin{bmatrix} 300 \\ -1400 \end{bmatrix};$$
$$\Sigma^* = \bar{\Omega} - M\bar{\Omega}M'.$$

Figure 9 shows the simulated states from the first specification. Here the correlation between the VAR innovations has changed widely over time and even switched signs. Figure 10 shows the factor loadings of the VAR states, the constant and the average yield curve. Due to the specification in δ_1, where the factor loading on CPI inflation is zero, and inflation has influenced longer maturity yields through its VAR coefficients interacted with short rate, hence the factor loading of CPI inflation has also a hump shape, which transmits to the yield curve as a curvature factor. Figure 11shows the net effects of the variance-covariance elements on yields in percentage term. As can be seen from this graph, each element in Ω_t has a humped effect on the yield curve. The lower left panel shows the sum of effects from the covariances $\omega_{12,t}$ and $\omega_{21,t}$. And its eventual total effect depends on the signs and levels of risk prices, and also their time-varying relative effects. In net, the volatility in $\omega_{22,t}$ dominates, but from time to time, the net effect can turn negative due to either high volatility in $\omega_{11,t}$ or variance $\omega_{12,t}$ and most probably due to their joint effect.

[Figure 9] **[Figure 10]** **[Figure 11]**

Figure 12 shows the simulated states from the second specification. Figure 13 and 14 shows the factor loadings. In this specification, both states in X_t display slope effects, and due to the weak correlation factor, the positive volatility effect from $\omega_{22,t}$ dominates that of $\omega_{11,t}$. Most of the time, the net effect of Ω_t is positive.

[**Figure 12**] [**Figure 13**] [**Figure 14**]

5.3 Variance-covariance effect and the curvature factor in yields

The simulation study has shown that the volatility factor and the net effects of the variance-covariance matrix Ω_t are hump shaped curvature factors. This gives strong indication that the empirical curvature factor extracted from the yield curve either by non-parametric methods or the Nielson-Siegle methods is closely related to the stochastic variance-covariance effects of the yield state VAR innovation. On the other hand, the first specification of $A^{+3}(2)$ model with short rate and inflation as state variables implies that rotation of state VAR factors can generate curvature effects per se, so the curvature factor is likely to be a mix of factors from the VAR states and the variance-covariance. Overtime, the different components may magnify itself through the curvature factor when its effect dominates the others.

Figure 15 displays the empirical curvature factor extracted from Diebold-Li data. It fluctuates widely along time, and often switch signs. The second panel shows the simulated effect of Ω_t in an $A^{+1}(1)$ model. The third panel shows the simulated net effect of Ω_t in an $A^{+3}(2)$ model from the first specifacation. From these graphs, we can see again that the shape of the volatility factor corresponds well to the Nelson-Siegel curvature factor; there is significant flucatuation in the variance-covariance effects from the simulated models as well.

[**Figure 15**]

6 Model estimation

Estimation of this class of models can be carried out with different techniques, depending on the modeling property of the stochastic volatility assumed in the model. If it is assumed to be a Gaussian process, then either MLE or MCMC estimation in a Gibbs sampler with Kalman

Filter step can be effective. For the latter, Ang, Dong and Piazzesi (2005) has given a detailed description. When the same method is applied to the ASV-ATSM model, the Kalman Filter step should also take into account the time-varying volatility of the upper part of the Q matrix in the (M-I) model representation.

If, instead, one assumes high skewness in the volatility distribution, hence uses a low degree of freedom WAR process, then one should use MCMC estimation in a Gibbs sampler with Particle Filter step, to deal with the non-Gaussian distribution in the WAR innovations. The model setting provides a clear conditional dynamic structure of the states and yields (see below), the WAR process has also a well-defined probability distribution. so that it provide a natural experiment to explore the recently advanced techiques in particle filters to approximate the discrete time state dynamics in a non-Gaussian setting.

$$
\begin{array}{ccccccccccc}
X_0 & \longrightarrow & X_1 & & \longrightarrow & X_2 & \longrightarrow & \cdots & \longrightarrow & X_T & \\
 & & & \searrow & & & \searrow & & & & \searrow \\
 & & \wedge & & & \wedge & & & & \wedge & \\
 & & \vdots & & Y_1 & \vdots & & Y_2 & \cdots & \vdots & & Y_T \\
 & & & \nearrow & & & \nearrow & & & & \nearrow \\
\Omega_0 & \longrightarrow & \Omega_1 & & \longrightarrow & \Omega_2 & \longrightarrow & \cdots & \longrightarrow & \Omega_T &
\end{array}
$$

In the following, I briefly depict the general procedure. To unfold the whole picture and to explore into details the esimation issue, I reserve this task to a following paper.

6.1 General Procedure

The model can be estimated by MCMC methods and Sequential Importance Sampling-Resampling (particle filter) in a Gibbs sampling algorithm.

For i = 1, choose $\Theta^{(0)}$, $\Omega_{1:T}^{(0)} = \bar{\Omega}$.

1) Given $\Theta^{(0)}$, draw states.

1-1) Given parameters $\Theta^{(0)}$, and assume steady state volatility $\Omega_{1:T}^{(0)} = \bar{\Omega}$, draw $X_{1:T}^{(1)}$ using forward-filtering backward-sampling via Kalman filter(Carter and Kohn (1994)).

1-2) Given parameters $\Theta^{(0)}$ and $X_{1:T}^{(1)}$, draw $\Omega_{1:T}^{(1)}$ by forward filtering and backward smoothing via simulation (Pitt and Shephard (1999) and Godsill, Doucet and West (2004)).

2) Given state sample $Z_{1:T}^{(1)} = \left\{X_{1:T}^{(1)}, vech\left(\Omega_{1:T}^{(1)}\right)\right\}$, draw $\Theta^{(1)}$.

For i = 2:N,

1-1) Conditional on $\Theta^{(i-1)}$ and $\Omega_{1:T}^{(i-1)}$, draw $X_{1:T}^{(i)}$ by forward-filtering backward-sampling via Kalman filter.

1-2) Conditional on $\Theta^{(i-1)}$ and $X_{1:T}^{(i)}$, draw $\Omega_{1:T}^{(i)}$ by forward-filtering backward-sampling via simulation.

2) Conditional on $Z_{1:T}^{(i)} = \left\{X_{1:T}^{(i)}, vech\left(\Omega_{1:T}^{(i)}\right)\right\}$, draw $\Theta^{(i)}$.

Step 1-1) is easily implemented, because given $\Omega_{1:T}$, $X_{1:T}$ evolves with Gaussian error. After a standard Kalman Filter is implemented forward, the state X_t can be sampled backwards. This procedure is proposed in Carter and Kohn (1994). Kim and Nelson(1999) also gives a detailed explanation.

Step 1-2) is implemented with Sequential Importance sampling-Resampling (SIR, or Particle filter) technique to deal with forward-filtering backward-smoothing procedure in a non-Gaussian setting. Auxilliary Particle Filter (APF, Pitt and Shephard (1999)) can be utilized to efficiently sample from the forward filtering proceducre; the backward-smoothing follows the method depicted in Godsill, Doucet and West(2004).

Step 2) is actually implemented by MCMC in a Gibbs sampling algorithm, which is similar to the procedure used in Ang, Dong and Piazzesi (2005), but with the some modifications taking into consideration of the new features of this model .

7 Conclusions

This paper proposes a term structure model where the short rate is driven by a VAR state dynamics, and the variance-covariance matrix of VAR innovations follows a Wishart autoregressive stochastice process. Under this model setting, the time-varying risk premia come from uncertainty in the variance-covariance of innovations to the VAR state factors. And this uncertainty maps into the long maturity yields through no-arbitrage restrictions. Hence the state of volatility, though evolving independently from the VAR state factors which directly determine the short rate, also drives the yield curve at the medium to long maturities, hence deemed as "auxiliary" factors in yields. The model is denoted as Auxiliary Stochastic Volatility Affine Term Structure Model (ASV-ATSM). If the innovations to the volatility-covolatility process are assumed to be Gaussian, it can be categorized in the Dai-Singleton(2002) Affine Term Structure

Model (ATSM) framework,but with a set of structures imposed on the parameter space and the dynamics of states. In the extreme case where the distribution of variance-covariance of innovations to the VAR collapses into a constant, the model converges to the essentially affine $A_0(m)$ model with constant risk price. In another case with Gaussian process of the stochastic volatility-covalitity, the model becomes an $A_m(m+K)$ model, where m is the number of elements driving the stochastic volatility of the K state factors in the short rate.

In this model, both the VAR dynamics and the variance-covariance of VAR innovations affect the yield curve, without much restrictions on the variance-covariance matrix. This flexibility helps to model not only the feature of linear projection of the yield curve level, but also the behaviors of stochastic volatility.

This class of models have some interesting features:

1) volatility is a curvature factor of the yield curve;

2) the time-varying risk premia are directly driven by uncertainties in the variance-covariance of innovations to the VAR states;

3) volatility of the VAR innovations has sizable effects on medium to long maturity yields;

4) simulation study shows that it can well explain the bond yield "conundrum" where although the underlying VAR states remain at the same level, difference in volatility can result in different shapes of the yield curve;

5) it provides a useful tool to jointly study the term structure and macro VAR with stochastic volatility.

Estimation strategies are briefly discussed in this paper. For the case where stochastic volatility-covolatility is represented by a Wishart autoregressive process with low degree of freedom, a MCMC in Gibbs sampler with Auxiliary Particle Filter step is effective in the estimation.

References

1. Ahn, D., Dittmar, R., and A., Gallant, 2002, "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, Vol.15, pp.243-288.

2. Ang, A., and G., Bekaert, 2002, "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20, 2, 163-182

3. Ang, A., S. Dong, and M. Piazzesi, 2007, "No-arbitrage Taylor Rules," Working paper, University of Chicago.

4. Ang, A., and M. Piazzesi, 2003, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 4, 745-787.

5. Ang, A., M. Piazzesi, and M. Wei 2006. "What Does the Yield Curve tell us about GDP Growth?" *Journal of Econometrics* 131, pp. 359-403.

6. *Diebold, F.X. and Li., C. ,2005, "Forecasting the Term Structure of Government Bond Yields." Journal of Econometrics, 131.*

7. Diebold, F.X., Glenn D. Rudebusch, and S. Boragan Aruoba, 2006, "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach." *Journal of Econometrics* 131, 309-338.

8. Godsill, S.J., A. Doucet, and M. West, 2004, "Monte Carlo Smoothing for Nonlinear Time Series," *Journal of American Statistical Association*, Vol.99, No.465.

9. Carter, C.K., and R. Kohn, 1994, "On Gibbs Sampling for State Space Models," *Biometrica*, 81, pp. 541-553.

10. Dai, Q., and K.J. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *The Journal of Finance*, Vol. 55 No. 5, pp. 1943-1978.

11. Duffee, G.R., 2002, "Term premia and interest rate forecasts in affine models." *Journal of Finance*, 57, 405-443.

12. Engle, R.F., and K.F. Kroner, 1995, "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, Vol. 11, No. 1, pp 122-150.

13. Gourieroux, C., 2006, "Continuous Time Wishart Process for Stochastic Risk," *Econometric Reviews*, Vol.25, pp.177-217.

14. Gourieroux, C., J. Jasiak, and R. Sufana, 2005, "The Wishart Autoregressive Process of Multivariate Stochastic Volatility," Working Papers 2005_2, York University, Department of Economics.

15. Gourierous, C., and R., Sufana, 2003, "Wishart Quadratic Term Structure Models," Les Cahiers du CREF of HEC Montreal Working Paper No. 03-10.

16. Justiniano, A., and G., Primiceri, 2007, "The Time Varying Volatility of Macroeconomic Fluctuations," Working paper, Northwestern University. (2006 version: NBER Working Paper No. 12022)

17. Kim, C., and C.R. Nelson, 1999, *State-Space Models with Regime Switching*, MIT Press, Cambridge, MA.

18. Longstaff, F.A., and E.S. Schwartz, 1992, "Interest Rate Volatility and the Term Structure: A Two-factor General Equalibrium Model," Journal of Finance 47, 1259-1282.

19. Rudebusch, G., E. Swanson, 2006, "The Bond Yield "Conundrum" from a Macro-Finance Perspective," Monetary and Economic Studies 24(S-1), 83-128

20. Rudebusch, G., and T. Wu, 2005, "A Macro-Finance Model of the Term Strucutre, Monetary Policy, and the Economy," forthcoming Economic Journal.

21. Pitt, M.K., and N. Shephard, 1999, "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, Vol. 94, No. 446, pp. 590-599.

22. Primiceri, G., 2005, "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, pp. 821-852

23. Singleton, K.J., 2006, *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*, Princeton University Press.

24. Stock, J. H., and M. W. Watson, 2003, "Has the Business Cycle Changed? Evidence and Explanations,"in FRB Kansas City Symposium, Jackson Hole, Wyoming

25. Vasicek, Oldrich, 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics* 5, 177-188.

APPENDIX A. Derivation of the no-arbitrage restrictions on the yield equations

A.1 Derivation when WAR as the stochastic volatility-covolatility process

First of all, to derive the model solution with WAR as the variance-covariance process, I need to use the lemma on conditional Laplace transform of WAR process. The proof of this lemma is presented in Gourieroux, Jasiak, and Sufana(2004).

Lemma 1. **Conditional Laplace transform of WAR process**

The conditional Laplace transform Ψ_t of the WAR process (3)

$$\Omega_t = M\Omega_{t-1}M' + J\Sigma + \eta_t, \quad (J\Sigma + \eta_t) \sim W(J, \Sigma)$$

can be written as:

$$\begin{aligned} \Psi_t(\Gamma) &= E_t\left[\exp Tr\left(\Gamma\Omega_{t+1}\right)|z_t\right] \\ &= E_t\left[\exp\left(z'_{t+1}\Gamma z_{t+1}\right)|z_t\right] \\ &= \frac{\exp\left[z'_t M'\Gamma(I-2\Sigma\Gamma)^{-1}Mz_t\right]}{\left[\det(I-2\Sigma\Gamma)\right]^{J/2}} \\ &= \frac{\exp Tr\left[M'\Gamma(I-2\Sigma\Gamma)^{-1}M\Omega_t\right]}{\left[\det(I-2\Sigma\Gamma)\right]^{J/2}} \end{aligned} \qquad (14)$$

where the argument of the Laplace transform is a symmetric matrix Γ and Tr denotes the trace operator. The Laplace transform is defined for a matrix Γ such that $\left\|2\Sigma^{1/2}\Gamma\Sigma^{1/2}\right\| < 1$.

A.1.1 Pricing kernel

First of all, the pricing kernel defines the equilibrium relationship between the price of yield of maturity n this month with the yield of maturity $n-1$ next month by linking them with the stochastic discount factor m.

$$\begin{aligned} P_t^{(n+1)} &= E_t\left[m_{t+1}P_{t+1}^{(n)}\right] \\ &= E_t\left[\exp\left\{-r_t - \tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda - \Lambda' v_{t+1}\right\}\exp\left\{A_n + B'_n X_{t+1} + C_n\left(\Omega_{t+1}\right)\right\}\right] \\ &= \exp\left\{-r_t + A_n\right\}E_t\left[\exp\left\{-\tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda - \Lambda'_t v_{t+1} + B'_n X_{t+1} + C_n\left(\Omega_{t+1}\right)\right\}\right] \\ &= \exp\left\{-\delta_0 - \delta'_1 X_t + A_n\right\} \\ &\quad \cdot E_t\left[\exp\left\{-\tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda - \Lambda' v_{t+1} + B'_n v_{t+1} + C_n\left(\Omega_{t+1}\right)\right\}\right] \\ &= \exp\left\{-\delta_0 + A_n + B'_n\mu + \left(B'_n\Phi - \delta'_1\right)X_t\right\} \\ &\quad \cdot E_t\left[\exp\left\{\left(-\Lambda' + B'_n\right)v_{t+1} - \tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + C_n\left(\Omega_{t+1}\right)\right\}\right] \\ &= \exp\left\{-\delta_0 + A_n + B'_n\mu + \left(B'_n\Phi - \delta'_1\right)X_t\right\} \cdot H_n(\Omega_{t+1}) \end{aligned}$$

where

$$
\begin{aligned}
H_n(\Omega_{t+1}) &\equiv E_t\left[\exp\left\{\left(-\Lambda' + B_n'\right)v_{t+1} - \tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + C_n\left(\Omega_{t+1}\right)\right\}\right] \\
&= E_t\left[E_{\Omega_{t+1}}\exp\left\{\left(-\Lambda' + B_n'\right)v_{t+1} - \tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + C_n\left(\Omega_{t+1}\right)\right\}\right] \\
&= E_t\left[\left\{E_{\Omega_{t+1}}\exp\left(-\Lambda' + B_n'\right)v_{t+1}\right\}\exp\left\{-\tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + C_n\left(\Omega_{t+1}\right)\right\}\right] \\
&= E_t\left[\exp\left\{E_{\Omega_{t+1}}\left[\left(-\Lambda' + B_n'\right)v_{t+1} + \tfrac{1}{2}var\left(\left(-\Lambda' + B_n'\right)v_{t+1}\right)\right]\right.\right. \\
&\qquad \left.\left.\cdot\exp\left\{-\tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + C_n\left(\Omega_{t+1}\right)\right\}\right\}\right]
\end{aligned}
$$

with

$$
E_{\Omega_{t+1}}\left[\left(-\Lambda' + B_n'\right)v_{t+1}\right] = 0,
$$

and

$$
\begin{aligned}
& E_{\Omega_{t+1}}\left[var\left(\left(-\Lambda' + B_n'\right)v_{t+1}\right)\right] \\
= \;& E_{\Omega_{t+1}}\left[\left(-\Lambda' + B_n'\right)v_{t+1}v_{t+1}'\left(-\Lambda + B_n\right)\right] = \left(-\Lambda' + B_n'\right)\Omega_{t+1}\left(-\Lambda + B_n\right).
\end{aligned}
$$

$$
\Downarrow
$$

$$
\begin{aligned}
H_n(\Omega_{t+1}) &= E_t\left[\exp\left\{\tfrac{1}{2}\left(-\Lambda' + B_n'\right)\Omega_{t+1}\left(-\Lambda + B_n\right)\right\}\exp\left\{-\tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + C_n\left(\Omega_{t+1}\right)\right\}\right] \\
&= E_t\left[\exp\left\{\tfrac{1}{2}\left(-\Lambda' + B_n'\right)\Omega_{t+1}\left(-\Lambda + B_n\right) - \tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + C_n\left(\Omega_{t+1}\right)\right\}\right] \\
&= E_t\left[\exp\left\{-\tfrac{1}{2}B_n'\Omega_{t+1}\Lambda - \tfrac{1}{2}\Lambda'\Omega_{t+1}B_n + \tfrac{1}{2}B_n'\Omega_{t+1}B_n + C_n\left(\Omega_{t+1}\right)\right\}\right] \\
&= E_t\left[\exp\left\{\sum_{j=1}^{J} z_{j,t+1}'\Psi_n z_{j,t+1} + C_n\left(\Omega_{t+1}\right)\right\}|z_t\right]
\end{aligned}
$$

where

$$
\Psi_n \equiv \left(-\frac{1}{2}\Lambda B_n' - \frac{1}{2}B_n\Lambda' + \frac{1}{2}B_n B_n'\right).
$$

Hence,

$$
\begin{aligned}
P_t^{(n+1)} &= \exp\left\{-\delta_0 + A_n + B_n'\mu + \left(B_n'\Phi - \delta_1'\right)X_t\right\} \\
&\quad \cdot E_t\left[\exp\left\{\sum_{j=1}^{J} z_{j,t+1}'\Psi_n z_{j,t+1} + C_n\left(\Omega_{t+1}\right)\right\}|z_t\right]
\end{aligned}
$$

A.1.2. Solution

Second, by utilising the boundary condition $C_1(\Omega_t) = 0$, one can deduce the coefficient restrictions iteratively.

A.1.2.1. Starting from $n = 1$:

$$\begin{aligned} H_1(\Omega_{t+1}) &= E_t\left[\exp\left\{\sum_{j=1}^{J} z'_{j,t+1}\Psi_1 z_{j,t+1} + 0\right\}|z_t\right] \\ &= \frac{\exp Tr\left[M'\Gamma_1(I-2\Sigma\Gamma_1)^{-1}M\Omega_t\right]}{\left[\det(I-2\Sigma\Gamma_1)\right]^{J/2}} \\ &= \exp\left\{D_1 + Tr\left[M'\Gamma_1\left(I-2\Sigma\Gamma_1\right)^{-1}M\Omega_t\right]\right\} \end{aligned}$$

For $n = 1$, $\Gamma_1 = \Psi_1$. Define $\exp D_n \equiv \left[\det\left(I - 2\Sigma\Gamma_n\right)\right]^{-J/2}$, then $D_n = -\frac{J}{2}\ln\left[\det\left(I - 2\Sigma\Gamma_n\right)\right]$.

$$\begin{aligned} P_t^{(2)} &= E_t\left[m_{t+1}P_{t+1}^{(1)}\right] \\ &= \exp\left\{-\delta_0 + A_1 + B_1'\mu + \left(B_1'\Phi - \delta_1'\right)X_t\right\}\cdot H_1(\Omega_{t+1}) \\ &= \exp\left\{-\delta_0 + A_1 + B_1'\mu + D_1 + \left(B_1'\Phi - \delta_1'\right)X_t + Tr\left[M'\Gamma_1\left(I-2\Sigma\Gamma_1\right)^{-1}M\Omega_t\right]\right\} \end{aligned}$$

Therefore,

$$\begin{aligned} A_2 &= -\delta_0 + A_1 + B_1'\mu + D_1, \\ B_2' &= B_1'\Phi - \delta_1', \\ C_2\left(\Omega_t\right) &= Tr\left[M'\Gamma_1\left(I-2\Sigma\Gamma_1\right)^{-1}M\Omega_t\right]. \end{aligned}$$

A.1.2.2. For $n = 2$:

$$\begin{aligned} P_t^{(3)} &= E_t\left[m_{t+1}P_{t+1}^{(2)}\right] \\ &= \exp\left\{-\delta_0 + A_2 + B_2'\mu + \left(B_2'\Phi - \delta_1'\right)X_t\right\}\cdot H_2(\Omega_{t+1}) \\ &= \exp\left\{-\delta_0 + A_2 + B_2'\mu + \left(B_2'\Phi - \delta_1'\right)X_t\right\} \\ &\quad \cdot E_t\left[\exp\left\{\sum_{j=1}^{J} z'_{j,t+1}\Psi_2 z_{j,t+1} + Tr\left[M'\Gamma_1\left(I-2\Sigma\Gamma_1\right)^{-1}M\Omega_{t+1}\right]\right\}|z_t\right] \\ &= \exp\left\{-\delta_0 + A_2 + B_2'\mu + \left(B_2'\Phi - \delta_1'\right)X_t\right\} \\ &\quad \cdot E_t\left[\exp\left\{\sum_{j=1}^{J} z'_{j,t+1}\left[\Psi_2 + M'\Gamma_1\left(I-2\Sigma\Gamma_1\right)^{-1}M\right]z_{j,t+1}\right\}|z_t\right] \end{aligned}$$

Define $\Gamma_2 \equiv \Psi_2 + M'\Gamma_1\left(I - 2\Sigma\Gamma_1\right)^{-1}M$,

$$\begin{aligned} P_t^{(3)} &= \exp\left\{-\delta_0 + A_2 + B_2'\mu + \left(B_2'\Phi - \delta_1'\right)X_t\right\}\cdot E_t\left[\exp\left\{\sum_{j=1}^{J} z'_{j,t+1}\Gamma_2 z_{j,t+1}\right\}|z_t\right] \\ &= \exp\left\{-\delta_0 + A_2 + B_2'\mu + \left(B_2'\Phi - \delta_1'\right)X_t\right\}\cdot\exp\left\{D_2 + Tr\left[M'\Gamma_2\left(I-2\Sigma\Gamma_2\right)^{-1}M\Omega_t\right]\right\} \\ &= \exp\left\{-\delta_0 + A_2 + B_2'\mu + D_2 + \left(B_2'\Phi - \delta_1'\right)X_t + Tr\left[M'\Gamma_2\left(I-2\Sigma\Gamma_2\right)^{-1}M\Omega_t\right]\right\} \end{aligned}$$

Therefore,

$$\begin{aligned} A_3 &= -\delta_0 + A_2 + B_2'\mu + D_2, \\ B_3' &= B_2'\Phi - \delta_1', \\ C_3(\Omega_t) &= Tr\left[M'\Gamma_2 (I - 2\Sigma\Gamma_2)^{-1} M\Omega_t\right]. \end{aligned}$$

A.1.2.3. Iterate forward, the general solution for $n > 1$:

$$\begin{aligned} A_{n+1} &= -\delta_0 + A_n + B_n'\mu + D_n, \\ &= (n+1)A_1 + B_1'\left[\sum_{i=0}^{n-1}\Phi^i\right]\mu + \sum_{i=1}^{n} D_i \\ B_{n+1}' &= B_n'\Phi - \delta_1', \\ C_{n+1}(\Omega_t) &= Tr\left[M'\Gamma_n (I - 2\Sigma\Gamma_n)^{-1} M\Omega_t\right]. \end{aligned}$$

with

$$\begin{aligned} A_1 &= -\delta_0 \\ B_1 &= -\delta_1 \\ G_1 &= \mathbf{0} \end{aligned}$$

and

$$\begin{aligned} G_{n+1} &\equiv M'\Gamma_n (I - 2\Sigma\Gamma_n)^{-1} M \text{ for } n > 0 \\ D_n &\equiv -\tfrac{J}{2}\ln\left[\det(I - 2\Sigma\Gamma_n)\right] \end{aligned}$$

with Γ_n defined as the following:

$$\Gamma_n = -\tfrac{1}{2}\Lambda B_n' - \tfrac{1}{2}B_n\Lambda' + \tfrac{1}{2}B_nB_n' + G_n$$

A.1.3. An alternative presentation for the no-arbitrage coefficients

In order to understand intuitively how these restrictions are imposed directly on the coefficients in the yield equation, we can write them in the following affined form.

Given that

$$p_{t,t+n} = A_n + B_n'X_t + C_n(\Omega_t),$$

$$y_{t,t+n} = a_n + b_n'X_t + c_n(\Omega_t) = \frac{-1}{n}\left(A_n + B_n'X_t + C_n(\Omega_t)\right),$$

we can derive

$$\begin{array}{lcl} b_{n+1} & = & \frac{1}{(n+1)}\left[\sum_{i=0}^{n}\left(\Phi'\right)^{i}\right]b_1 \\ a_{n+1} & = & = a_1 + \frac{1}{(n+1)}b_1'\left[\sum_{i=0}^{n-1}\Phi^{i}\right]\mu - \frac{1}{(n+1)}\sum_{i=1}^{n}D_i \\ c_{n+1}\left(\Omega_t\right) & = & -\frac{1}{(n+1)}Tr\left[G_{n+1}\Omega_t\right] \end{array} .$$

A.1.4. Special case: Independent One-dimension Wishart(Chi-square) or Gaussian Autoregressive Process

A.1.4.1. Chi-square Autoregressive Process

When Ω_t is assumed to be strictly diagonal, i.e., no correlation risk, then the diagonal elements are independent Wishart (Chi-square) autoregressive process of dimension 1:

$$\Omega_t = \begin{bmatrix} \omega_{11,t} & & \\ & \ddots & \\ & & \omega_{KK,t} \end{bmatrix}_{K\times K} .$$

Each ω_{ii} follows a WAR(1) as:

$$\omega_{ii,t} = m_i^2\omega_{ii,t-1} + J_i\sigma_i^2 + \eta_{i,t}, \quad J_i\sigma_i^2 + \eta_{i,t} \sim W_1\left(J_i, \sigma_i^2\right) ,$$

where $W_1\left(J_i, \sigma_i^2\right)$ is equivalent to $\sigma_i^2\chi\left(J_i\right)$.

Assume A and B are each a $K \times 1$ vector, with a_i and b_i as their ith elements, then

$$A'\Omega_t B = \sum_{i=1}^{K} a_i\omega_{ii}b_i .$$

With Ω_t restricted as such, Σ, M, Ψ_n, Γ_n, and G_n are also diagonal, with their ith diagonal elements as:

$$\begin{array}{lcl} G_{1,i} & = & 0 \\ G_{n+1,i} & = & m_i^2\Gamma_{ni}\left(I - 2\sigma_i^2\Gamma_{n,i}\right)^{-1} \\ \Gamma_{n,i} & = & -\lambda_i B_{n,i} + \frac{1}{2}B_{n,i}^2 + G_{n,i} \\ D_n & = & -\sum_{i=1}^{K}\frac{J_i}{2}\ln\left(I - 2\sigma_i^2\Gamma_{n,i}\right) \\ Tr\left(G_{n+1}\Omega_t\right) & = & \sum_{i=1}^{K} G_{n+1,i}\omega_{ii,t} \end{array} .$$

A.1.4.2. Cholesky Decomposition and Ω_t with Chi-square Autoregressive Process

Suppose Ω_t can be represented by a Cholesky decomposition:

$$\Omega_t = H\tilde{\Omega}_t H',$$

where $\tilde{\Omega}_t$ is diagonal at any time (no correlation risk), and $E(\tilde{\Omega}_t) = I$ so that $E(\Omega_t) = HH'$. H is a lower triangular matrix with h_{ji} as its element in the jth row and ith colomn.

Then $A'\Omega_t B = \tilde{A}'\tilde{\Omega}_t\tilde{B}$, where A and B are each a $K \times 1$ vector, $\tilde{A}' = A'H$, $\tilde{B} = H'B$. The ith element of them are respectively: $\tilde{A}_i = \tilde{A}'_i = \sum_{j=i}^{K} a_j h_{ji}$, $\tilde{B}_i = \tilde{B}'_i = \sum_{j=i}^{K} b_i h_{ji}$.

A.2 Derivation when Gaussian matrix autoregressive process represents the stochastic volatility-covolatility process

A.2.1 Lemma.

The conditional Laplace transform Ψ_t of the Gaussian matrix autoregressive process

$$\Omega_t = M\Omega_{t-1}M' + \Sigma^* + \eta_t,\ vec\left(\eta_t\right) \sim N_{K^2}\left(0, V_\eta\right) \tag{15}$$

is:

$$\begin{aligned} \Psi_t\left(\Gamma\right) &= E_t\left[\exp Tr\left(\Gamma\Omega_{t+1}\right)\right] \\ &= \exp\left\{tr\left(\Gamma\Sigma^*\right) + \tfrac{1}{2}vec\left(\Gamma'\right)' V_\eta vec\left(\Gamma'\right) + tr\left(M'\Gamma M\Omega_t\right)\right\} \end{aligned} \tag{16}$$

Proof:

$$\begin{aligned} \Psi_t\left(\Gamma\right) &= E_t\left[\exp Tr\left(\Gamma\Omega_{t+1}\right)\right] \\ &= E_t\left[\exp\left\{tr\left[\Gamma\left(M\Omega_t M' + \Sigma^* + \eta_{t+1}\right)\right]\right\}\right] \\ &= \exp\left\{tr\left[\Gamma\left(M\Omega_t M' + \Sigma^*\right)\right]\right\} \cdot E_t\left[\exp\left\{tr\left(\Gamma\eta_{t+1}\right)\right\}\right] \\ &= \exp\left\{tr\left(\Gamma\Sigma^*\right) + tr\left(M'\Gamma M\Omega_t\right)\right\} \cdot E_t\left[\exp\left\{tr\left(\Gamma\eta_{t+1}\right)\right\}\right] \end{aligned}$$

According to $tr(A \cdot B) = vec(A')' \cdot vec(B)$

$$tr\left(\Gamma\eta_{t+1}\right) = vec\left(\Gamma'\right)' vec\left(\eta_{t+1}\right)$$

then

$$
\begin{aligned}
E_t\left[\exp\left\{tr\left(\Gamma\eta_{t+1}\right)\right\}\right] &= E_t\left[\exp\left\{vec\left(\Gamma'\right)' vec\left(\eta_{t+1}\right)\right\}\right] \\
&= \exp\left\{E_t\left[vec\left(\Gamma'\right)' vec\left(\eta_{t+1}\right)\right] + \tfrac{1}{2}var\left[vec\left(\Gamma'\right)' vec\left(\eta_{t+1}\right)\right]\right\} \\
&= \exp\left\{0 + \tfrac{1}{2}vec\left(\Gamma_1'\right)' V_\eta vec\left(\Gamma_1'\right)\right\}
\end{aligned}
$$

$$
\Longrightarrow \Psi_t\left(\Gamma\right) = \exp\left\{tr\left(\Gamma\Sigma^*\right) + \tfrac{1}{2}vec\left(\Gamma'\right)' V_\eta vec\left(\Gamma'\right) + tr\left(M'\Gamma M\Omega_t\right)\right\}
$$

A.2.2 Matrix transformation

- A useful matrix transformation result is needed:If A, B, C, D are $K \times 1$ vectors each, and Ω is $K \times K$ symmetric positive definite matrix, then:

$$
A'\Omega B + C'\Omega D = tr\left(\begin{bmatrix} A' \\ B' \end{bmatrix}\Omega\begin{bmatrix} C & D \end{bmatrix}\right) = tr\left(\begin{bmatrix} A & B \end{bmatrix}'\Omega\begin{bmatrix} C & D \end{bmatrix}\right)
$$

- Define $\Psi_{L,n} \equiv \begin{bmatrix} -\frac{1}{2}B_n' \\ -\frac{1}{2}\Lambda' \\ \frac{1}{2}B_n' \end{bmatrix}_{3\times K}$, and $\Psi_{R,n} \equiv \begin{bmatrix} \Lambda & B_n & B_n \end{bmatrix}_{K\times 3}$.

One can show that

$$
\Psi_{R,n}\Psi_{L,n} = \Psi_n = -\frac{1}{2}\Lambda B_n' - \frac{1}{2}B_n\Lambda' + \frac{1}{2}B_nB_n'.
$$

A.2.3. Derivation results

Following similar steps as in Appendix A.1, and using the above results, the derivation can be easily carried out and the restrictions have the following form:

$$
\begin{aligned}
A_1 &= -\delta_0 \\
B_1 &= -\delta_1 \\
C_n(\Omega_t) &= tr\left(G_n\Omega_t\right) \\
G_1 &= \mathbf{0} \\
\Psi_n &= -\tfrac{1}{2}\Lambda B_n' - \tfrac{1}{2}B_n\Lambda' + \tfrac{1}{2}B_nB_n' \\
\Gamma_1 &= \Psi_1 \\
\Gamma_{n+1} &= \Psi_{,n+1} + M'\Gamma_nM \\
G_{n+1} &= M'\Gamma_nM \\
D_n &= tr\left(\Gamma_n\Sigma^*\right) + \tfrac{1}{2}vec\left(\Gamma_n'\right)'V_\eta vec\left(\Gamma_n'\right)
\end{aligned}
$$

Notice, that the main differences are in the expression of G_{n+1} and D_n, which reflects the normal distribution of $vec\left(\eta_t\right)$.

A.3 Derivation of model extension with leverage effects (volatility-in-mean)

Specification: volatility in mean

$$
\begin{aligned}
X_t &= \mu + \Phi X_{t-1} + f\left(\Omega_t\right) + v_t, v_t \sim N(0,\Omega_t) \\
\text{with } X_{t,i} &= \mu_i + \Phi_iX_{t-1} + \psi_i'\Omega_t\psi_i + v_{t,i},
\end{aligned}
$$

where A_i denotes the ith row of A, and ψ_i is a $K\times 1$ vector.

A3.1 Pricing kernel

First of all, the pricing kernel defines the equilibrium relationship between the price of yield of maturity n this month with the yield of maturity $n-1$ next month by linking them with the stochastic discount factor m.

$$
P_t^{(n+1)} = \quad \exp\left\{-\delta_0 + A_n + B_n'\mu + \left(B_n'\Phi - \delta_1'\right)X_t\right\}\cdot H_n(\Omega_{t+1})
$$

where

$$
\begin{aligned}
H_n(\Omega_{t+1}) &\equiv E_t\left[\exp\left\{\left(-\Lambda' + B_n'\right)v_{t+1} - \tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + B_n'f\left(\Omega_{t+1}\right) + C_n\left(\Omega_{t+1}\right)\right\}\right] \\
&= E_t\left[\exp\left\{E_{\Omega_{t+1}}\left[\left(-\Lambda' + B_n'\right)v_{t+1} + \tfrac{1}{2}var\left(\left(-\Lambda' + B_n'\right)v_{t+1}\right)\right]\right.\right. \\
&\quad \left.\left.\cdot\exp\left\{-\tfrac{1}{2}\Lambda'\Omega_{t+1}\Lambda + B_n'f\left(\Omega_{t+1}\right) + C_n\left(\Omega_{t+1}\right)\right\}\right\}\right]
\end{aligned}
$$

With

$$B_n' f(\Omega_{t+1}) = \sum_{i=1}^{K} B_{n,i} \psi_i' \Omega_{t+1} \psi_i$$

Then

$$H_n(\Omega_{t+1}) = \quad E_t \left[\exp \left\{ \sum_{j=1}^{J} z_{j,t+1}' \Psi_n z_{j,t+1} + C_n(\Omega_{t+1}) \right\} |z_t \right]$$

where

$$\Psi_n \equiv \left(-\frac{1}{2} \Lambda B_n' - \frac{1}{2} B_n \Lambda' + \frac{1}{2} B_n B_n' + \sum_{i=1}^{K} B_{n,i} \psi_i \psi_i' \right)$$

Hence,

$$\begin{aligned} P_t^{(n+1)} = \quad & \exp\left\{ -\delta_0 + A_n + B_n' \mu + \left(B_n' \Phi - \delta_1' \right) X_t + B_1' f(\Omega_t) \right\} \\ & \cdot E_t \left[\exp \left\{ \sum_{j=1}^{J} z_{j,t+1}' \Psi_n z_{j,t+1} + C_n(\Omega_{t+1}) \right\} |z_t \right] \end{aligned}$$

A3.2 Solution

Iterate forward, the solution can be derived similarly,

$$\begin{aligned} A_1 &= -\delta_0 \\ B_1 &= -\delta_1 \\ C_1(\Omega_t) &= \mathbf{0} \\ G_1 &= \mathbf{0} \\ \Psi_n &= \left(-\tfrac{1}{2} \Lambda B_n' - \tfrac{1}{2} B_n \Lambda' + \tfrac{1}{2} B_n B_n' + \textstyle\sum_{i=1}^{K} B_{n,i} \psi_i \psi_i' \right) \\ \Gamma_1 &= \Psi_1 \\ G_{n+1} &= \textstyle\sum_{i=1}^{K} tr \left[\left(B_{n,i} \psi_i \right) \psi_i' \Omega_t \right] + M' \Gamma_n \left(I - 2 \Sigma \Gamma_n \right)^{-1} M \\ \Gamma_{n+1} &= \Psi_{n+1} + G_{n+1} \\ D_n &= -\tfrac{J}{2} \ln \left[\det \left(I - 2 \Sigma \Gamma_n \right) \right] \end{aligned}$$

APPENDIX B. Implied forward rate and excess returns

B.1. Forward rate

$$
\begin{aligned}
f_{t,n}^{(1)} &= p_{t,t+n-1} - p_{t,t+n} \\
&= A_{n-1} + B_{n-1}' X_t + Tr\left(G_{n-1}\Omega_t\right) \\
&\quad - A_n - B_n' X_t - Tr\left(G_n \Omega_t\right) \\
&= \left(A_{n-1} - A_n\right) + \left(B_{n-1}' - B_n'\right) X_t \\
&\quad + Tr\left[\left(G_{n-1} - G_n\right)\Omega_t\right]
\end{aligned}
$$

B.2. Excess returns

$$
\begin{aligned}
rx_{t+1}^n &= p_{t+1,t+n-1} - p_{t,t+n} - y_{t,t+1} \\
&= A_{n-1} + B_{n-1}' X_{t+1} + Tr\left[G_{n-1} E_t\left(\Omega_{t+1}\right)\right] \\
&\quad - A_n - B_n' X_t - Tr\left(G_n\Omega_t\right) + A_1 + B_1' X_t \\
&= A_{n-1} + B_{n-1}'\left(\mu + \Phi X_t + v_{t+1}\right) + Tr\left[G_{n-1}\left(M\Omega_t M' + J\Sigma + \eta_{t+1}\right)\right] \\
&\quad - A_n - B_n' X_t - Tr\left(G_n\Omega_t\right) + A_1 + B_1' X_t \\
&= A_{n-1} + A_1 - A_n + B_{n-1}'\left(\mu + \Phi X_t\right) + B_1' X_t - B_n' X_t + JTr\left[G_{n-1}\Sigma\right] \\
&\quad + Tr\left[\left(M' G_{n-1} M - G_n\right)\Omega_t\right] + \underbrace{B_{n-1}' v_{t+1} + Tr\left[G_{n-1}\eta_{t+1}\right]}_{g(v_{t+1},\eta_{t+1})} \\
&= -D_{n-1} + B_{n-1}'\Phi X_t + \left(B_1' - B_{n-1}'\Phi - B_1'\right) X_t + JTr\left(G_{n-1}\Sigma\right) \\
&\quad + Tr\left[\left(M' G_{n-1} M - G_n\right)\Omega_t\right] + g(v_{t+1},\eta_{t+1}) \\
&= -D_{n-1} + JTr\left(G_{n-1}\Sigma\right) + Tr\left[\left(M' G_{n-1} M - G_n\right)\Omega_t\right] + g(v_{t+1},\eta_{t+1}) \\
&= const. + Tr\left[\left(M' G_{n-1} M - G_n\right)\Omega_t\right] + g(v_{t+1},\eta_{t+1})
\end{aligned}
$$

Expected excess return:

$$
\begin{aligned}
E_t\left[rx_{t+1}^n\right] &= E_t\left[p_{t+1,t+n-1}\right] - p_{t,t+n} - y_{t,t+1} \\
&= const. + Tr\left[\left(M' G_{n-1} M - G_n\right)\Omega_t\right]
\end{aligned}
$$

Excess returns in general:

$$\begin{aligned}
rx_{t+s}^{n} &= p_{t+s,t+n-s} - p_{t,t+n} + p_{t,t+s} \\
&= A_{n-s} + B_{n-s}' X_{t+s} + Tr\left[G_{n-s}\Omega_{t+s}\right] - A_n - B_n' X_t - Tr\left(G_n \Omega_t\right) \\
&\quad + A_s + B_s' X_t + Tr\left(G_s \Omega_t\right) \\
&= A_{n-s} + B_{n-s}' \left(\sum_{i=0}^{s} \Phi^i \mu + \Phi^s X_t + \sum_{i=1}^{s} \Phi^{s-i} v_{t+i} \right) \\
&\quad + Tr\left\{ G_{n-s} \left[M^s \Omega_t (M^s)' + J\Sigma(s) + \sum_{i=1}^{s} M^{s-i} \eta_{t+i} \left(M^{s-i}\right)' \right] \right\} \\
&\quad - A_n - B_n' X_t - Tr\left(G_n \Omega_t\right) + A_s + B_s' X_t + Tr\left(G_s \Omega_t\right) \\
&= A_{n-s} - A_n + A_s + B_{n-s}' \left(\sum_{i=0}^{s} \Phi^i \mu + \Phi^s X_t \right) - B_n' X_t + B_s' X_t \\
&\quad + JTr\left[G_{n-s}\Sigma(s)\right] + Tr\left\{ \left[(M^s)' G_{n-s} M^s - G_n + G_s \right] \Omega_t \right\} \\
&\quad \underbrace{+ B_{n-s}' \sum_{i=1}^{s} \Phi^{s-i} v_{t+i} + Tr\left\{ G_{n-s} \left[\sum_{i=1}^{s} M^{s-i} \eta_{t+i} \left(M^{s-i}\right)' \right] \right\}}_{g\left(\{v_{t+i}\},\{\eta_{t+i}\}\right)} \\
&= A_{n-s} - A_n + A_s + B_{n-s}' \left(\sum_{i=0}^{s} \Phi^i \mu \right) + \left(B_{n-s}' \Phi^s - B_n' + B_s' \right) X_t \\
&\quad + JTr\left[G_{n-s}\Sigma(s)\right] + Tr\left\{ \left[(M^s)' G_{n-s} M^s - G_n + G_s \right] \Omega_t \right\} + g\left(\{v_{t+i}\}, \{\eta_{t+i}\} \right) \\
&= A_{n-s} - A_n + A_s + B_{n-s}' \left(\sum_{i=0}^{s} \Phi^i \mu \right) + JTr\left[G_{n-s}\Sigma(s)\right] \\
&\quad Tr\left\{ \left[(M^s)' G_{n-s} M^s - G_n + G_s \right] \Omega_t \right\} + g\left(\{v_{t+i}\}, \{\eta_{t+i}\} \right) \\
&= const. + Tr\left\{ \left[(M^s)' G_{n-s} M^s - G_n + G_s \right] \Omega_t \right\} + g\left(\{v_{t+i}\}, \{\eta_{t+i}\} \right)
\end{aligned}$$

where the deleting of X_t terms is due to the following facts:

$$B_{n+1}' = B_1' \left[\sum_{i=0}^{n} \Phi^i \right]$$

$$\begin{aligned}
B_{n-s}' \Phi^s - B_n' + B_s' &= b_1' \left[\sum_{i=0}^{n-s-1} \Phi^i \right] \Phi^s - b_1' \left[\sum_{i=0}^{n-1} \Phi^i \right] + b_1' \left[\sum_{i=0}^{s-1} \Phi^i \right] \\
&= b_1' \left(\sum_{i=s}^{n-1} \Phi^i - \sum_{i=0}^{n-1} \Phi^i + \sum_{i=0}^{s-1} \Phi^i \right) = 0
\end{aligned}$$

Expected excess returns in general:

$$\begin{aligned}
E_t\left[rx_{t+s}^{n} \right] &= E_t\left[p_{t+s,t+n-s} \right] - p_{t,t+n} + p_{t,t+s} \\
&= const. + Tr\left\{ \left[(M^s)' G_{n-s} M^s - G_n + G_s \right] \Omega_t \right\}
\end{aligned}$$

APPENDIX C. Matrix transformation for the derivation of the compact state-space model

Some useful transformation and results to simplify the model representation.

1. Vectorization of the trace products of two symmetric matrices

Using a property of the *vec* operator:

$$vec(A')' \cdot vec(B) = tr(B \cdot A) = trace(A \cdot B) = vec(B')' \cdot vec(A),$$

if A and B are both symmetric matrices with dimension n, then

$$tr(A \cdot B) = vec(A)' \cdot vec(B).$$

2. Vectorized presentation of the Wishart Autoregressive process

Using two properties of the *vec* operator,

$$\begin{aligned} vec(A+B) &= vec(A) + vec(B) \\ vec(ABC) &= (C' \otimes A)vec(B) \end{aligned}$$

the vectorized presentation of $\Omega_t = M\Omega_{t-1}M' + \Sigma^* + \eta_t$ can be written as:

$$vec\,(\Omega_t) = (M \otimes M)\,vec\,(\Omega_{t-1}) + vec\,(\Sigma^*) + vec\,(\eta_t)\,. \tag{17}$$

In the steady state:

$$vec\,(\Omega) = (M \otimes M)\,vec\,(\Omega) + vec\,(\Sigma^*)$$

$$\Downarrow$$

$$vec\,(\Omega) = (I_{k^2} - M \otimes M)^{-1} vec\,(\Sigma^*) \tag{18}$$

Figure 1. Factor loadings of yields with one state in X and one volatility factor

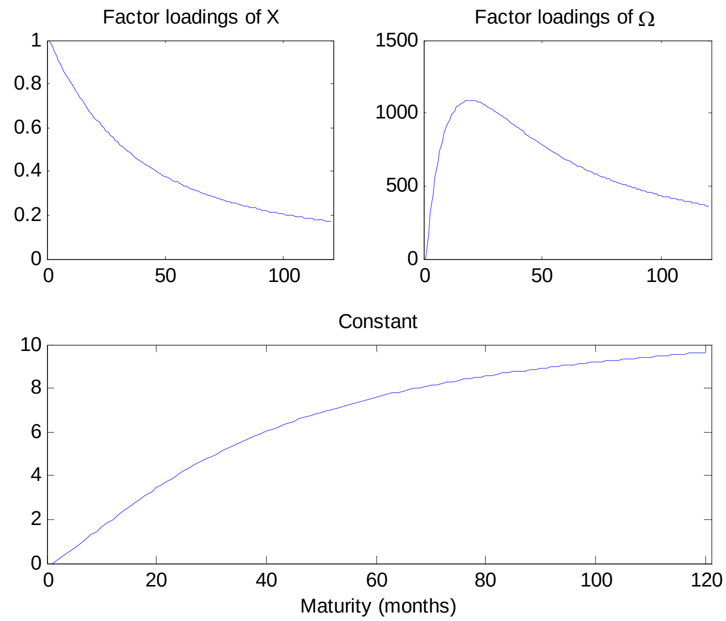


Figure 2. Parameters affacting volatility factor loading

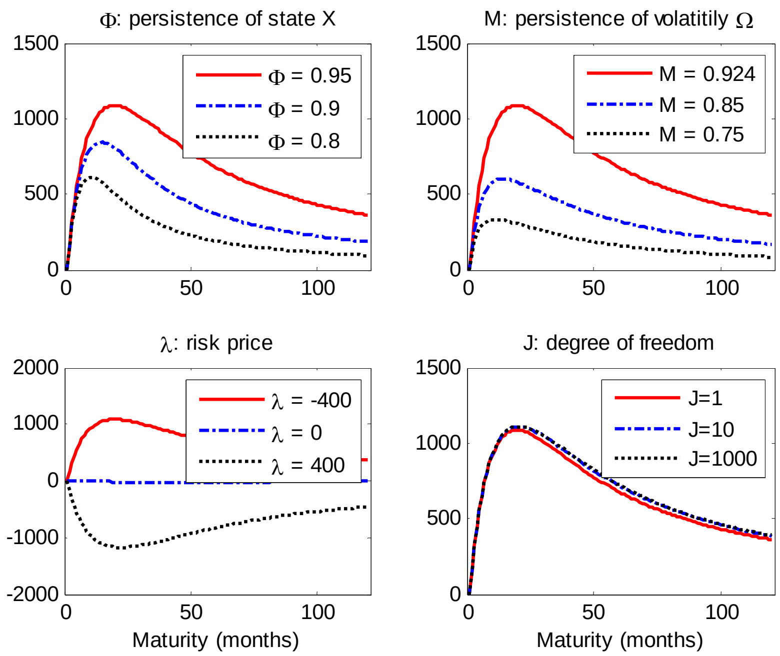


Figure 3. $A^{+1}(1)$ model simulation of states

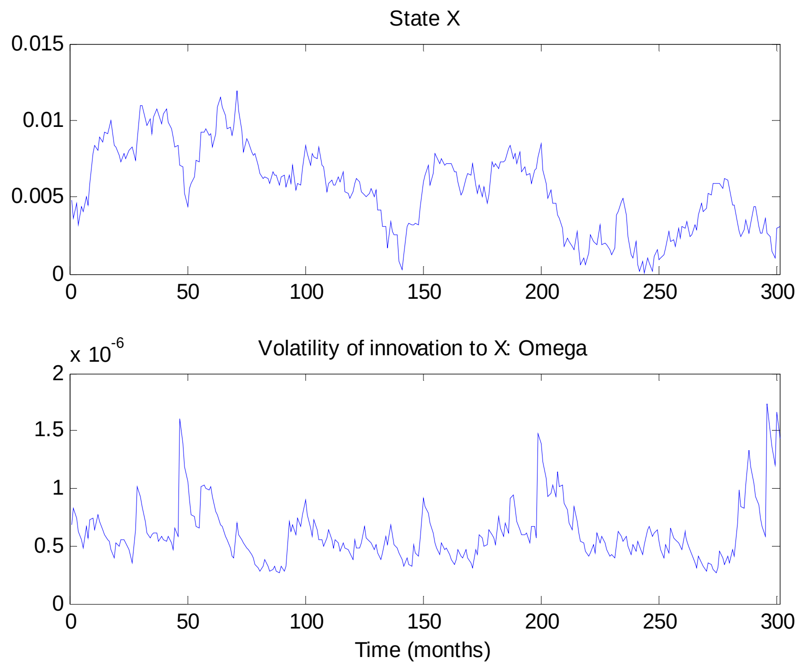


Figure 4. Simulated yields from an $A^{+1}(1)$ model

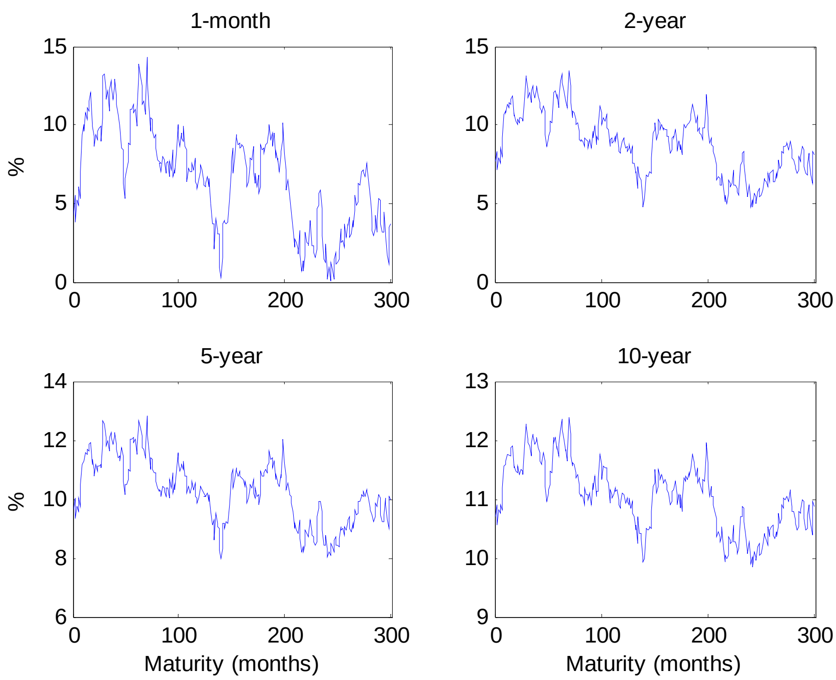


Figure 5. Simulated yield curve from an $A^{+1}(1)$ model (I)

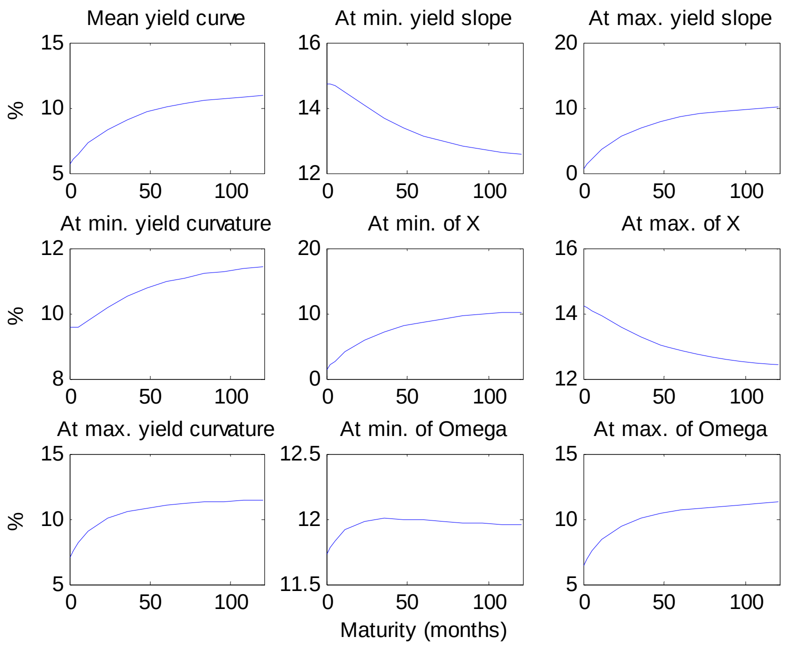


Figure 6. Simulated yield curve from an $A^{+1}(1)$ model (II)

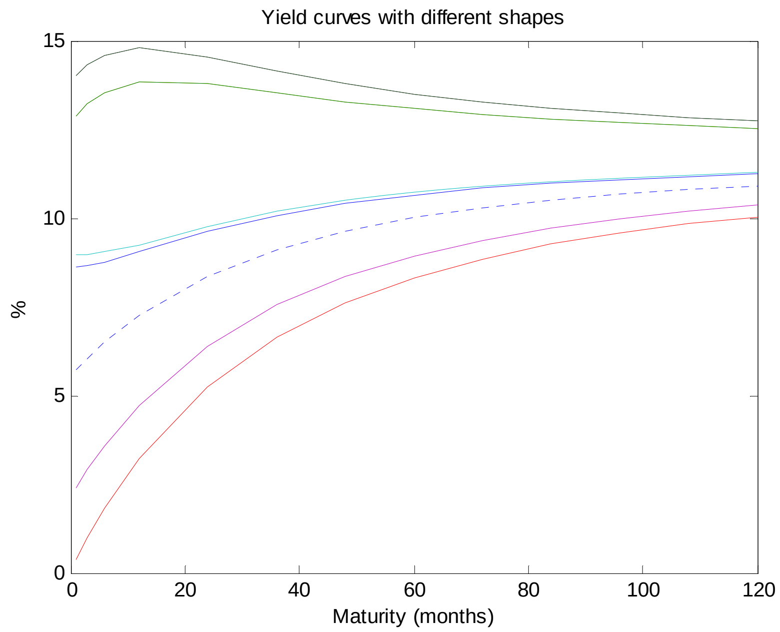


Figure 7. Simulated yield curve with same state X but different Ω_t

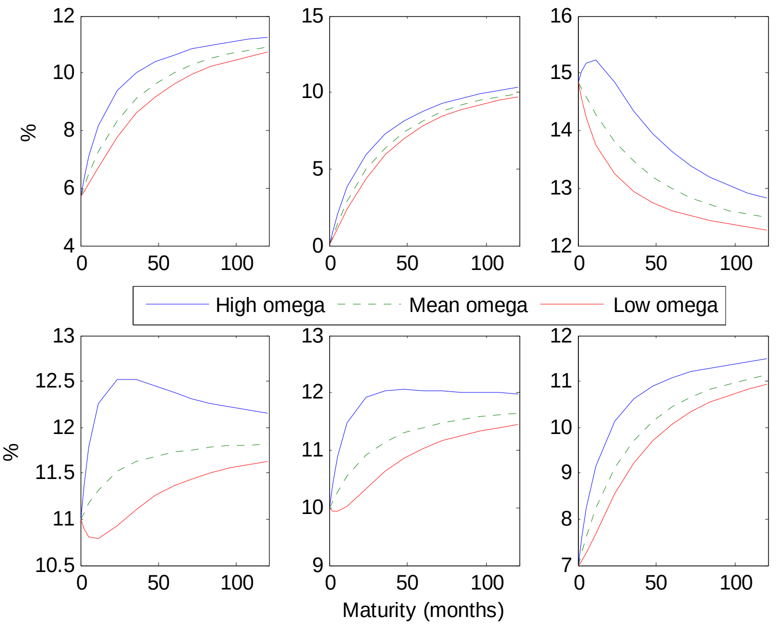


Figure 8. Comparison of factor loadings of $A^{+1}(1)$ and $A_0(1)$ model

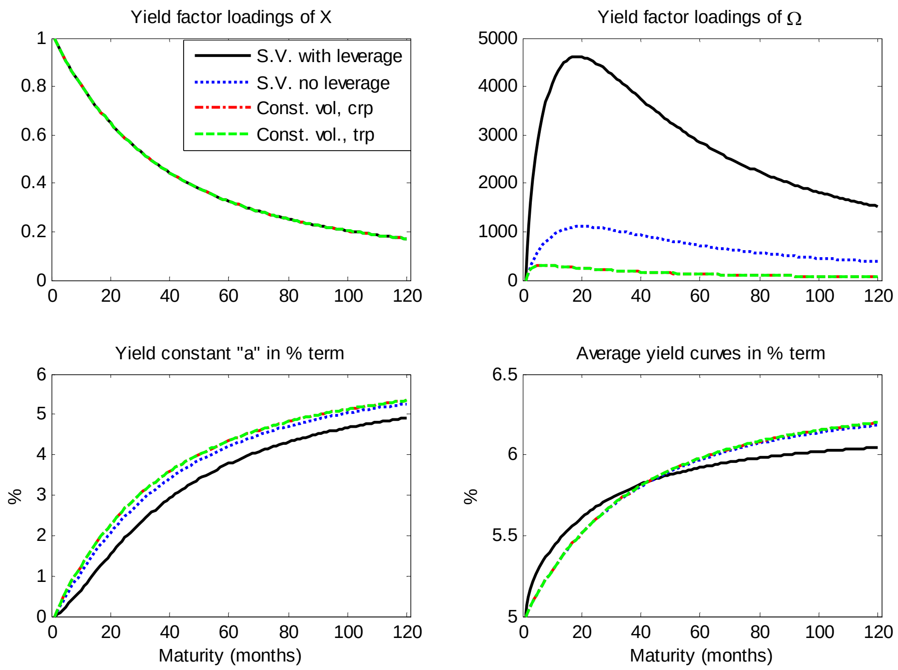


Figure 9. Simulated states from an $A^{+3}(2)$ model

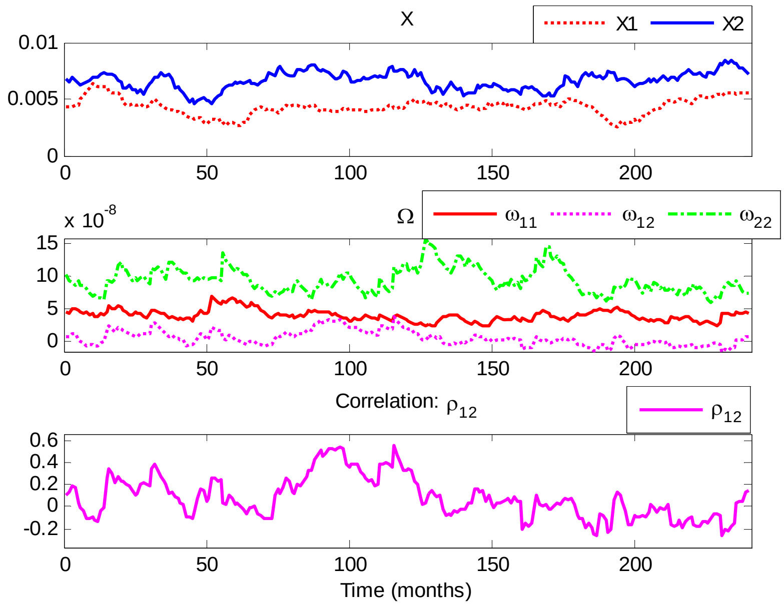


Figure 10. **Figure 11.**

Factor loadings from simulated $A^{+3}(2)$ model : specification I

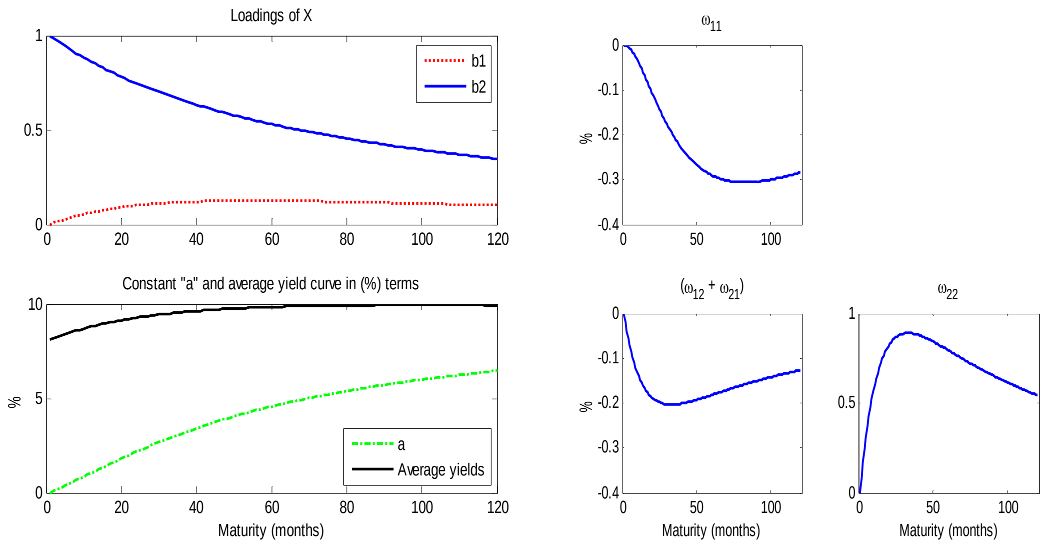


Figure 12. Simulated states from an $A^{+3}(2)$ model

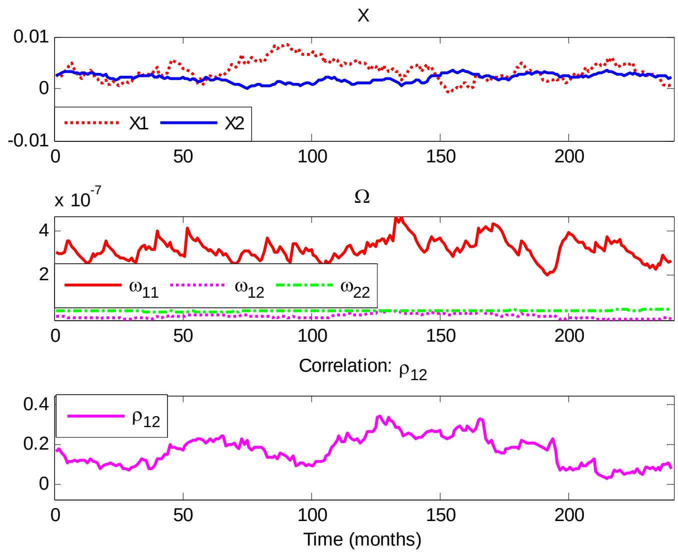


Figure 13. **Figure 14.**

Factor loadings from simulated $A^{+3}(2)$ model : specification II

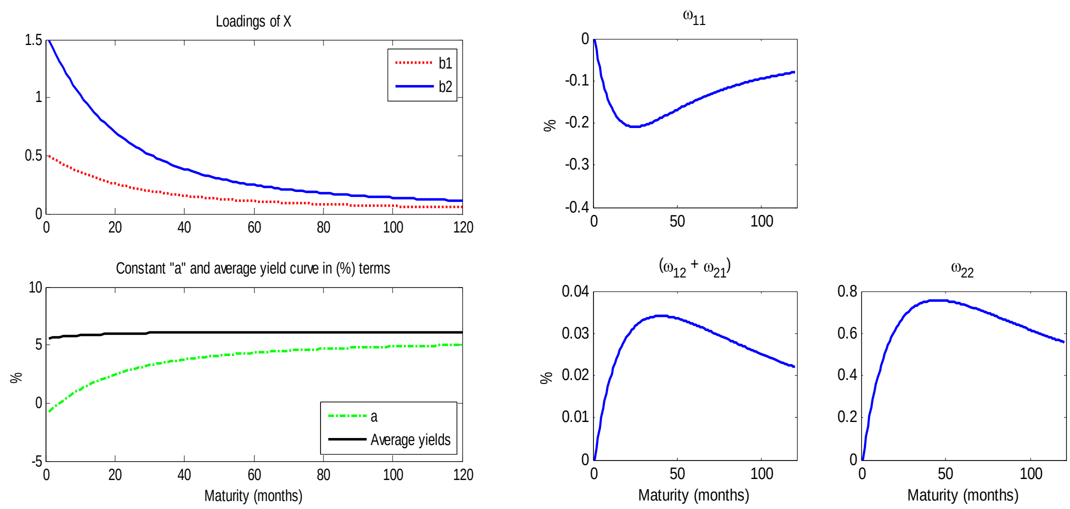
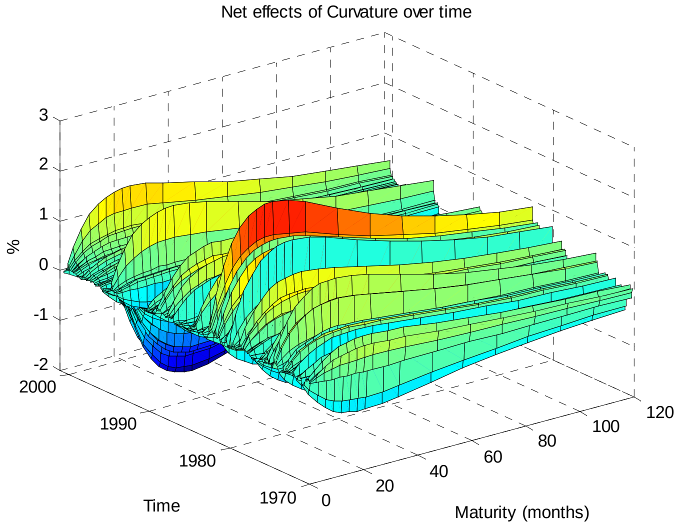
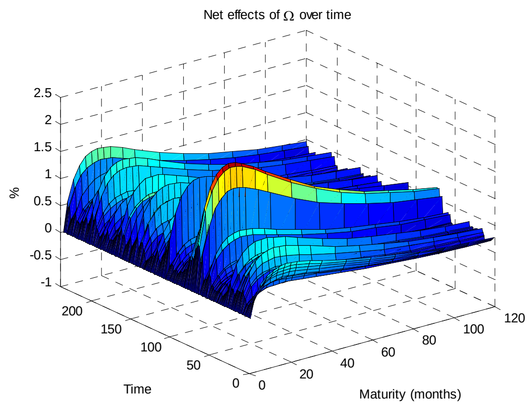


Figure 15

(a) Empirical Nelson-Siegel curvature factor (1970:01 - 2000:12)



(b) A^{+1}(1)



(c) A^{+3}(2)