A Multifrequency Theory of the Interest Rate Term Structure

Liuren Wu at Baruch College

Joint work with Laurent Calvet and Adlai Fisher

Ziff Brothers Investments, April 2nd, 2010
Shocks to the interest rate term structure

- Shocks of all frequencies come at the interest rate dynamics/term structure:
  - **Long term**: *Inflation* shocks tend to move the term structure in parallel; *Real GDP growth* shocks tend to move short rates more than long rates.
  - **Intermediate term**: *Monetary policy* shocks are often imposed at the short end and they dissipate through the yield curve via expectations.
  - **Short term**: *Supply/demand* (transactions) shocks enter the yield curve at a particular maturity and dissipate through the yield curve via hedging and yield curve statistical arbitrage trading.

- A successful term structure model must capture the effects of shocks of all frequencies.
The literature

- **Theory:** Dynamic term structure models with $N$ factors are well-developed, with analytical tractability. Examples include the affine class (Duffie, Kan, Pan, and Singleton) and the quadratic class (Leippold and Wu).

- **Practice:** The commonly estimated models are all low-dimensional, mostly with three factors.
  - Three-factor models are successful in capturing major variations in the interest rate level, the term structure slope, and curvature.
  - The remaining movements can be economically significant (in four-leg trades, Bali, Heidari, & Wu).
  - Three-factor models fail miserably in
    - predicting future interest rate movements (Duffee),
    - capturing the cross-correlation between non-overlapping forwards (Dai & Singleton),
    - pricing interest-rate options (Heidari & Wu, Li & Zhao).
Why not estimate a high-dimensional model?

- **Curse of dimensionality:**
  - A generic affine three-factor model has 20-30 parameters (more for quadratic models).
  - The number of parameters increases quadratically with dimensionality.
  - Many of these parameters cannot be effectively identified.
  - These models suffer from the “double whammy” of being
    - too little — It cannot match all the features of the data.
    - too much — It has too many parameters to be effectively identified.

- We propose a class of models that can include all identifiable frequencies, but no curse of dimensionality — The models are *dimension invariant*.
  - Parameter identification is not an issue.
  - Dimension is a choice (number and distribution of frequencies), but not a concern.
The instantaneous interest rate $r_t$ follows a \textit{cascade} dynamics,

\begin{align}
    r_t &= x_{n,t}, \\
    dx_{j,t} &= \kappa_j (x_{j-1,t} - x_{j,t}) dt + \sigma_{j,t} dW_{j,t}, \quad j = n, n-1, \cdots, 1, \\
    x_{0,t} &= \theta_r. \\
\end{align}

- Start the short rate at the highest identifiable frequency $x_{n,t}$.
- Let the short rate mean reverts to a stochastic tendency $x_{n-1,t}$.
  — By design, the tendency $x_{n-1,t}$ moves slower than $x_{n,t}$.
- The tendency mean reverts to another, even slower tendency ...
- The lowest frequency reverts to a constant mean $\theta_r$, which is also the mean of the short rate.
- Intuitively, the tendencies are like exponentially weighted moving averages with increasingly long windows.
- $n \rightarrow \infty$ is also an option.
The general affine Gaussian models (Duffie & Kan, 96):
\[ r_t = a + b^\top X_t, \quad dX_t = K(c - X_t)dt + \Sigma dW. \]

Factors can rotate. For example, equivalently,
\[ r_t = a' + (b')^\top Z_t, \quad dZ_t = -K'Z_t dt + dW, \]
with
\[ a' = a + b^\top c, \quad b' = \Sigma b, \quad c' = \Sigma^{-1}c, \quad K' = \Sigma^{-1}K. \]

- Economic meaning for each factor is elusive.
- Many of the parameters are not identifiable.
  — Need careful specification analysis (Dai & Singleton, 2000).

The *cascade* structure ranks the factors according to frequency.
— a natural separation/filtration of the different frequency components in the interest rate movements — no more rotation.

Economic meaning of each factor becomes clearer — helpful for designing models to match data.
— 1/\( \kappa \) has the unit of time.

- From time series, the highest identifiable frequency is the observation frequency. The lowest frequency is the sample length.
- From term structure, maturity range determines frequency range.
We achieve **dimension invariance** by parameterizing the distribution of the different frequencies.

We assume that the mean reversion speeds of different frequencies scale via a **power law**: \( \kappa_j = \kappa_r b^{(j-1)}, \quad b > 1 \).

- Using a functional form to approximate a series of discrete coefficients is a common trick used in econometrics to improve identification.
- Example: **Geometric distributed lags model** assumes that the effects of a variable \( x_t \) diminishes as the lag \( j \) becomes larger: \( \beta_j = \beta_0 \lambda^j, \quad \lambda < 0 \).

Approximate power laws are popular in physics, and are also observed in financial data (Mandelbrot, Calvet & Fisher, Gabaix).

By fixing the mean reversion at the lowest frequency \( \kappa_r \), and progressively adding more higher frequency components, we are essentially increasing sampling frequency while fixing the sample length.

One can also start with the highest frequency, and progressively add more lower frequency components — increasing the sample length while fixing sampling frequency.
The focus of this paper is on the modeling of the interest rate term structure.

Volatility variation is largely “unspanned” by the term structure. Hence, its specification does not matter much for term structure modeling.

We assume IID risks with constant volatility: $\sigma_{j,t} = \sigma_r$.

For option pricing, variance dynamics specification is important. A potentially useful dimension-invariant cascade specification for variance (for future work),

\[
\begin{align*}
\sigma_{j,t}^2 & = v_{K,t} \\
v_{k,t} & = \kappa_k^v (v_{k-1,t} - v_k,t) \, dt + \omega \sqrt{v_k,t} dZ_k,t, \quad k = K, K - 1, \ldots, 1, \\
v_{0,t} & = \theta_v, \\
\rho & = \mathbb{E}[dW_j,t dZ_k,t]/dt, \quad \kappa_k^v = \beta^{k-1} \kappa_1^v, \quad \beta > 1.
\end{align*}
\]
Dimension invariance assumption on risk premium

- Risk premium is not important for term structure modeling, either.
  - Duffie & Kan only define the “affine” dynamics under $\mathbb{Q}$. $\mathbb{P}$-dynamics (and hence risk premium) can be anything (pretty much) ...

- We assume constant and identical market prices for risks of all frequencies: $\gamma_{j,t} = \gamma_r$.

- For “bond risk premia” (e.g., Cochrane, Piazzesi) and “expectation hypotheses”, risk premium is all they care about.
  - The different frequency components that we identify can be used as the instruments to explain the bond risk premium (in a second step).
    - What is a “tent shape” in our frequency decomposition?
    - Do the volatility frequency components identified from the “vol cube” have anything to say about bond risk premium?
  - Excess Bond Return $t+\Delta t = a + \sum_j b_j x_{j,t} + \sum_j c_k v_{k,t} + e_{t+\Delta t}$
    - Parameterize the distribution of $(b_j, c_k)$ to achieve dimension-invariance and enhance identification.

- Bottom line for term structure modeling: Five parameters $(\theta_r, \sigma_r, \kappa_r, b, \gamma_r)$, regardless of the number of frequencies ($n$).
The values of zero-coupon bonds are exponential-affine in \( X_t = \{x_{j,t}\}_{j=1}^n \),

\[
P(X_t, \tau) = \mathbb{E}_t^P \left[ \exp \left( - \int_t^T r_s ds \right) \mathcal{E} \left( - \int_t^T \gamma_s \cdot dX_s \right) \right] = e^{-b(\tau)^\top X_t - c(\tau)},
\]

The instantaneous forward rate is affine in the state vector,

\[
f(X_t, \tau) = a(\tau)^\top X_t + e(\tau),
\]

The intercept has 3 components: long-run mean, risk premium, convexity: \( e(\tau) = \)

\[
\left\{ \begin{array}{l}
\kappa_r \theta_r \sum_{i=1}^n \alpha_{i,1} (1 - e^{-\kappa_i \tau}) \\
-\gamma_r \sigma_r^2 \sum_{j=1}^n \sum_{i=j}^n \alpha_{i,j} (1 - e^{-\kappa_i \tau}) \\
-\frac{\sigma_r^2}{2} \sum_{j=1}^n \sum_{i=j}^n \alpha_{i,j} \alpha_{k,j} (1 - e^{-\kappa_k \tau} - e^{-\kappa_i \tau} + e^{-(\kappa_i + \kappa_k) \tau})
\end{array} \right.
\]

The loading coefficient \( a(\tau) \) are convolutions of exponentials.
The forward rate response functions to shocks from different frequency components, $a_j(\tau)$:

- Highest frequency
- Intermediate
- Lowest frequency

Numerical example: $\kappa_r = 1/30$, $\kappa_n = 52$, $n = 15$, $b = 1.69$. 
Data

- Six LIBOR (at 1, 2, 3, 6, 12 months),
- Nine swap rates (at 2, 3, 4, 5, 7, 10, 15, 20, 30 years).
- Weekly sampled (Wednesday) from January 4, 1995 to December 26, 2007. 678 observations for each series. All together 10,170 observations.
Estimation

- Cast the model into a state space form:
  - Regard $X_t$ as the hidden state, regard the LIBOR and swap rates as observations with errors.

- Given parameters, use unscented Kalman filter to infer the states $X_t$ from the observations at each date.

- Construct the log likelihood by assuming that the forecasting errors on LIBOR and swap rates are normally distributed.

- Estimate the 5 parameters by maximizing the likelihood of forecasting errors.
Normally, this is the first thing one decides on before one can pin down the parameter space.

Under our model, the parameter space is invariant to the dimensionality decision. We worry about the dimensionality the last.

Since we have 15 interest rate series, we estimate 15 models with \( n = 1, 2, 3, \ldots, 15 \).

The estimations of these models are equally easy and fast.

The extensive estimation exercise serves at least two purposes:

- Determine how many frequency components the data ask for — This normally depends on the data. More maturities would naturally ask for more frequency components.
- Analyze how high-dimensional models differ from low-dimensional models in performance.
### Parameter estimates and likelihood ratio tests

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\kappa_r$</th>
<th>$\theta_r$</th>
<th>$\sigma_r$</th>
<th>$\theta_r^Q$</th>
<th>$b$</th>
<th>$\sigma_e^2$</th>
<th>$\mathcal{L}$</th>
<th>$\mathcal{V}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2092</td>
<td>0.0436</td>
<td>0.0065</td>
<td>0.0688</td>
<td>0.0000</td>
<td>0.1574</td>
<td>4086</td>
<td>47.91</td>
</tr>
<tr>
<td>3</td>
<td>0.0526</td>
<td>0.0000</td>
<td>0.0101</td>
<td>0.0662</td>
<td>7.3138</td>
<td>0.0047</td>
<td>19928</td>
<td>20.70</td>
</tr>
<tr>
<td>5</td>
<td>0.0441</td>
<td>0.0000</td>
<td>0.0125</td>
<td>0.0507</td>
<td>2.8266</td>
<td>0.0010</td>
<td>25551</td>
<td>15.99</td>
</tr>
<tr>
<td>7</td>
<td>0.0283</td>
<td>0.0000</td>
<td>0.0129</td>
<td>0.0419</td>
<td>2.6150</td>
<td>0.0004</td>
<td>27898</td>
<td>11.93</td>
</tr>
<tr>
<td>8</td>
<td>0.0275</td>
<td>0.0000</td>
<td>0.0133</td>
<td>0.0632</td>
<td>2.5271</td>
<td>0.0004</td>
<td>28445</td>
<td>11.00</td>
</tr>
<tr>
<td>9</td>
<td>0.0278</td>
<td>0.0000</td>
<td>0.0141</td>
<td>0.0650</td>
<td>2.2351</td>
<td>0.0003</td>
<td>28801</td>
<td>9.18</td>
</tr>
<tr>
<td>10</td>
<td>0.0313</td>
<td>0.0000</td>
<td>0.0140</td>
<td>0.0507</td>
<td>2.2010</td>
<td>0.0003</td>
<td>28972</td>
<td>6.68</td>
</tr>
<tr>
<td>11</td>
<td>0.0305</td>
<td>0.0000</td>
<td>0.0144</td>
<td>0.0966</td>
<td>1.9603</td>
<td>0.0003</td>
<td>29036</td>
<td>6.06</td>
</tr>
<tr>
<td>12</td>
<td>0.0359</td>
<td>0.0000</td>
<td>0.0147</td>
<td>0.0876</td>
<td>1.9130</td>
<td>0.0002</td>
<td>29194</td>
<td>4.41</td>
</tr>
<tr>
<td>13</td>
<td>0.0383</td>
<td>0.0000</td>
<td>0.0149</td>
<td>0.0833</td>
<td>1.8953</td>
<td>0.0002</td>
<td>29283</td>
<td>3.33</td>
</tr>
<tr>
<td>14</td>
<td>0.0409</td>
<td>0.0000</td>
<td>0.0151</td>
<td>0.0781</td>
<td>1.8757</td>
<td>0.0002</td>
<td>29332</td>
<td>2.32</td>
</tr>
<tr>
<td>15</td>
<td>0.0572</td>
<td>0.0000</td>
<td>0.0156</td>
<td>0.0559</td>
<td>1.7400</td>
<td>0.0002</td>
<td>29377</td>
<td>—</td>
</tr>
</tbody>
</table>

- **Vuong test** (last column): *More is significantly better.*
- Spacing ($b$) is finer when more is allowed.
- Parameters ($\kappa_r, \sigma_r$) stabilize as $n$ increases.
Circles: $\kappa_i$ as free parameters; Solid line: power-law scaling
## In-sample fitting performance: Pricing error statistics

<table>
<thead>
<tr>
<th>Maturity</th>
<th>A. Three-factor model</th>
<th>B. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Rmse</td>
</tr>
<tr>
<td>1 m</td>
<td>-0.68</td>
<td>7.47</td>
</tr>
<tr>
<td>2 m</td>
<td>0.63</td>
<td>3.82</td>
</tr>
<tr>
<td>3 m</td>
<td>1.61</td>
<td>5.03</td>
</tr>
<tr>
<td>6 m</td>
<td>0.39</td>
<td>6.78</td>
</tr>
<tr>
<td>9 m</td>
<td>-1.74</td>
<td>6.88</td>
</tr>
<tr>
<td>1 y</td>
<td>-3.06</td>
<td>6.74</td>
</tr>
<tr>
<td>2 y</td>
<td>2.11</td>
<td>6.17</td>
</tr>
<tr>
<td>3 y</td>
<td>1.97</td>
<td>6.90</td>
</tr>
<tr>
<td>4 y</td>
<td>0.87</td>
<td>6.32</td>
</tr>
<tr>
<td>5 y</td>
<td>-0.21</td>
<td>5.85</td>
</tr>
<tr>
<td>7 y</td>
<td>-1.89</td>
<td>5.55</td>
</tr>
<tr>
<td>10 y</td>
<td>-2.35</td>
<td>5.17</td>
</tr>
<tr>
<td>15 y</td>
<td>0.88</td>
<td>3.87</td>
</tr>
<tr>
<td>20 y</td>
<td>1.91</td>
<td>5.35</td>
</tr>
<tr>
<td>30 y</td>
<td>-0.76</td>
<td>9.67</td>
</tr>
<tr>
<td>Average</td>
<td>-0.02</td>
<td>6.11</td>
</tr>
</tbody>
</table>
Similar to Nelson-Siegel (basis function is exponentials), with two advantages:

- Dynamic consistency.
- No longer limit to a three-factor structure — Near-perfect fitting is a must for stripping swap rate curves.
### Predictive variation

\[ 1 - \frac{\text{Mean Squared Forecasting Error}}{\text{Mean Squared Interest Rate Change}} \]

<table>
<thead>
<tr>
<th>Model</th>
<th>A. AR(1)</th>
<th>B. Three-factor model</th>
<th>C. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h )</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>LIBOR maturity in months:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>25.85</td>
<td>43.84</td>
<td>57.50</td>
</tr>
<tr>
<td>2</td>
<td>23.83</td>
<td>36.65</td>
<td>47.28</td>
</tr>
<tr>
<td>3</td>
<td>22.82</td>
<td>32.19</td>
<td>41.34</td>
</tr>
<tr>
<td>6</td>
<td>20.85</td>
<td>25.00</td>
<td>31.90</td>
</tr>
<tr>
<td>9</td>
<td>20.22</td>
<td>19.35</td>
<td>23.79</td>
</tr>
<tr>
<td>12</td>
<td>21.45</td>
<td>17.53</td>
<td>20.58</td>
</tr>
<tr>
<td>2</td>
<td>-0.71</td>
<td>32.92</td>
<td>42.84</td>
</tr>
<tr>
<td>1</td>
<td>-1.94</td>
<td>15.23</td>
<td>23.31</td>
</tr>
<tr>
<td>3</td>
<td>-50.31</td>
<td>-12.95</td>
<td>1.57</td>
</tr>
<tr>
<td>6</td>
<td>-87.43</td>
<td>-42.16</td>
<td>-24.57</td>
</tr>
<tr>
<td>9</td>
<td>-67.23</td>
<td>-38.76</td>
<td>-28.15</td>
</tr>
<tr>
<td>2</td>
<td>21.71</td>
<td>40.82</td>
<td>52.16</td>
</tr>
<tr>
<td>3</td>
<td>17.65</td>
<td>28.50</td>
<td>37.00</td>
</tr>
<tr>
<td>1</td>
<td>8.78</td>
<td>21.86</td>
<td>29.17</td>
</tr>
<tr>
<td>2</td>
<td>5.77</td>
<td>12.56</td>
<td>16.94</td>
</tr>
<tr>
<td>3</td>
<td>1.30</td>
<td>4.99</td>
<td>7.06</td>
</tr>
<tr>
<td>1</td>
<td>6.85</td>
<td>3.71</td>
<td>3.07</td>
</tr>
</tbody>
</table>

**AR(1) is the best;**  
**3-factor model cannot beat random walk.**
### Out-of-sample forecasting performance

<table>
<thead>
<tr>
<th>Model</th>
<th>A. AR(1)</th>
<th>B. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
<td>Predictive variation</td>
<td>Predictive variation</td>
</tr>
<tr>
<td>$h$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>LIBOR maturity in months:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.57</td>
<td>-3.22</td>
</tr>
<tr>
<td>2</td>
<td>-1.50</td>
<td>-3.24</td>
</tr>
<tr>
<td>3</td>
<td>-1.98</td>
<td>-3.71</td>
</tr>
<tr>
<td>6</td>
<td>-3.36</td>
<td>-5.62</td>
</tr>
<tr>
<td>9</td>
<td>-4.52</td>
<td>-7.14</td>
</tr>
<tr>
<td>12</td>
<td>-4.90</td>
<td>-7.78</td>
</tr>
</tbody>
</table>

AR(1) is the worst;  
15-factor model beats random walk in sample and out of sample!
Where does the forecasting strength come from?

- AR(1) regression neither uses the term structure information nor is it parsimonious.
  - To exploit the term structure information, need a VAR(1) structure.
  - One AR(1) on each series, $15 \times 2 = 30$ parameters already!
  - Forget about a general VAR(1).

- Our model can be regarded as a constrained VAR(1):
  - Exploits information on the term structure.
  - Parsimony generates out-of-sample stability for all our models.

- ... as simple as possible, but not simpler.
  - Low-dimensional models cannot even fit — The forecast is almost surely wrong over short horizons.
    - If the fitting error is 6 bps, the forecasting error over the next second will also be 6 bps — no hope of beating random walk.

- Our high-dimensional model is:
  - simple and stable: Similar in and out of sample performance.
  - flexible and fits perfectly: The forecast starts at the right place.
We propose a class of dimension-invariant cascade multifrequency dynamic term structure models:

- The cascade factor structure eliminates factor rotation, pins down the meaning of each factor, and provides a natural separation/filtration of different frequency components.
- We achieve dimension invariance by parameterizing the frequency distribution.
  - Power law scaling on mean reversion — enjoys empirical support.
  - IID risk and risk premium — simplification for term structure modeling.

Model estimation and performance analysis reveal several advantages over traditional “general” specifications:

- No more curse of dimensionality: high-dimensional models are just as easy to be estimated/identified as low-dimensional models.
- High-dimensional models do perform better in several fronts.
- The dimension-invariant cascade multifrequency framework can be readily applied to (i) option pricing, and (ii) risk premium analysis.