Can Interest Rate Volatility be Extracted from the Yield Curve? A Test Of Unspanned Stochastic Volatility

Pierre Collin-Dufresne, Robert Goldstein and Chris S. Jones

Motivation

- Empirical Evidence from Yield data in favor of Multifactor Term Structure Models.
  – Litterman and Scheinkman (91), Knez et al. (94), Dai and Singleton (DS 00)...

- Empirical Evidence that interest rate volatility is stochastic.
  – Brenner, Harjes and Kroner (95), Anderson and Lund (97), Benzoni et al. (2003)...

- Mounting evidence suggests that term structure factors are not sufficient to explain the dynamics of fixed income derivatives.
  – Longstaff et al. (00), Brandt and Yaron (00), Jagannathan, Kaplin and Sun (99), Fan, Gupta and Ritchken (01), Collin-Dufresne and Goldstein (02), Heiddari and Wu (02)

- Collin-Dufresne and Goldstein (02) propose 'Unspanned Stochastic Volatility (USV)'
  – Bond markets alone are incomplete:
  – Other fixed income derivatives needed to hedge volatility-risk

What is 'Unspanned Stochastic Volatility'?

- Numerous Multifactor Stochastic Volatility Models of the TS.
  - Fong and Vasicek (91), Longstaff and Schwartz (92), Chen and Scott (93), Balduzzi et al. (96), Chen (96), DS (00)...

- All of these models fall within 'Affine Class'

⇒ In general, stochastic volatility can be reinterpreted as changes in yields
  - (Duffie and Kan (DK 96))

⇒ If enough bonds with different maturities are traded, markets are complete.

⇒ All Fixed Income Derivatives can be hedged with bonds alone.

Example: Term Structure Stochastic Volatility Model

- Term structure Derivatives: Longstaff and Schwartz (1992):

$$\begin{aligned} dr &= \kappa_r(\theta_r - r)\,dt + \sqrt{V}\,dz_1^{\mathcal{Q}} \\ dV &= \kappa_V(\theta_V - V)\,dt + \sigma\sqrt{V}dz_2^{\mathcal{Q}} \end{aligned}$$

Note: $P^T(t) = P^T(t, r_t, V_t) = e^{A(T-t)+B(T-t)\,r_t+C(T-t)V_t}$

   – Longstaff and Schwartz (1992)
   – More generally, Duffie and Kan (1996)

⇒ Volatility risk <u>can</u> be hedged with appropriate position in any two bonds.
⇒ Volatility plays dual role (cross-section and time series):
   1) It is a linear combination of yields
   2) It is the quadratic variation of the spot rate

USV breaks this dual role

Contrast: Equity Stochastic Volatility Model

- Equity Derivatives: Heston (1993):

$$\begin{aligned} \frac{dS}{S} &= r\,dt + \sqrt{V}\,dz_1^{\mathcal{Q}} \\ dV &= \kappa(\theta - V)\,dt + \sigma\sqrt{V}dz_2^{\mathcal{Q}} \end{aligned}$$

  – $dz_1^{\mathcal{Q}}$ drives innovations in stock price

  – $dz_2^{\mathcal{Q}}$ drives innovations in volatility

$\Rightarrow$ Volatility risk cannot be hedged by any portfolio of stock and bond

Difference: Have modeled traded asset, and its volatility.

USV in an Heath-Jarrow and Morton Framework

- Simple to model USV in an HJM framework:

$$\begin{aligned} df^T(t) &= \mu_T(t)\, dt + a^T(t, \Sigma_t)\, dz_1^Q(t) \\ d\Sigma_t &= \mu_\Sigma\, dt + \sigma_\Sigma dz_2^Q(t)\,, \end{aligned}$$

By definition $P^T(t) = e^{-\int_t^T f^s(t) ds}$. Applying Itô-Doeblin formula:

$$\frac{dP^T(t)}{P^T(t)} = r_t\, dt - \left(\int_t^T a^s(t, \Sigma_t) ds\right) dz_1^Q(t)$$

⇒ **Any** HJM model for which $dz_1 dz_2 \neq dt$ is a USV model.

- USV simple to obtain by directly modeling traded asset prices.

Andreasen et al.(97), De Jong and Santa-Clara (99), Kimmel (02), Collin-Dufresne and Goldstein (02).

- However, the model is time-inhomogeneous and takes the term structure as given:
  ⇒ has no cross-sectional implications for bond prices.

USV in a Standard Time-homogeneous Affine Model

- CDG (2002) give necessary and sufficient conditions for USV in an affine model:
  - Two-dimensional Markov models cannot display USV (and be arbitrage-free).
  - Three dimensional (non-Gaussian) affine models (with two or three factors) can display USV if parameters satisfy certain restrictions.

- Are these restrictions reasonable?

  ⇒ USV model is likely to fit derivatives better in-sample, because some volatility parameters can only be estimated from derivatives data.

  ⇒ Do restrictions significantly reduce the ability of the model at capturing term structure dynamics and cross section?

- Since affine USV is nested within an unrestricted affine model, why impose it?

  ⇒ 'Likelihood-ratio' type test?

  ⇒ Will see that 'standard approach' does not nest USV

<u>Outline of Talk</u>

1. A Canonical Representation for Affine Models

2. The Maximal $A_1(3)$ and $A_1(4)$ USV Model

3. Estimation of the Restricted and Unrestricted Model

4. Empirical Results

Affine Term Structure Models (DK 1996)

Introduce 'Latent' set of N state variables $\{X\}$ with dynamics:

$$dX(t) = \mathcal{K}^Q \left(\theta^Q - X(t)\right) \, dt + \Sigma \sqrt{S(t)} \, dW^Q(t)$$

where $\mathcal{K}^Q$, Σ are $(N \times N)$ matrices, and S diagonal matrix with components

$$S_{ii}(t) = \alpha_i + \beta_i^\top X(t) \, .$$

The spot rate is specified as an affine function of X:

$$r(t) = \delta_0 + \delta_x^\top \, X(t) \, ,$$

Due to Markov structure, all variables of interest can be written as $V\left(\{X(t)\}\right)$

Affine Term Structure Models

Note that this specification imposes that the

1) Risk-neutral drift $\mathcal{K}^Q\left(\theta^Q - X(t)\right)$

2) Covariance Matrix $\Sigma\sqrt{S(t)}\sqrt{S(t)}\Sigma^T = \Sigma\, S(t)\, \Sigma^T$

3) Spot rate $r(t) = \delta_0 + \delta_x^\top X(t)$

are all affine (i.e., linear plus constant) functions of $\{X\}$:

These strong restrictions are imposed in order to obtain tractability.
Whether or not these models can capture empirical observation is an open question.

For the most part, when the affine model has been found to fail in some respect, there has been a proposal to modify/generalize the framework that improves empirical fit while maintaining tractability.

Tractability of Affine framework:

Many fixed income securities (e.g., caps, swaptions) obtain tractable solutions.
In particular, bond prices take a simple exponential affine structure: ($\tau \equiv (T - t)$):

$$P(t,\tau) = e^{A(\tau) - B(\tau)^\top X(t)},$$

where $A(\tau)$ and $B(\tau)$ satisfy the (deterministic) ODE's (and initial conditions):

$$\frac{dA(\tau)}{d\tau} = -\theta^{Q\top}\mathcal{K}^{Q\top} B(\tau) + \tfrac{1}{2}\Sigma_{i=1}^{N}\left[\Sigma^\top B(\tau)\right]_i^2 \alpha_i - \delta_0 : \qquad A(0) = 0$$

$$\frac{dB(\tau)}{d\tau} = -\mathcal{K}^{Q\top} B(\tau) - \tfrac{1}{2}\Sigma_{i=1}^{N}\left[\Sigma^\top B(\tau)\right]_i^2 \beta_i + \delta_x : \qquad B(0) = 0$$

which can be quickly estimated for all maturities τ on a computer.

Tractability of Affine framework:

Note: Bond yields $Y(t,\tau)$ via $Y(t,\tau) = -\left(\frac{1}{\tau}\right) \operatorname{Log}\left(P(t,\tau)\right)$ linear in $\{X\}$

$$Y(t,\tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)^{\top}}{\tau} X(t).$$

2 important points:

1) Empirically, by observing yields $\{Y\}$, can 'back out' the latent variables $\{X\}$ if parameter vector $\{\Theta\}$ (and hence $A(\tau)$ and $B(\tau)^{\top}$) is known.

⇒ Empirically, usually assume N yields are measured without error
⇒ Below, we demonstrate this assumption is strongly rejected empirically

2) Theoretically, seems to suggest that one can 'rotate' state vector (and its dynamics) from the latent variables $\{X\}$ to the observed yields $\{Y\}$.

⇒ In fact, cannot do so in general and maintain tractability.
⇒ Due to $A(\tau)$ and $B(\tau)^{\top}$ not known in closed-form

Problems with Latent Variable specification:

Q? Why would we want to rotate from latent variables to observables?

1) Latent variables identified only after parameter vector is known
Ex: 1 factor model:

$$r(t) = \delta_0 + \delta_x X(t) \qquad \Rightarrow \qquad X(t) = \left(\frac{1}{\delta_x}\right)(r(t) - \delta_0)$$

$r(t)$ (= yield at zero maturity) identified independent of parameters,
but definition of $X(t)$ changes every time a new parameter vector is tested.

$\Rightarrow$ Difficult to come up with a reasonable first guess for $\{\Theta\}$
$\Rightarrow$ Increase the number of local maxima?

2) Difficult to identify whether model is 'maximal' (Dai and Singleton (2000))

$\Rightarrow$ Models which appear to be well-specified in fact contain some parameters which are not identifiable.

<u>Rotation to Observable State Vector:</u>

Q? Is there a tractable way to rotate from latent variables to observables?

A! Yes! But rather than yields of finite maturity, need Taylor series expansion of yields around zero maturity:

$$
\begin{aligned}
Y(t,\tau) &= Y(t,0) + \tau\, \partial_{\tau=0} Y(t,\tau) + \left(\frac{\tau^2}{2!}\right) \partial^2_{\tau=0} Y(t,\tau) + \ldots \\
A(\tau) &= A(0) + \tau\, \partial_{\tau=0} A(\tau) + \left(\frac{\tau^2}{2!}\right) \partial^2_{\tau=0} A(\tau) + \ldots \\
B(\tau) &= B(0) + \tau\, \partial_{\tau=0} B(\tau) + \left(\frac{\tau^2}{2!}\right) \partial^2_{\tau=0} B(\tau) + \ldots .
\end{aligned}
$$

Q? Why does this lead to tractability?

A! Because the ODE's that define $A(\tau)$ and $B(\tau)$ are effectively expansions about $\tau = 0$

Rotation to Observable State Vector:

Using $A(0) = B(0) = 0$ and collecting terms of the same order τ, we find:

$$
\begin{aligned}
Y^n(t) &\equiv \partial^n_{\tau=0} Y(t,\tau) \\
&= \frac{1}{n+1}\left(-\partial^{n+1}_{\tau=0} A(\tau) + \sum_{i=1}^{N} \partial^{n+1}_{\tau=0} B_i(\tau) X_i(t)\right) \quad \forall n = 0, 1, 2 \ldots
\end{aligned}
$$

The first few Taylor series components have simple interpretations:

$$
\begin{aligned}
Y^0(t) &= r(t) \\
Y^1(t) &= (1/2)\, \mu^Q(t) \\
Y^2(t) &= (1/3)\, \left(E^Q_t[d\mu^Q(t)] - V(t)\, dt\right)
\end{aligned}
$$

where $\mu^Q = \mathrm{E}[dr(t)]/dt$ and $V(t) = (dr(t))^2/dt$.

Rotation to Observable State Vector:

We call a representation of the state vector **canonical** if it is written in terms of:

1) Taylor series components of the term structure at zero (i.e., the Y^n),
2) The quadratic co-variations of the Y^n (i.e., $V^{i,j} = dY^i dY^j/dt$).

Advantages:

1) Factors are intuitive (level, slope, curvature, spot rate volatility, etc.)

2) Theoretical observability of factors guarantees that model is 'maximal'
   – all parameters are identifiable from fixed income securities

3) Model-insensitive estimates of state vector readily available
   – State vector empirically 'observable'
   – Can use, eg., OLS to estimate 'first guess' at parameter vector

4) Model remains affine and tractable

5) Identifies which parameters identifiable from bond prices alone under 'USV'

Example of Identification: 2-factor Gaussian Model

Ex: The 'maximum' 2-factor Gaussian model under the Q measure

$$\begin{aligned} dr_t &= (\alpha_r + \beta_{rr}\, r_t + \beta_{rx}\, x_t)dt + \sigma_r\, dZ^Q_{r,t} \\ dx_t &= (\alpha_x + \beta_{xr}\, r_t + \beta_{xx}\, x_t)dt + \sigma_x\, dZ^Q_{x,t} \end{aligned}$$

'Canonical representation' obtained by rotating from (r_t, x_t) to (r_t, μ^Q_t), where

$$\mu^Q = \alpha_r + \beta_{rr}\, r_t + \beta_{rx}\, x_t$$

Therefore

$$\begin{aligned} dr_t &= \mu^Q dt + \sigma_r dZ^Q_{r,t} \\ d\mu^Q_t &= (\gamma_\mu - \kappa_{\mu r}\, r_t - \kappa_{\mu\mu}\, \mu^Q_t)dt + \sigma_\mu\, dZ^Q_{\mu,t} \end{aligned}$$

- Only 6 parameters in canonical representation (instead of 9).
- μ^Q_t is observable (it is twice the slope of the term structure at zero).

Model-Insensitive Estimation of the State Variables

- Simulate a two factor $A_2(2)$ model (parameters from Duffie and Singleton (1997)).
- Sample 10 years of weekly data maturities $\{0.5, 1, 2, 5, 7, 10\}$ years.
- Add i.i.d. noise with either 2bp or 5bp standard errors.
- Estimate the level ($Y^0 = r$) and slope ($Y^1 = \mu^Q$) of the term structure (at $\tau = 0$) using quadratic and cubic polynomials fitted with OLS.
- Regress the estimates obtained from the polynomial fits on the true value of the simulation: (See Table 1)

$$\text{true } r_t = \alpha + \beta \times \text{estimated } r_t + \epsilon_t$$

$$\text{true } \mu_t^Q = \alpha' + \beta' \times \text{estimated } \mu_t^Q + \epsilon_t',$$

Canonical Representation and Maximality: the $A_1(3)$ model

- Consider the 3-factor model of short rate $r = Y_1 + Y_2 + Y_3$ with:

$$
\begin{aligned}
dY_1 &= \kappa_{11}(\theta_1 - Y_1)dt + \sigma_{11}\sqrt{Y_1}dZ_1 \\
dY_2 &= (\kappa_{21}Y_1 + \kappa_{22}Y_2 + \kappa_{23}Y_3)dt + \sigma_{21}\sqrt{Y_1}dZ_1 + \sigma_{22}\sqrt{\alpha_2 + \beta_2 Y_1}dZ_2 + \sigma_{23}\sqrt{\alpha_3 + \beta_3 Y_1}dZ_3 \\
dY_3 &= (\kappa_{31}Y_1 + \kappa_{32}Y_2 + \kappa_{33}Y_3)dt + \sigma_{31}\sqrt{Y_1}dZ_1 + \sigma_{32}\sqrt{\alpha_2 + \beta_2 Y_1}dZ_2 + \sigma_{33}\sqrt{\alpha_3 + \beta_3 Y_1}dZ_3
\end{aligned}
$$

- Rotating to (r, μ^Q, V) we obtain:

$$
\begin{aligned}
dV_t &= (\gamma_V - \kappa_V V_t)dt + \sigma_V\sqrt{V_t - \psi_1}\, dZ_1 \\
dr_t &= \mu_t^Q dt + \sigma_1\sqrt{V_t - \psi_1}\, dZ_1 + \sqrt{\sigma_2^2 V_t - \psi_2}\, dZ_2 + \sqrt{\sigma_3^2 V_t - \psi_3}\, dZ_3 \\
d\mu_t^Q &= (m_0 + m_r r_t + m_\mu \mu_t^Q + m_V V_t)dt + \nu_1\sqrt{V_t - \psi_1}\, dZ_1 + \nu_2\sqrt{\sigma_2^2 V_t - \psi_2}\, dZ_2 + \nu_3\sqrt{\sigma_3^2 V_t - \psi_3}\, dZ_3,
\end{aligned}
$$

where by definition of V_t:

$$
\begin{aligned}
\sigma_1^2 + \sigma_2^2 + \sigma_3^2 &= 1 \\
\sigma_1^2\psi_1 + \psi_2 + \psi_3 &= 0.
\end{aligned}
$$

$\Rightarrow$ Maximal model has 14 parameters instead of 19 (Confirms DS (2000)).

Maximal $A_1(3)$ model with USV

4 Necessary and sufficient conditions for USV in $A_1(3)$ case (CDG (2002)):

$$\begin{aligned} m_r &= -2c_V^2 & \quad m_\mu &= 3c_V \\ m_V &= 1 & \quad \sigma_V^\mu &= (c_V)^2 \end{aligned}$$

where

$$\begin{aligned} dr\, d\mu^Q &= (c_0 + c_V V)\, dt \\ (d\mu^Q)^2 &= (\sigma_0^\mu + \sigma_V^\mu V)\, dt \end{aligned}$$

The maximal model with USV is:

$$\begin{aligned} dV_t &= (\gamma_V - \kappa_V V_t)dt + \sigma_V \sqrt{V_t - \psi_1}\, dZ_1 \\ dr_t &= \mu_t dt + \sigma_1 \sqrt{V_t - \psi_1}\, dZ_1 + \sqrt{(1-\sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_2}\, dZ_3 + \sqrt{-\psi_2}\, dZ_2 \\ d\mu_t &= (m_0 - 2c_V^2 r_t + 3c_V \mu_t + V_t)\, dt + c_V \sigma_1 \sqrt{V_t - \psi_1}\, dZ_1 + c_V \sqrt{(1-\sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_2}\, dZ_3 + \nu_2 \sqrt{-\psi_2}\, dZ_2 \end{aligned}$$

where

for stationarity: $\kappa_v > 0\ ,\quad c_V < 0$

for admissibility: $\gamma_v - \kappa_v \psi_1 > 0\ ,\quad -\psi_2 > 0\ ,\quad 1 > \sigma_1^2\ ,\quad \psi_1 + \psi_2 > 0$

USV imposes 5 restrictions (model has 9 parameters under the Q measure):

$$\gamma_v,\ \kappa_v,\ \sigma_v,\ \psi_1,\ \sigma_1,\ c_V,\ \psi_2,\ m_0,\ \nu_2$$

Bond Pricing in the $A_1(3)$ USV model

$$P(t,T) = \exp\left(A(T-t) - B_r(T-t)\, r_t - B_\mu(T-t)\ \mu_t^Q\right)$$

where:

$$\begin{aligned}
B_r(\tau) &= \frac{-3 + 4e^{c_V \tau} - e^{2c_V \tau}}{2c_V} \\
B_\mu(\tau) &= \frac{(1 - e^{c_V \tau})^2}{2c_V^2} \\
A(\tau) &= \frac{1}{96 c_V^5}\Big(-3e^{4c_V\tau}(2c_V c_0 - \sigma_0^\mu) + 16 e^{3c_V\tau}(3c_V c_0 - \sigma_0^\mu) + 25\sigma_0^\mu \\
&\quad -48 e^{c_V\tau}(-5c_V c_0 - 2c_V^2 m_0 + \sigma_0^\mu) + 12 e^{2c_V\tau}(2c_V(-6c_0 - c_V m_0) + 3\sigma_0^\mu) \\
&\quad +6c_V(-23c_0 + 12 c_V m_0 + 2(-6c_V c_0 - 4c_V^2 m_0 + \sigma_0^\mu)\tau)\Big)
\end{aligned}$$

- Note that the USV model is a two-factor model of the cross-section of bond prices, but a three factor model of the time series of bonds.
- γ_V, κ_V are not identifiable from bond prices alone: In contrast to claim of DS, 'maximality' must be defined relative to all fixed income derivatives.
- Note: Even if volatility is an arbitrary Markov process, obtain same affine yields even though dynamics of state vector are not affine!

Maximal $A_1(4)$ model with USV

Similarly we derive the maximal 4-factor USV model which has state variables (r, μ, V, θ) where $\theta = 3Y^2$ (local curvature):

$$
\begin{aligned}
dV_t &= (\gamma_V - \kappa_V V_t)dt + \sigma_V \sqrt{V_t - \psi_1}\, dZ_1^Q(t) \\
dr_t &= \mu_t^Q\, dt + \sigma_1 \sqrt{V_t - \psi_1}\, dZ_1^Q(t) + \sqrt{(1-\sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4}\, dZ_2^Q(t) \\
&\quad + \sqrt{-\psi_3}\, dZ_3^Q(t) + \sqrt{-\psi_4}\, dZ_4^Q(t) \\
d\mu_t^Q &= (\theta_t^Q + V_t)\, dt + c_{r\mu}\sigma_1 \sqrt{V_t - \psi_1}\, dZ_1^Q(t) + c_{r\mu}\sqrt{(1-\sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4}\, dZ_2^Q(t) \\
&\quad + \nu_3 \sqrt{-\psi_3}\, dZ_3^Q(t) + \nu_4 \sqrt{-\psi_4}\, dZ_4^Q(t) \\
d\theta_t^Q &= \left(a_0 - 2c_{r\mu}^2 (3c_{r\mu} - a_\theta)\, r_t + (7c_{r\mu}^2 - 3c_{r\mu} a_\theta)\, \mu_t^Q + a_\theta\, \theta^Q + 3c_{r\mu}\, V_t\right) dt \\
&\quad + c_{r\mu}^2 \sigma_1 \sqrt{V_t - \psi_1}\, dZ_1^Q(t) + c_{r\mu}^2 \sqrt{(1-\sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4}\, dZ_2^Q(t) \\
&\quad + \eta_3 \sqrt{-\psi_3}\, dZ_3^Q(t) + \eta_4 \sqrt{-\psi_4}\, dZ_4^Q(t).
\end{aligned}
$$

Note: the $A_1(4)$ USV model has a total of 14 risk-neutral parameters $(\gamma_V, \kappa_V, \sigma_V, \psi_1, \psi_3, \psi_4, \nu_3, \nu_4, \eta_3, \eta_4, \sigma_1, a_0, c_{r\mu}, a_\theta)$, as opposed to 22 for the unrestricted model.

Bond Pricing in the $A_1(4)$ USV model

The zero coupon bond price is given by:

$$P(t,T) = \exp\left(A(T-t) - B_r(T-t)\, r_t - B_\mu(T-t)\, \mu_t^Q - B_\theta(T-t)\, \theta_t^Q\right),$$

where the deterministic functions $A(\tau)$, $B_r(\tau)$, $B_\mu(\tau)$, and $B_\theta(\tau)$ are obtained in closed form.

- Note that the $A_1(4)$ USV model is a 3-factor Gaussian model of the cross-section of bond prices, but a 4-factor model of the time series of bonds.

Specification of Risk-Premia

Specify risk-premia process so that the dynamics of the state vector for the unrestricted $A_1(3)$ under the historical measure are:

$$dV_t = \left(\left(\gamma_V + \lambda_{V0} - \lambda_V \psi_1\right) - \left(\kappa_V - \lambda_V\right) V_t \right) dt + \sigma_V \sqrt{V_t - \psi_1}\, dZ_1(t)$$

$$\begin{aligned} dr_t &= \left(\lambda_{r0} + \lambda_{rr}\, r_t + \left(1 + \lambda_{r\mu}\right) \mu_t^Q + \lambda_{rV} V_t\right) dt \\ &+\sigma_1 \sqrt{V_t - \psi_1}\, dZ_1(t) + \sqrt{\sigma_2^2 V_t - \psi_2}\, dZ_2(t) + \sqrt{\sigma_3^2 V_t - \psi_3}\, dZ_3(t) \end{aligned}$$

$$\begin{aligned} d\mu_t^Q &= \left(\left(m_0 + \lambda_{\mu 0}\right) + \left(m_r + \lambda_{\mu r}\right) r_t + \left(m_\mu + \lambda_{\mu\mu}\right) \mu_t^Q + \left(m_V + \lambda_{\mu V}\right) V_t \right) dt \\ &+\nu_1 \sqrt{V_t - \psi_1}\, dZ_1(t) + \nu_2 \sqrt{\sigma_2^2 V_t - \psi_2}\, dZ_2(t) + \nu_3 \sqrt{\sigma_3^2 V_t - \psi_3}\, dZ_3(t) \end{aligned}$$

- All drift parameters in r_t, μ_t^Q, V dynamics are risk-adjusted.

- Extends slightly Duffee, but need to guarantee that Feller condition holds under both the physical and risk neutral measures for existence of EMM (Liptser and Shiryaev).

Empirical Methodology

- Use weekly swap rate data (maturities $\{2, 3, 4, 5, 7, 10\}$ and six month LIBOR from Jan. 7, 1988 to Nov. 27, 2002.
- Adjust LIBOR quote for non-synchronicity with USD swap rates.
- Estimate unrestricted model using Quasi Maximum Likelihood (QML), similar to Chen and Scott (1993), Pearson and Sun (1994).
  - Standard procedure fits 3 specific yields perfectly to invert for state variables.
  - Remaining yields observed with 'measurement' errors.
  - Log-Likelihood is a combination of transition density of state variables and (Gaussian) likelihood for the errors.
  - When transition density is not known explicitly use a Gaussian (QML) approximation based on the exact first two moments which can be computed explicitly (Fisher and Gilles (1996), Duffee (2002))

‘Improvements’ to QML estimation

- Use **principal components** instead of yields to invert for state and ‘errors’
  - Guarantees to fit perfectly first three PC which explain over 95% of the variance of yields (Litterman and Scheinkman (1991).
  - ‘Orthogonalizes’ the (unconditional) matrix of measurement errors.
  - Dispenses with the arbitrariness of the yields fitted exactly.
  - Retains simplicity of inversion for the state (PC’s are linear in state variables).
- Tested cumulant expansion approximation to the transition density based on explicit higher order moments to improve estimation of transition density.

Noninvertibility of Yields

When state vector cannot be inverted from bond prices, use a simulated QML approach based on *Efficient Importance Sampler* of Richard and Zhang (1996,97) (see also Sandmann and Koopman (1998)), Pennachi (1991), Brandt and He (2002)

Ex 1): USV implies V cannot be determined from bond yields.

Ex 2): If assumed that yields are measured with errors, then state vector not invertible from yields.

Let $\mathbf{P} = \{\mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_T\}$ denote the time series of PC's of the yield curve.

Likelihood function, $\mathrm{p}(\mathbf{P}|\theta)$, may be written as the integral

$$\int \mathrm{p}(\mathbf{P}, \mathbf{V}|\theta)\, d\mathbf{V}$$

where $\mathbf{V} = \{V_1, V_2, ..., V_T\}$ denotes the time series of the variance process.

Noninvertibility of Yields

The integral is evaluated using simulation.

From 'importance sampling', an approximate auxiliary model $\mathrm{p}^a(\mathbf{V}|\mathbf{P},\theta)$ is specified

$$\int \mathrm{p}(\mathbf{P},\mathbf{V}|\theta)\,d\mathbf{V} = \int \mathrm{p}(\mathbf{P},\mathbf{V}|\theta)\frac{\mathrm{p}^a(\mathbf{V}|\mathbf{P},\theta)}{\mathrm{p}^a(\mathbf{V}|\mathbf{P},\theta)}d\mathbf{V} = \mathrm{E}^a\left[\frac{\mathrm{p}(\mathbf{P},\mathbf{V}|\theta)}{\mathrm{p}^a(\mathbf{V}|\mathbf{P},\theta)}\right],$$

The closer the auxiliary model is to the actual model, the less sensitive the ratio is to the simulated variance path, and the more quickly the expectation will converge.

The EIS approach essentially chooses the auxiliary density $\mathrm{p}^a(\mathbf{V}|\mathbf{P},\theta)$ (within a certain parametric class) to minimize the variation in

$$\ln \mathrm{p}(\mathbf{V}|\mathbf{P},\theta) - \ln \mathrm{p}^a(\mathbf{V}|\mathbf{P},\theta)$$

Test Five Specifications

1. Unrestricted $A_1(3)$ model
   - assume 2 PCs are observed without error — "2PC"
   - simulate paths of V
   - "invert" V and the 2 PCs for r, μ^Q, and V
2. Unrestricted $A_1(3)$ model
   - assume 3 PCs are observed without error — "3PC"
   - "invert" PCs for r, μ^Q, and V
3. $A_1(3)$ model with USV restrictions
   - assume 2 PCs are observed without error — "USV"
   - "invert" PCs for r and μ^Q
   - simulate paths of V
4. Unrestricted $A_1(2)$ model
   - assume 2 PCs are observed without error — "$A_1(2)$"
   - "invert" PCs for r and V (there is no μ^Q)
5. Unrestricted $A_1(4)$ model
   - assume 3 PCs are observed without error — "$A_1(4)$USV"
   - simulate paths of V
   - "invert" V and the 3 PCs for r, μ^Q, θ, and V

– Models 2, 3, and 4 are restricted versions of 1 (Table 4).

– Model 5 nests Model 3.

Empirical Results

- Ignoring QML approximation error, all three-factor restricted models rejected by LR test.
- All models predict short rate and slope accurately (Table 7).
- Only $A_1(3)$ and $A_1(4)$USV capture dynamics of Curvature (Figure 1, Table 7).
- But $A_1(3)$ (unrestricted and 3PC) predict volatility that is negatively correlated with model-independent GARCH volatility, as well as with volatility extracted from short rate implied by the model itself! (Figure 2, Table 7)

$\Rightarrow$ $A_1(3)$ cannot both capture curvature and dynamics of short rate volatility (V plays double role in unrestricted $A_1(3)$ model).

- $A_1(3)$ USV does a better job at capturing volatility dynamics, but not quite as good for curvature (Table 7, Figure 2).
- Only $A_1(4)$ USV can capture both time series property of short rate volatility and dynamics of TS Curvature factor.
- Superiority of $A_1(4)$ model confirmed by out of sample yield changes (table 9) and squared yield changes (table 10), as well as predictability regression (Figure 3) and Maturity/volatility relation (Table 4).

Conclusion

- Propose a canonical representation for affine models in which:
  - the state variables have simple physical interpretations such as level, slope and curvature at the short end,
  - their dynamics remain affine and tractable,
  - the model is by construction 'maximal' in the sense of Dai and Singleton (00),
  - model-insensitive estimates of the state variables are readily available.
- Offer a complete characterization of the 'maximal' $A_1(3)$ and $A_1(4)$ USV model.
- Empirical estimation of the various models show:
  - USV restrictions do not significantly affect cross-sectional fit, but
  - Substantially improve the time-series properties of the model (in and out of sample forecasts of squared changes in yields).
  - Even though USV is nested within the unrestricted model, imposing the USV restriction explicitly improves the estimated time series of volatility, because USV breaks the dual role played by volatility in the unrestricted model.
  - To capture dynamics of level, slope, and curvature, as well as stochastic short rate volatility need four distinct factors ($A_1(4)$).