A Multifrequency Theory of the Interest Rate Term Structure

Liuren Wu

Baruch College
Joint work with Laurent Calvet and Adlai Fisher

Workshop on Financial Econometrics, The Fields Institute
April 23, 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 model should capture the effects of shocks of all frequencies.
- Curse of dimensionality often forces us to focus on a particular segment of the frequency spectrum...
- We propose a class of models that can include all identifiable frequencies, but with no curse of dimensionality — The models are *dimension invariant*.

A cascade multifrequency interest rate dynamics

- The instantaneous interest rate r_t follows a *cascade* dynamics,

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

  - 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 θ_r, which is also the mean of the short rate.
  - Intuitively, the tendencies are like exponentially weighted moving averages with increasingly long windows.
  - $n \to \infty$ is also an option.

Cascade v. general affine

- The general affine Gaussian models (Duffie & Kan, 96):
  $r_t = a + \mathbf{b}^\top X_t, dX_t = K(\mathbf{c} - X_t)dt + \Sigma dW$.
  - Factors can rotate. For example, equivalently,
    $r_t = a' + (\mathbf{b}')^\top Z_t, dZ_t = -K' Z_t dt + dW$, with
    $a' = a + \mathbf{b}^\top \mathbf{c}, \mathbf{b}' = \Sigma \mathbf{b}, \mathbf{c}' = \Sigma^{-1}\mathbf{c}, K' = \Sigma^{-1} K$.
    - 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.

Dimension invariant assumption on frequency distribution

- 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 an variable x_t diminishes as the lag j becomes larger: $\beta_j = \beta_0 \lambda^j$.
  - The focus of this paper is on the interest rate term structure modeling. Volatility variation is largely "unspanned" by the term structure.
    $\Rightarrow$ We assume *IID* risks with constant volatility: $\sigma_{j,t} = \sigma_r$.
  - Risk premium is not important for term structure modeling, either.
    $\Rightarrow$ We assume *constant and identical* market prices for risks of all frequencies: $\gamma_{j,t} = \gamma_r$.
- **The result**: Five parameters $(\theta_r, \sigma_r, \kappa_r, b, \gamma_r)$, regardless of the number of frequencies (n).

Dimension invariant extensions

- **Option pricing:**
  - Variance dynamics specification becomes important. A potentially useful dimension-invariant cascade specification for variance:

$$
\begin{aligned}
\sigma_{j,t}^2 &= v_{K,t} \\
dv_{k,t} &= \kappa_k^v \left(v_{k-1,t} - v_{k,t}\right) dt + \omega \sqrt{v_{K,t}} dZ_{k,t}, \quad k = K, K-1, \cdots, 1, \\
v_{0,t} &= \theta_v, \\
\rho &= \mathbb{E}[dW_{j,t} dZ_{k,t}]/dt, \quad \kappa_k^v = \beta^{k-1} \kappa_1^v, \quad \beta > 1.
\end{aligned}
$$

- **Bond risk premia/expectation hypotheses**:
  - The different frequency components that we identify can be used as the instruments to explain the bond risk premium.
    - 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.

Bond pricing

- The values of zero-coupon bonds are exponential-affine in $X_t = \{x_{j,t}\}_{j=1}^n$,

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

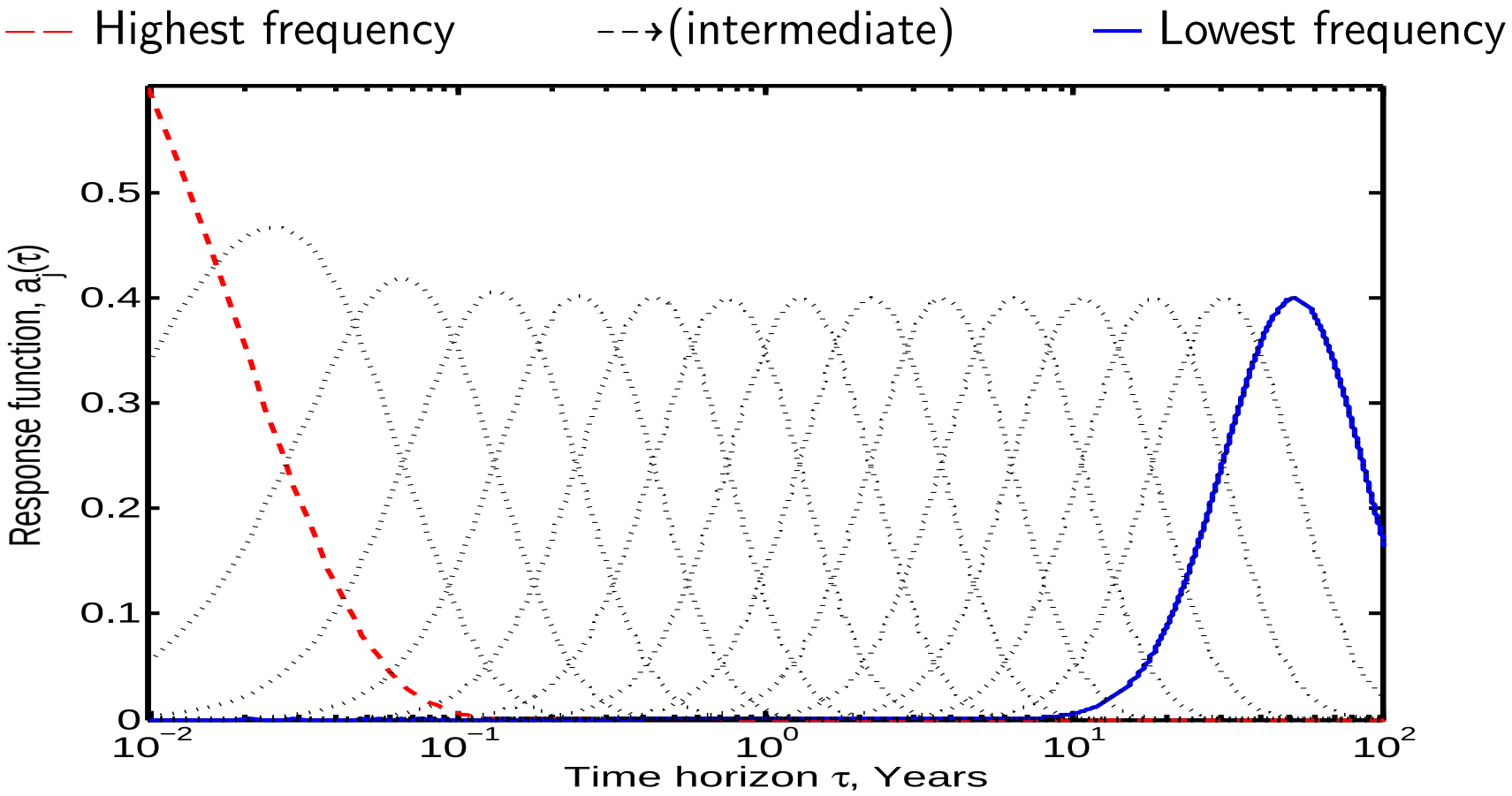
- The instantaneous forward rate is affine in the state vector,

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

  - The intercept has 3 components: long-run mean, risk premium, convexity: $e(\tau) =$
$$\begin{cases} \kappa_r \theta_r \sum_{i=1}^n \alpha_{i,1}\left(1 - e^{-\kappa_i \tau}\right) \\ -\gamma_r \sigma_r^2 \sum_{j=1}^n \sum_{i=j}^n \alpha_{i,j}\left(1 - e^{-\kappa_i \tau}\right) \\ -\frac{\sigma_r^2}{2} \sum_{j=1}^n \sum_{i=j}^n \sum_{k=j}^n \alpha_{i,j}\alpha_{k,j}\left(1 - e^{-\kappa_k \tau} - e^{-\kappa_i \tau} + e^{-(\kappa_i + \kappa_k)\tau}\right) \end{cases}$$
  - The loading coefficients $a(\tau)$ are convolutions of exponentials.

The forward rate response functions

to shocks from different frequency components, $a_j(\tau)$:



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

One can think of $a(\tau)$ as basis functions and X_t as time-varying weights.

Data and estimation

- 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; 10,170 data points.
- 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.

Dimensionality

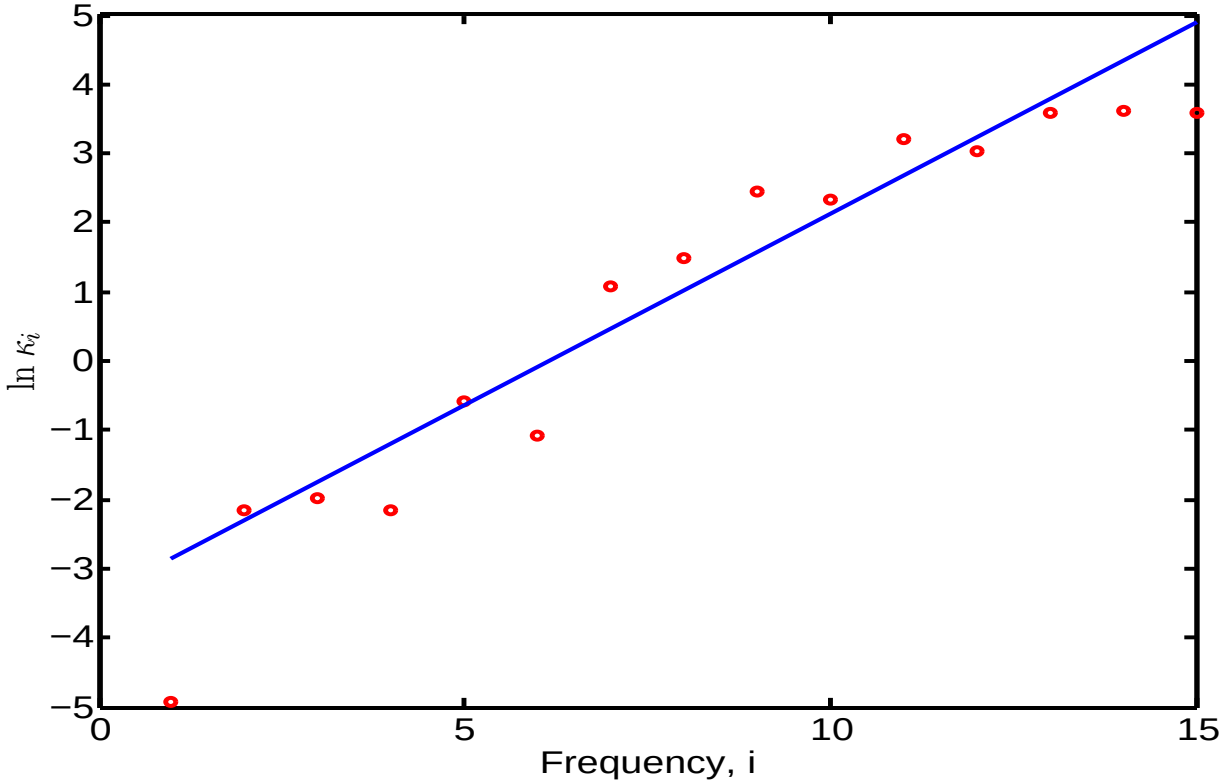
- Normally, this is the first thing one decides on before one can pin down the parameter space.
- Under our framework, the parameter space is invariant to the dimensionality decision. We worry about the dimensionality the last.
- Since we have 15 interest rate series, we estimate 15 models with $n = 1, 2, 3, \cdots 15$.
- 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. 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

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

- Vuong test (last column): *More is significantly better.*
- Spacing (b) is finer when more is allowed.
- Parameters (κ_r, σ_r) stabilize as n increases.

Power law scaling: Theory and evidence



Circles: κ_i as free parameters; Solid line: power-law scaling

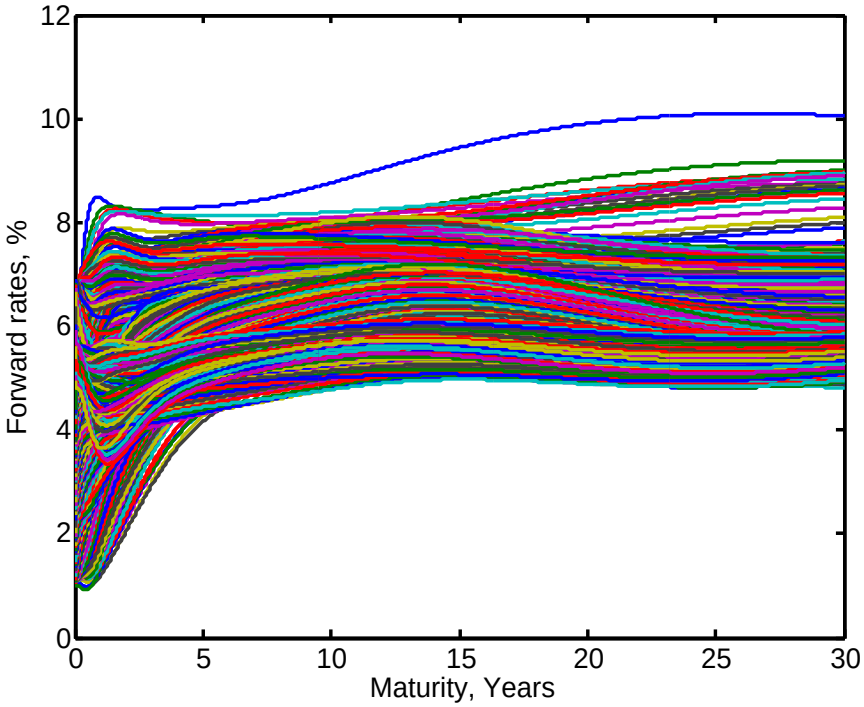
Power-scaling is reasonable.

In-sample fitting performance: Pricing error statistics

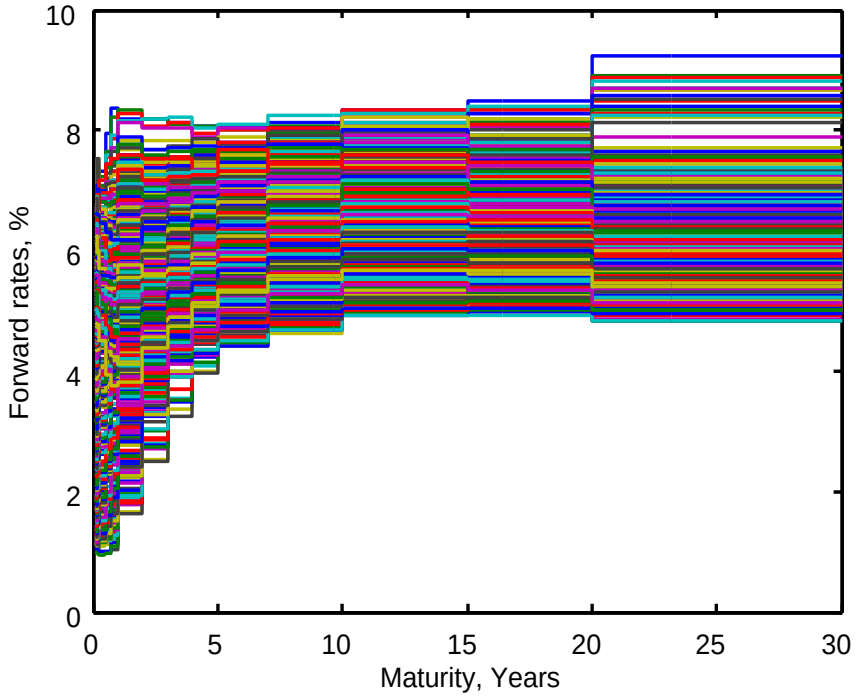
Model	A. Three-factor model					B. 15-factor model				
Maturity	Mean	Rmse	Auto	Max	VR	Mean	Rmse	Auto	Max	VR
1 m	-0.68	7.47	0.86	43.93	99.83	0.02	0.62	0.36	5.40	100.00
2 m	0.63	3.82	0.69	37.42	99.96	0.01	1.76	0.52	16.31	99.99
3 m	1.61	5.03	0.85	42.54	99.93	-0.11	1.79	0.60	18.96	99.99
6 m	0.39	6.78	0.93	24.05	99.86	0.04	1.06	0.59	8.78	100.00
9 m	-1.74	6.88	0.89	32.06	99.86	0.38	0.92	0.69	4.31	100.00
1 y	-3.06	6.74	0.79	33.00	99.88	-0.49	1.21	0.06	4.71	100.00
2 y	2.11	6.17	0.81	24.38	99.86	0.28	1.09	-0.02	4.52	100.00
3 y	1.97	6.90	0.88	34.12	99.78	-0.19	0.75	0.36	3.88	100.00
4 y	0.87	6.32	0.90	33.48	99.76	-0.04	0.81	0.16	8.08	100.00
5 y	-0.21	5.85	0.90	27.63	99.76	0.07	0.73	0.20	4.60	100.00
7 y	-1.89	5.55	0.92	17.32	99.77	0.08	0.70	0.35	6.86	100.00
10 y	-2.35	5.17	0.89	18.65	99.78	-0.12	0.95	0.23	9.00	99.99
15 y	0.88	3.87	0.86	13.14	99.82	0.00	0.72	0.29	4.68	99.99
20 y	1.91	5.35	0.90	17.64	99.66	0.08	0.79	0.33	6.90	99.99
30 y	-0.76	9.67	0.95	31.88	98.68	-0.09	0.71	0.23	4.82	99.99
Average	-0.02	**6.11**	0.87	28.75	99.75	-0.00	**0.98**	0.33	7.45	99.99

Application: Yield curve stripping

Model-generated forward curves



Piece-wise constant assumption



- 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.

In-sample forecasting performance

Predictive variation: $1 - \frac{\text{Mean Squared Forecasting Error}}{\text{Mean Squared Interest Rate Change}}$

Model	A. AR(1)			B. Three-factor model			C. 15-factor model			
h	1	2	3	1	2	3	1	2	3	weeks
LIBOR maturity in months:										
1	25.85	43.84	57.50	-0.71	32.92	42.84	21.71	40.82	52.16	
2	23.83	36.65	47.28	-1.94	15.23	23.31	17.65	28.50	37.00	
3	22.82	32.19	41.34	-50.31	-12.95	1.57	8.78	21.86	29.17	
6	20.85	25.00	31.90	-87.43	-42.16	-24.57	5.77	12.56	16.94	
9	20.22	19.35	23.79	-67.23	-38.76	-28.15	1.30	4.99	7.06	
12	21.45	17.53	20.58	-39.25	-26.45	-21.32	6.85	3.71	3.07	

AR(1) is the best;
3-factor model cannot beat random walk.

Out-of-sample forecasting performance

Model	A. AR(1)			B. 15-factor model						
Statistics	Predictive variation			Predictive variation			t-statistics against RW			
h	1	2	3	1	2	3	1	2	3	(weeks)
LIBOR maturity in months:										
1	-1.57	-3.22	-4.89	24.24	38.91	52.76	1.73	3.49	4.80	
2	-1.50	-3.24	-5.04	19.59	28.00	40.31	1.68	3.48	5.03	
3	-1.98	-3.71	-5.45	9.80	21.90	32.77	1.69	4.75	6.33	
6	-3.36	-5.62	-7.49	8.45	14.70	21.36	2.46	4.58	5.83	
9	-4.52	-7.14	-9.17	4.71	7.74	11.53	2.26	3.53	4.15	
12	-4.90	-7.78	-9.87	7.94	4.78	4.85	3.63	2.33	2.02	

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) neither uses the term structure information, nor 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.
- 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

- 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.