A Multifrequency Theory of the Interest Rate Term Structure

Laurent Calvet, Adlai Fisher, and Liuren Wu

HEC, UBC, & Baruch College

February 3, 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 model structure with no curse of dimensionality.
  ⇒ The model *dimension invariant* — 5 parameters regardless of dimension.
  - Parameter identification is not an issue.
  - Dimension is a choice, but not a concern.
  - We can choose the dimension as high as needed to match the data.
The instantaneous interest rate $r_t$ follows a *cascade* dynamics,

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

1. Start the short rate at the highest identifiable frequency $x_{n,t}$.
2. 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 $x_{n,t}$.
3. The tendency mean reverts to another, even slower tendency ...
4. The lowest frequency reverts to a constant mean $\theta_r$.
5. *IID* risks and identical market prices: $\sigma_j = \sigma_r$, $\gamma_j = \gamma_r$.
6. The mean reversion speeds of different frequencies scale via a *power law*:
   $$
   \kappa_j = \kappa_r b^{(j-1)}, \quad b > 1.
   $$

$\Rightarrow$ The model becomes *dimension invariant*.

Five parameters ($\theta_r, \sigma_r, \kappa_r, b, \gamma_r$), regardless of the number of factors ($n$).
Comparison to the literature: Cascade v. general affine

- A subclass of the general affine Gaussian models (Duffie & Kan, 96):
  \[ r_t = a + b^\top X_t, \, dX_t = K(c - X_t)dt + \Sigma dW. \]

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

Comparison to the literature: Power law scaling

- Power-law scaling is a common phenomenon observed in many areas of natural science.
- Approximate power laws are often observed in financial data (Mandelbrot, Calvet & Fisher, Gabaix).
- Together with the iid risk/market price assumption, we use power-law scaling to achieve extreme parsimony and dimension invariant.

  - 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 an variable $x_t$ diminishes as the lag $j$ becomes larger:
    $$ \beta_j = \beta_0 \lambda^j, \quad \lambda < 0. $$
An alternative representation of the short rate dynamics

\[ r_t = \theta_r + \sum_{j=1}^{n} a_j(t)(x_{j,0} - \theta_r) + \sigma_r \sum_{j=1}^{n} \int_0^t a_j(t-s) dW_{j,s}. \]

- \(a_j(\tau)\) — the response function of the short rate to a unit shock from the \(j\)th frequency component at \(\tau\)-time ago.

- It can be solved as convolutions of exponential density functions:
  \[ a_j(\tau) = (K_j \ast \ldots \ast K_n)(\tau)/\kappa_j, \quad K_j(\tau) = \kappa_j e^{-\kappa_j \tau}, \quad \tau > 0. \]

- The short rate response to \(W_{n,t}\) starts at one and decays exponential, \(a_n(\tau) = e^{-\kappa_n \tau}\). The decay is fast with higher mean reversion.

- The response to \(W_{n-1,t}\) is hump shaped,
  \[ a_{n-1}(\tau) = \frac{\kappa_n}{\kappa_n - \kappa_{n-1}} \left( e^{-\kappa_{n-1} \tau} - e^{-\kappa_n \tau} \right). \]

  with the maximum response occurring at \(\bar{\tau}_{n-1} = \ln b/(\kappa_{n-1}(b-1))\).

- All lower frequency shocks generate hump-shaped responses, with the maxima occurring at progressively longer horizons.
Short rate response functions
to shocks from different frequency components, $a_j(\tau)$:

$W_n$ (highest)  $\rightarrow$ (intermediate)  $W_1$ (lowest)

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

Liuren Wu (Baruch)  Cascade Dynamics with Power Law Scaling  February 3, 2010  9 / 24
Unconditional variance and the limit behavior

- The unconditional variance of the short rate has an upper bound,
  \[
  \text{Var}(r_t) = \sigma_r^2 \sum_{j=1}^{n} \int_0^{\infty} a_j^2(s)ds \leq \sigma_r^2 \sum_{j=1}^{n} \int_0^{\infty} a_j(s)ds = \sigma_r^2 \sum_{j=1}^{n} \frac{1}{\kappa_j} < \infty.
  \]
  as long as \( \kappa_j > 0 \).
- Under power law scaling, the upper bound becomes,
  \[
  \text{Var}(r_t) \leq \sigma_r^2 \sum_{j=1}^{n} \frac{1}{\kappa_j} = \frac{\sigma_r^2}{\kappa_r} \frac{1 - b^{-n}}{1 - b^{-1}}.
  \]
- The short rate has finite variance regardless of how many factors we add to the dynamics.
- The limit exists as \( n \to \infty \).
- The unconditional variance contribution of higher frequency components becomes progressively smaller.
Bond pricing

- 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 short rate response function $a(\tau)$ across different time lags also determines the contemporaneous response of the forward rate curve.

$a(\tau)$ — fixed basis functions
$X_t$ — time-varying weights.
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.
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.
Dimensionality

- 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 $n = 1, 2, 3, \cdots 15$.

- 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 one uses. 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>

- More is significantly better.
- Spacing is finer when more is allowed.
- Parameters 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>Model</th>
<th>A. Three-factor model</th>
<th>B. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maturity</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>
Application: Yield curve stripping

Model-generated forward curves

Piece-wise constant assumption

Liuren Wu (Baruch)
No parameter estimation is needed for practical application of stripping.

- Use $\kappa_i$ or $(\kappa_r, b)$ to match maturities on the curve.
- Use $\theta_r$ and $\gamma_r$ to match the short and long-end of the curve.
- Fix $\sigma_r$ to historical values (effects are small on the term structure).

Fixing the parameters amounts to fixing the basis functions.

- Choose the states/weights $X_t$ to match the curve.

Similar to Nelson-Siegel, with two advantages:

- Dynamic consistency.
- No longer limit to a three-factor structure — Near-perfect fitting is a must for stripping swap rate curves.
Cross-correlation between weekly changes between six-month LIBOR and other interest rate series

Low-dimensional models tend to generate overly high cross-correlations.
In-sample forecasting performance

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></th>
<th>B. Three-factor model</th>
<th></th>
<th>C. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIBOR maturity in months:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>25.85</td>
<td>43.84</td>
<td>57.50</td>
<td>-0.71</td>
<td>32.92</td>
</tr>
<tr>
<td>2</td>
<td>23.83</td>
<td>36.65</td>
<td>47.28</td>
<td>-1.94</td>
<td>15.23</td>
</tr>
<tr>
<td>3</td>
<td>22.82</td>
<td>32.19</td>
<td>41.34</td>
<td>-50.31</td>
<td>-12.95</td>
</tr>
<tr>
<td>6</td>
<td>20.85</td>
<td>25.00</td>
<td>31.90</td>
<td>-87.43</td>
<td>-42.16</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>Statistics</th>
<th>A. AR(1)</th>
<th>B. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predictive variation</td>
<td>Predictive variation</td>
<td>t-statistics against RW</td>
</tr>
<tr>
<td>h</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>LIBOR maturity in months:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.57</td>
<td>-3.22</td>
<td>-4.89</td>
</tr>
<tr>
<td>2</td>
<td>-1.50</td>
<td>-3.24</td>
<td>-5.04</td>
</tr>
<tr>
<td>3</td>
<td>-1.98</td>
<td>-3.71</td>
<td>-5.45</td>
</tr>
<tr>
<td>6</td>
<td>-3.36</td>
<td>-5.62</td>
<td>-7.49</td>
</tr>
<tr>
<td>9</td>
<td>-4.52</td>
<td>-7.14</td>
<td>-9.17</td>
</tr>
<tr>
<td>12</td>
<td>-4.90</td>
<td>-7.78</td>
<td>-9.87</td>
</tr>
</tbody>
</table>

**AR(1) is the worst;**

**15-factor model beats random walk in and out!**

Liuren Wu (Baruch)

Cascade Dynamics with Power Law Scaling

February 3, 2010
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 6bps — 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.
Concluding remarks

- Within the DTSM framework, we make several key assumptions:
  - A cascade factor structure: Eliminate factor rotation. Pin down the meaning of each factor. Provide a natural separation/filtration of different frequency components.
  - IID risk and risk premium: Two parameters to control the risk and risk premium of all risks.
  - Power law scaling: Two parameters to control the mean reversion speeds of all frequencies.

- The result is a class of dimension-invariant models:
  - The number of parameters is invariant to the number of factors.

- No more curse of dimensionality: high-dimensional models are just as easy to be estimated as low-dimensional models.

- Evidence: High-dimensional models do provide superior performance in several fronts.