Statistical Arbitrage Based on No-Arbitrage Dynamic Term Structure Models

Liuren Wu

Zicklin School of Business, Baruch College
1 General Ideas
2 Bond market terminologies
3 Dynamic term structure models
4 Model design and estimation
5 Statistical arbitrage trading
The securities have direct claims to future cash flows.

Valuation is based on forecasts of future cash flows and risk:
- DCF (Discounted Cash Flow Method): Discount time-series forecasted future cash flow with a discount rate that is commensurate with the forecasted risk.

Investment: Buy if market price is lower than model value; sell otherwise.

Both valuation and investment depend crucially on forecasts of future cash flows (earnings, growth rates) and risks (beta, credit risk).

This is not what we do.
Payoffs are linked directly to the price of an “underlying” security.

Valuation is mostly based on replication/hedging arguments.
- Find a portfolio that includes the underlying security, and possibly other related derivatives, to replicate the payoff of the target derivative.
- The value of the derivative security should be equal to the replicating cost; otherwise, there is an arbitrage.
- Models of this type are called “no-arbitrage” models.

Key: No forecasts are involved. Valuation is based on *cross-sectional comparison*.
- It is not about whether the underlying security price will go up or down, but about the relative pricing relation between the underlying and the derivatives under all possible scenarios.
- If you can replicate the derivative for a cost of $1, you are willing to sell the derivative for (a little more than) $1, regardless of whether the derivative or underlying price will go up or down in the future.

Investment: Find value mismatches between the target security and the replicating portfolio.
Readings behind the technical jargons: \( P \) v. \( Q \)

- **\( P \):** Actual probabilities that earnings will be high or low, estimated based on historical data and other insights about the company.
  
  Valuation is all about getting the forecasts right and assigning the appropriate price ("risk premium") for the forecasted risk — *fair wrt future cashflows/risks and your risk preference.*

- **\( Q \):** "Risk-neutral" probabilities that we can use to aggregate expected future payoffs and discount them back with riskfree rate, regardless of how risky the cash flow is.
  
  It is related to real-time scenarios, but has nothing to do with real-time probability.
  
  Since the intention is to hedge away risk under all scenarios and discount back with riskfree rate, we do not really care about the actual probability of each scenario happening.
  
  We just care about what all the possible scenarios are and whether our hedging works under all scenarios.

  \( Q \) is not about getting close to the actual probability, but about being *fair wrt the prices of securities that you use for hedging/replicating.*
Consider a non-dividend paying stock in a world with zero riskfree interest rate. Currently, the market price for the stock is $100. What should be the forward price for the stock with one year maturity?

- The forward price is $100.
  - Standard forward pricing argument says that the forward price should be equal to the cost of buying the stock and carrying it over to maturity.
  - The buying cost is $100, with no storage or interest cost or dividend benefit.

How should you value the forward differently if you have inside information that the company will be bought tomorrow and the stock price is going to double?

- Shorting a forward at $100 is still safe for you if you can buy the stock at $100 to hedge.
Investing in derivative securities without forecasts

- If you can **really** forecast the cashflow (with inside information), you probably do not care much about hedging or no-arbitrage modeling.
  - What is risk to the market is an opportunity for you.
  - You just lift the market and try not getting caught for inside trading.

- But if you do not have insights on cash flows (earnings growth etc) and still want to invest in derivatives, the focus does not need to be on forecasting, but on cross-sectional consistency.
  - The no-arbitrage pricing models can be useful.
Basic idea:

- Interest rates across different maturities are related.

- A dynamic term structure model (DTSM) provides a functional form for this relation that excludes arbitrage.
  - The model usually consists of specifications of risk-neutral factor dynamics ($X$) and the short rate as a function of the factors, e.g., $r_t = a_r + b_r^\top X_t$.

- Nothing about the forecasts: The “risk-neutral dynamics” are estimated to match historical term structure shapes.

- A model is well-specified if
  - It can fit most of the term structure shapes reasonably well.
  - Hedging against the modeled risk factors generates stable (riskfree) portfolios.
A 3-factor affine model with adjustments for discrete Fed policy changes:

Pricing errors on USD swap rates in bps

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>MAE</th>
<th>Std</th>
<th>Auto</th>
<th>Max</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 y</td>
<td>0.80</td>
<td>2.70</td>
<td>3.27</td>
<td>0.76</td>
<td>12.42</td>
<td>99.96</td>
</tr>
<tr>
<td>3 y</td>
<td>0.06</td>
<td>1.56</td>
<td>1.94</td>
<td>0.70</td>
<td>7.53</td>
<td>99.98</td>
</tr>
<tr>
<td>5 y</td>
<td>-0.09</td>
<td>0.68</td>
<td>0.92</td>
<td>0.49</td>
<td>5.37</td>
<td>99.99</td>
</tr>
<tr>
<td>7 y</td>
<td>0.08</td>
<td>0.71</td>
<td>0.93</td>
<td>0.52</td>
<td>7.53</td>
<td>99.99</td>
</tr>
<tr>
<td>10 y</td>
<td>-0.14</td>
<td>0.84</td>
<td>1.20</td>
<td>0.46</td>
<td>8.14</td>
<td>99.99</td>
</tr>
<tr>
<td>15 y</td>
<td>0.40</td>
<td>2.20</td>
<td>2.84</td>
<td>0.69</td>
<td>16.35</td>
<td>99.90</td>
</tr>
<tr>
<td>30 y</td>
<td>-0.37</td>
<td>4.51</td>
<td>5.71</td>
<td>0.81</td>
<td>22.00</td>
<td>99.55</td>
</tr>
</tbody>
</table>

- Superb pricing performance: R-squared is greater than 99%. Maximum pricing errors is 22bps.
- Pricing errors are transient compared to swap rates (0.99):
  Average half life of the pricing errors is 3 weeks.
  The average half life for swap rates is 1.5 years.
Investing in interest rate swaps based on DTSMs

- If you can forecast interest rate movements,
  - Long swap if you think rates will go down.
  - Forget about dynamic term structure model: It does not help your interest rate forecasting.

- If you cannot forecast interest rate movements (it is hard), use the dynamic term structure model not for forecasting, but as a decomposition tool:

  \[ y_t = f(X_t) + e_t \]

  - What the model captures \((f(X_t))\) is the persistent component, which is difficult to forecast.
  - What the model misses (the pricing error \(e\)) is the more transient and hence more predictable component.

- Form swap-rate portfolios that
  - neutralize their first-order dependence on the persistent factors.
  - only vary with the transient residual movements.

- Result: The portfolios are strongly predictable, even though the individual interest-rate series are not.
For a three-factor model, we can form a 4-swap rate portfolio that has zero exposure to the factors.

- The portfolio should have duration close to zero
  - No systematic interest rate risk exposure.
  - The fair value of the portfolio should be relatively flat over time.

The variation of the portfolio’s market value is mainly induced by short-term liquidity shocks...

Long/short the swap portfolio based on its deviation from the fair model value.

- Provide liquidity to where the market needs it and receives a premium from doing so.
The time-series of 10-year USD swap rates

- The left panel represents the 10-year swap rate hedged with 2, 5, and 30-year swap rates. The right panel is the 10-year swap rate unhedged.
- It is much easier to predict the hedged portfolio (left panel) than the unhedged swap contract (right panel).
- Some regime shifting after the crisis (e.g., zero bound)...

Other economies
Backtesting results from a simple investment strategy

Holding each investment for 4 weeks.
Caveats

- Convergence takes time: We take a 4-week horizon.

- Accurate hedging is vital for the success of the strategy. The model needs to be estimated with dynamic consistency:
  - Parameters are held constant. Only state variables vary.
  - Appropriate model design is important: parsimony, stability, adjustment for some calendar effects.
  - Daily fitting of a simpler model (with daily varying parameters) is dangerous.

- Spread trading (one factor) generates low Sharpe ratios.

- Butterfly trading (2 factors) is also not guaranteed to succeed.

- Reference: Bali, Heidari, Wu, Predictability of Interest Rates and Interest-Rate Portfolios.
Another example: Trading the linkages between sovereign CDS and currency options

- When a sovereign country’s default concern (over its foreign debt) increases, the country’s currency tend to depreciate, and currency volatility tend to rise.
  - “Money as stock” corporate analogy.

- Observation: Sovereign credit default swap spreads tend to move positively with currency’s
  - option implied volatilities (ATMV): A measure of the return volatility.
  - risk reversals (RR): A measure of distributional asymmetry.

Co-movements between CDS and ATMV/RR

Mexico

Brazil

CDS Spread, %

Implied Volatility Factor, %

Risk Reversal Factor, %
A no-arbitrage model that prices both CDS and currency options

- Model specification:
  - At normal times, the currency price (dollar price of a local currency, say peso) follows a diffusive process with stochastic volatility.
  - When the country defaults on its foreign debt, the currency price jumps by a large amount.
  - The arrival rate of sovereign default is also stochastic and correlated with the currency return volatility.

- Under these model specifications, we can price both CDS and currency options via no-arbitrage arguments. The pricing equations is tractable. Numerical implementation is fast.

- Estimate the model with dynamic consistency: Each day, three things vary: (i) Currency price (both diffusive moves and jumps), (ii) currency volatility, and (iii) default arrival rate.

- All model parameters are fixed over time.
The hedged portfolio of CDS and currency options

Suppose we start with an option contract on the currency. We need four other instruments to hedge the risk exposure of the option position:

1. The underlying currency to hedge infinitesimal movements in exchange rate
2. A risk reversal (out of money option) to hedge the impact of default on the currency value.
3. A straddle (at-the-money option) to hedge the currency volatility movement.
4. A CDS contract to hedge the default arrival rate variation.

The portfolio needs to be rebalanced over time to maintain neutral to the risk factors.

- The value of hedged portfolio is much more transient than volatilities or cds spreads.
Back-testing results

Wu (Baruch) Statistical Arbitrage
Similar linkages between corporate CDS and stock options

All series are standardized to have similar scales in the plots.

Reference: Carr and Wu, Stock Options and Credit Default Swaps: A Joint Framework for Valuation and Estimation
Each bar represents one hedged portfolio. Each hedged portfolio includes 5 instruments: two CDS contracts, two options at two maturities, and the underlying stock.
If you have a working crystal ball, others’ risks become your opportunities.
- Forget about no-arbitrage models; lift the market.

No-arbitrage type models become useful when
- You cannot forecast the future accurately: Risk persists.
- Hedge risk exposures.
- Perform statistical arbitrage trading on derivative products that profit from short-term market dislocations.

Caveats
- When hedging is off, risk can overwhelm profit opportunities.
- Accurate hedging requires modeling all risk dimensions.
  - Interest rates do not just move in parallel, but also experience systematic moves in slopes and curvatures.
  - Capital structure arbitrage: Volatility and default rates are not static, but vary strongly over time in unpredictable ways.

*Fulfilling these requirements is a quant’s job!*
Plan of this lecture series

- We illustrate the idea using the first example: Interest rate swap statistical arbitrage using dynamic term structure models.
  - The math and implementation are both simpler for term structure models than for option pricing models.
  - The idea is the same.

- Things we need to learn before we can put on a trade:
  - Bond market terminologies
  - Theories on dynamic term structure models (DTSM)
    - Not a summary of existing models, but an approach to design models that are simple, parsimonious, and yet flexible enough to match the data behavior.
  - Model estimation:
    - Data sources, market conventions.
    - Maximum likelihood estimation with unscented Kalman filter.
  - Implementation details on portfolio construction, P&L calculation.
Notations and terminologies

- $B(t, T)$ — the time-$t$ value of a zero-coupon bond that pays $1$ at time $T$.

- $y(t, T)$ — the time-$t$ continuously compounded spot interest rate with expiry rate $T$.

  \[ y(t, T) = -\frac{\ln B(t, T)}{T - t}, \quad B(t, T) = \exp(-y(t, T)(T-t)), \quad B(t, t) = 1 \]

- $f(t, T, S)$ — The time-$t$ continuously compounded forward rate prevailing at future time interval $(T, S)$:

  \[ f(t, T, S) = -\frac{\ln B(t, S)/B(t, T)}{S - T}, \quad \frac{B(t, S)}{B(t, T)} = e^{-f(t, T, S)(S-T)} \]

- $f(t, T)$ — the time-$t$ instantaneous forward rate prevailing at $T$:

  \[ f(t, T) = f(t, T, T) = -\frac{\partial \ln B(t, T)}{\partial T}, \quad B(t, T) = e^{-\int_t^T f(t, s)ds} \]

- $r(t)$ — the instantaneous interest rate defined by the limit:

  \[ r(t) = f(t, t) = \lim_{T \downarrow t} y(t, T). \]
The pricing kernel

- If there is no arbitrage in a market, there must exist at least one strictly positive process $M_t$ such that the deflated gains process associated with any trading strategy is a martingale:

$$M_t P_t = \mathbb{E}_t^P [M_T P_T] ,$$

where $P_t$ denotes the time-$t$ value of a trading strategy, $\mathbb{P}$ denotes the statistical probability measure.

- Consider the strategy of buying a zero-coupon bond at time $t$ and hold it to maturity at $T$, we have

$$M_t B(t, T) = \mathbb{E}_t^P [M_T B(T, T)] = \mathbb{E}_t^P [M_T] .$$

- If the market is complete, this process $M_t$ is unique.

- If the available securities cannot complete the market, there can be multiple processes that satisfy equation (1).

- If we cannot find a single positive process that can price all securities, there is arbitrage.

- $M_t$ is called the state-price deflator. The ratio $M_{t, T} = M_T / M_t$ is called the stochastic discount factor, or the pricing kernel.
From pricing kernel to exchange rates

- Let $M_{t,T}^h$ denote the pricing kernel in economy $h$ that prices all securities in that economy with its currency denomination.
- The $h$-currency price of currency-$f$ ($h$ is home currency) is linked to the pricing kernels of the two economies by,

$$\frac{S_{fh}^T}{S_{fh}^t} = \frac{M_{t,T}^f}{M_{t,T}^h}$$

- The log currency return over period $[t, T]$, $\ln S_{T}^{fh}/S_{t}^{fh}$ equals the difference between the log pricing kernels of the two economies.
- Let $S$ denote the dollar price of pound (e.g. $S_t = \$2.06$), then $\ln S_{T}/S_{t} = \ln M_{t,T}^{pound} - \ln M_{t,T}^{dollar}$.
- If markets are completed by primary securities (e.g., bonds and stocks), there is one unique pricing kernel per economy. The exchange rate movement is uniquely determined by the ratio of the pricing kernels.
- If the markets are not completed by primary securities, exchange rates (and currency options) help complete the markets by requiring that...
Multiplicative decomposition of the pricing kernel

In a discrete-time representative agent economy with an additive CRRA utility, the pricing kernel equals the ratio of the marginal utilities of consumption,

\[ M_{t,t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)} = \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} , \quad \gamma - \text{relative risk aversion.} \]

In continuous time, it is convenient to perform the following multiplicative decomposition on the pricing kernel:

\[ M_{t,T} = \exp \left( - \int_t^T r_s ds \right) \mathcal{E} \left( - \int_t^T \gamma_s dX_s \right) \]

\( r \) is the instantaneous riskfree rate, \( \mathcal{E} \) is the stochastic exponential martingale operator, \( X \) denotes the risk sources in the economy, and \( \gamma \) is the market price of the risk \( X \).

- If \( X_t = W_t \), \( \mathcal{E} \left( - \int_t^T \gamma_s dW_s \right) = e^{- \int_t^T \gamma_s dW_s - \frac{1}{2} \int_t^T \gamma_s^2 ds} \).
- In a continuous time version of the representative agent example, \( dX_s = d\ln c_t \) and \( \gamma \) is relative risk aversion.

\( r \) is normally a function of \( X \).
Recall the multiplicative decomposition of the pricing kernel:

\[ M_{t,T} = \exp \left( - \int_t^T r_s ds \right) \mathcal{E} \left( - \int_t^T \gamma_s^\top dX_s \right) \]

Given the pricing kernel, the value of the zero-coupon bond can be written as

\[ B(t, T) = \mathbb{E}_t^P [M_{t,T}] = \mathbb{E}_t^P \left[ \exp \left( - \int_t^T r_s ds \right) \mathcal{E} \left( - \int_t^T \gamma_s^\top dX_s \right) \right] = \mathbb{E}_t^Q \left[ \exp \left( - \int_t^T r_s ds \right) \right] \]

where the measure change from \( \mathbb{P} \) to \( \mathbb{Q} \) is defined by the exponential martingale \( \mathcal{E} (\cdot) \).

The risk sources \( X \) and their market prices \( \gamma \) matter for bond pricing through the correlation between \( r \) and \( X \).
Consider the following exponential martingale that defines the measure change from $\mathbb{P}$ to $\mathbb{Q}$:

$$
\frac{d\mathbb{Q}}{d\mathbb{P}} \bigg|_t = \mathcal{E} \left( - \int_0^t \gamma_s dW_s \right),
$$

- If the $\mathbb{P}$ dynamics is: $dS_t^i = \mu_t^i S_t^i dt + \sigma_t^i S_t^i dW_t^i$, with $\rho_t^i dt = \mathbb{E}[dW_t^i dW_t^j]$, then the dynamics of $S_t^i$ under $\mathbb{Q}$ is:
  $$
  dS_t^i = \mu_t^i S_t^i dt + \sigma_t^i S_t^i dW_t^i + \mathbb{E}[\gamma_t dW_t^i, \sigma_t^i S_t^i dW_t^i] = 
  (\mu_t^i - \gamma^i) S_t^i dt + \sigma_t^i S_t^i dW_t^i
  $$
- If $S_t^i$ is the price of a traded security, we need $r = \mu_t^i - \gamma^i \rho_t^i$. The risk premium on the security is $\mu_t^i - r = \gamma^i \rho_t^i$.

How do things change if the pricing kernel is given by:

$$
M_{t,T} = \exp \left( - \int_t^T r_s ds \right) \mathcal{E} \left( - \int_t^T \gamma_s^\top dW_s \right) \mathcal{E} \left( - \int_t^T \gamma_s^\top dZ_s \right),
$$

where $Z_t$ is another set of Brownian motions independent of $W_t$ or $W_t^i$.

- $Z$ does not affect bond pricing if it is not correlated with $r$. 
Let $X$ denote a pure jump process with its compensator being $\nu(x, t)$ under $\mathbb{P}$.

Consider a measure change defined by the exponential martingale:

$$\frac{d\mathbb{Q}}{d\mathbb{P}}|_t = \mathcal{E}(-\gamma X_t),$$

The compensator of the jump process under $\mathbb{Q}$ becomes:

$$\nu(x, t)^{\mathbb{Q}} = e^{-\gamma x} \nu(x, t).$$

Example: Merton (176)'s compound Poisson jump process,

$$\nu(x, t) = \lambda \frac{1}{\sigma_J \sqrt{2\pi}} e^{-\frac{(x-\mu_J)^2}{2\sigma_J^2}}. \text{ Under } \mathbb{Q}, \text{ it becomes}$$

$$\nu(x, t)^{\mathbb{Q}} = \lambda \frac{1}{\sigma_J \sqrt{2\pi}} e^{-\gamma x - \frac{(x-\mu_J)^2}{2\sigma_J^2}} = \lambda^{\mathbb{Q}} \frac{1}{\sigma_J \sqrt{2\pi}} e^{-\frac{(x-\mu_J^{\mathbb{Q}})^2}{2\sigma_J^2}} \text{ with}$$

$$\mu^{\mathbb{Q}} = \mu_J - \gamma \sigma_J^2 \text{ and } \lambda^{\mathbb{Q}} = \lambda e^{\frac{1}{2} \gamma (\gamma \sigma_J^2 - 2\mu_J)}.$$

Dynamic term structure models

A long list of papers propose different dynamic term structure models:

- **Specific examples:**
  - Vasicek, 1977, JFE: The instantaneous interest rate follows an Ornstein-Uhlenbeck process.
  - Cox, Ingersoll, Ross, 1985, Econometrica: The instantaneous interest rate follows a square-root process.
  - Many multi-factor examples ...

- **Classifications (back-filling)**
  - Duffie, Kan, 1996, Mathematical Finance: Spot rates are affine functions of state variables.
  - Leippold, Wu, 2002, JFQA: Spot rates are quadratic functions of state variables.
  - Filipovic, 2002, Mathematical Finance: How far can we go?
The “back-filling” procedure of DTSM identification

- The traditional procedure:
  - First, we make assumptions on factor dynamics ($X$), market prices ($\gamma$), and how interest rates are related to the factors $r(X)$, based on what we think is reasonable.
  - Then, we derive the fair valuation of bonds based on these dynamics and market price specifications.

- The back-filling (reverse engineering) procedure:
  - First, we state the form of solution that we want for bond prices (spot rates).
  - Then, we try to figure out what dynamics specifications generate the pricing solutions that we want.
    - The dynamics are not specified to be reasonable, but specified to generate a form of solution that we like.

- It is good to be able to go both ways.
  - It is important not only to understand existing models, but also to derive new models that meet your work requirements.
What we want: Zero-coupon bond prices are exponential affine functions of state variables.
- Continuously compounded spot rates are affine in state variables.
- It is simple and tractable. We can use spot rates as factors.

Let $X$ denote the state variables, let $B(X_t, \tau)$ denote the time-$t$ fair value of a zero-coupon bond with time to maturity $\tau = T - t$, we have

$$B(X_t, \tau) = \mathbb{E}^Q_t \left[ \exp \left( - \int_t^T r(X_s) ds \right) \right] = \exp \left( -a(\tau) - b(\tau)^\top X_t \right)$$

Implicit assumptions:
- By writing $B(X_t, \tau)$ and $r(X_t)$, and solutions $A(\tau), b(\tau)$, I am implicitly focusing on time-homogeneous models. Calendar dates do not matter. This assumption is for (notational) simplicity more than anything else.
- With calendar time dependence, the notation can be changed to, $B(X_t, t, T)$ and $r(X_t, t)$. The solutions would be $a(t, T), b(t, T)$.

Questions to be answered:
- What is the short rate function $r(X_t)$?
- What's the dynamics of $X_t$ under measure $Q$?
Diffusion dynamics

- To make the derivation easier, let’s focus on diffusion factor dynamics: $dX_t = \mu(X)dt + \sigma(X)dW_t$ under $\mathbb{Q}$.
- We want to know: What kind of specifications for $\mu(X), \sigma(X)$ and $r(X)$ generate the affine solutions?
- For a generic valuation problem,

$$f(X_t, t, T) = \mathbb{E}_t^\mathbb{Q} \left[ \exp \left( - \int_t^T r(X_s)ds \right) \Pi_T \right],$$

where $\Pi_T$ denotes terminal payoff, the value satisfies the following partial differential equation:

$$f_t + \mathcal{L}f = rf, \quad \mathcal{L}f - \text{infinitesimal generator}$$

with boundary condition $f(T) = \Pi_T$.
- Apply the PDE to the bond valuation problem,

$$B_t + B_X^\top \mu(X) + \frac{1}{2} \sum B_{XX} \cdot \sigma(X)\sigma(X)^\top = rB$$

with boundary condition $B(X_T, 0) = 1$. 
Starting with the PDE,

\[ B_t + B_X^\top \mu(X) + \frac{1}{2} \sum B_{XX} \cdot \sigma(X)\sigma(X)^\top = rB, \quad B(X_T, 0) = 1. \]

If \( B(X_t, \tau) = \exp(-a(\tau) - b(\tau)^\top X_t) \), we have

\[ B_t = B \left( a'(\tau) + b'(\tau)^\top X_t \right), \quad B_X = -Bb(\tau), \quad B_{XX} = Bb(\tau)b(\tau)^\top, \quad y(t, \tau) = \frac{1}{\tau} \left( a(\tau) + b(\tau)^\top X_t \right), \quad r(X_t) = a'(0) + b'(0)^\top X_t = a_r + b_r^\top X_t. \]

Plug these back to the PDE,

\[ a'(\tau) + b'(\tau)^\top X_t - b(\tau)^\top \mu(X) + \frac{1}{2} \sum b(\tau)b(\tau)^\top \cdot \sigma(X)\sigma(X)^\top = a_r + b_r^\top X_t. \]

Question: What specifications of \( \mu(X) \) and \( \sigma(X) \) guarantee the above PDE to hold at all \( X \)?

- Power expand \( \mu(X) \) and \( \sigma(X)\sigma(X)^\top \) around \( X \) and then collect coefficients of \( X^p \) for \( p = 0, 1, 2, \ldots \). These coefficients have to be zero separately for the PDE to hold at all times.
Back filling

\[ a'(\tau) + b'(\tau)^T X_t - b(\tau)^T \mu(X) + \frac{1}{2} \sum b(\tau)b(\tau)^T \cdot \sigma(X)\sigma(X)^T = a_r + b_r^T X_t \]

- Set \( \mu(X) = a_m + b_m X + c_m XX^T + \cdots \) and
  \[ [\sigma(X)\sigma(X)^T]_i = \alpha_i + \beta_i^T X + \eta_i XX^T + \cdots, \]
  and collect terms:
  - constant \( a'(\tau) - b(\tau)^T a_m + \frac{1}{2} \sum b(\tau)b(\tau)^T \cdot \alpha_i = a_r \)
  - \( X \) \( b'(\tau)^T - b(\tau)^T b_m + \frac{1}{2} \sum b(\tau)b(\tau)^T \cdot \beta_i = b_r^T \)
  - \( XX^T \) \( -b(\tau)^T c_m + \frac{1}{2} \sum b(\tau)b(\tau)^T \cdot \eta_i = 0 \)

- The quadratic and higher-order terms are almost surely zero.
- We thus have the conditions to have exponential-affine bond prices:
  \[ \mu(X) = a_m + b_m X, \quad [\sigma(X)\sigma(X)^T]_i = \alpha_i + \beta_i^T X, \quad r(X) = a_r + b_r^T X. \]
- We can solve the coefficients \([a(\tau), b(\tau)]\) via the following ordinary differential equations:
  \[ a'(\tau) = a_r + b(\tau)^T a_m - \frac{1}{2} \sum b(\tau)b(\tau)^T \cdot \alpha_i \]
  \[ b'(\tau) = b_r + b_m b(\tau) - \frac{1}{2} \sum b(\tau)b(\tau)^T \cdot \beta_i \]
From \( \mathbb{Q} \) to \( \mathbb{P} \), not from \( \mathbb{P} \) to \( \mathbb{Q} \)

- The affine conditions,

\[
\begin{align*}
\mu(X) &= a_m + b_m X, \\
[\sigma(X)\sigma(X)^\top]_i &= \alpha_i + \beta_i^\top X, \\
r(X) &= a_r + b_r^\top X.
\end{align*}
\]

are about the \( \mathbb{Q} \)-dynamics, not about the \( \mathbb{P} \)-dynamics.

- Traditionally, researchers start with the \( \mathbb{P} \)-dynamics (the real thing), and specify risk preferences (market prices of risks). From these two, they derive the \( \mathbb{Q} \)-dynamics and derivative (bond) prices.

- But the back-filling exercise shows that the real requirement for tractability (e.g., affine) is on \( \mathbb{Q} \)-dynamics.

Hence, we should also back fill the specification:

- Start with \( \mathbb{Q} \)-dynamics specification to generate tractable derivative pricing solutions.
- Fill in any market price of risk specification.
- It does not matter if the \( \mathbb{P} \)-dynamics is complex or not.
A one-factor example

Suppose the short rate follows the following $\mathbb{P}$-dynamics

$$dr_t = (a + br_t + cr_t^2 + dr_t^3 + er_t^4) \, dt + \sigma_r \, dW_t$$

and we specify the martingale component of the pricing kernel as

$$\mathcal{E} \left( - \int_t^T \gamma_s \, dW_t \right), \text{ with}$$

$$\gamma_s = \gamma_0 + \gamma_1 r_t + \left( \frac{c}{\sigma_r} \right) r_t^2 + \left( \frac{d}{\sigma_r} \right) r_t^3 + \left( \frac{d}{\sigma_r} \right) r_t^4.$$ 

Can you price zero-coupon bonds based on this specification?

Reversely, if the $\mathbb{Q}$-dynamics are given by

$$\mu(X) = a_m + b_m X, \quad [\sigma(X)\sigma(X)^\top]_i = \alpha_i + \beta_i^\top X$$

and the measure change is defined by $\mathcal{E} \left( - \int_t^T \gamma_s^\top \, dW_t \right)$, we have the drift of the $\mathbb{P}$-dynamics as

$$\mu(X)^\mathbb{P} = a_m + b_m X + \sigma(X) \gamma_t$$

The market price of risk $\gamma_t$ can be anything...
Can you identify the conditions for quadratic models: Bond prices are exponential quadratic in state variables?

Can you identify the conditions for “cubic” models: Bond prices are exponential cubic in state variables?

Affine bond prices: Recently Xavier Gabaix derive a model where bond prices are affine (not exponential affine!) in state variables.
The models that we have derived can have many many factors.

How many factors you use depend on the data (how many sources of independent variation) – A PCA analysis is useful.

The super general specification has many parameters that are not identifiable. You need to analyze the model carefully to get rid of redundant parameters.

Once you know that you have the capability to go complex, you actually want to go as simple as possible.

*Everything should be made as simple as possible, but not simpler.*

— Albert Einstein.
Data

- We estimate a DTSM using US dollar LIBOR and swap rates:
  - LIBOR maturities: 1, 2, 3, 6, and 12 months
  - Swap maturities: 2, 3, 5, 7, 10, 15, and 30 years.
- Quoting conventions: actual/360 for LIBOR; 30/360 with semi-annual payment for swaps.

\[
\begin{align*}
\text{LIBOR}(X_t, \tau) & = \frac{100}{\tau} \left( \frac{1}{B(X_t, \tau)} - 1 \right), \text{(simple compounding rates)} \\
\text{SWAP}(X_t, \tau) & = 200 \times \frac{1 - B(X_t, \tau)}{\sum_{i=1}^{2\tau} B(X_t, i/2)}, \text{(par bond coupon rates)}.
\end{align*}
\]

- LIBOR and swaps on other currencies have slightly different day-counting conventions.
- Sample period: I want it as long as possible, but data are not fully available before 1995.
- Sampling frequency: For model estimation, I often sample data weekly to avoid weekday effects.
Data: Time series and term structure

LIBOR and swap rates, %

USLBOR and swap rates

Wu (Baruch) Statistical Arbitrage
### Data: Summary statistics

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>Auto</th>
</tr>
</thead>
<tbody>
<tr>
<td>1m</td>
<td>4.329</td>
<td>1.807</td>
<td>-0.703</td>
<td>-1.077</td>
<td>0.998</td>
</tr>
<tr>
<td>2m</td>
<td>4.365</td>
<td>1.813</td>
<td>-0.710</td>
<td>-1.067</td>
<td>0.998</td>
</tr>
<tr>
<td>3m</td>
<td>4.399</td>
<td>1.818</td>
<td>-0.708</td>
<td>-1.055</td>
<td>0.998</td>
</tr>
<tr>
<td>6m</td>
<td>4.474</td>
<td>1.814</td>
<td>-0.705</td>
<td>-0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>9m</td>
<td>4.550</td>
<td>1.800</td>
<td>-0.689</td>
<td>-0.935</td>
<td>0.997</td>
</tr>
<tr>
<td>12m</td>
<td>4.638</td>
<td>1.780</td>
<td>-0.659</td>
<td>-0.874</td>
<td>0.996</td>
</tr>
<tr>
<td>2y</td>
<td>4.887</td>
<td>1.576</td>
<td>-0.547</td>
<td>-0.691</td>
<td>0.994</td>
</tr>
<tr>
<td>3y</td>
<td>5.100</td>
<td>1.417</td>
<td>-0.421</td>
<td>-0.651</td>
<td>0.993</td>
</tr>
<tr>
<td>4y</td>
<td>5.265</td>
<td>1.301</td>
<td>-0.297</td>
<td>-0.686</td>
<td>0.991</td>
</tr>
<tr>
<td>5y</td>
<td>5.397</td>
<td>1.213</td>
<td>-0.187</td>
<td>-0.751</td>
<td>0.990</td>
</tr>
<tr>
<td>7y</td>
<td>5.594</td>
<td>1.096</td>
<td>-0.015</td>
<td>-0.873</td>
<td>0.989</td>
</tr>
<tr>
<td>10y</td>
<td>5.793</td>
<td>1.003</td>
<td>0.136</td>
<td>-0.994</td>
<td>0.987</td>
</tr>
</tbody>
</table>

Average weekly autocorrelation for swap rates ($\phi$) is 0.991:

Half-life $= \ln \phi / 2 / \ln \phi \approx 78$ weeks (1.5 years)

*Interest rates are highly persistent; forecasting is difficult.*
Compute the covariance ($\Omega$) of the weekly changes in LIBOR/swap rates.

Perform eigenvalue/eigenvector decomposition on the covariance matrix. $[V,D] = \text{eigs}(\Omega)$.

The eigenvalues (after normalization) represent the explained variation.

The eigenvector defines the loading of each principal component.

Example: We can form a mimicking portfolio of the first principal component by $P_1 = V(:,1)\top R$, where $R$ denotes the $(12 \times 1)$ vector.
Model design: Three-factor affine DTSMs

- Affine specifications:
  - Risk-neutral ($Q$) factor dynamics:
    \[ dX_t = \kappa (\theta - X_t) \, dt + \sqrt{\Omega_t} \, dW_t^Q, \quad [\Omega_t]_{ii} = \alpha_i + \beta_i^\top X_t. \]
  - Short rate function: \( r(X_t) = a_r + b_r^\top X_t \)

- Bond pricing: Zero-coupon bond prices:
  \[ B(X_t, \tau) = \exp \left( -a(\tau) - b(\tau)^\top X_t \right). \]

- Affine forecasting dynamics ($P$):
  \[ dX_t = \kappa^P (\theta^P - X_t) \, dt + \sqrt{\Omega_t} \, dW_t^P \]
  - Does not matter for bond pricing.
  - Specification is up to identification.

- Dai, Singleton (2000): \( A_m(3) \) classification with \( m = 0, 1, 2, 3 \).

\[ [\Omega_t]_{ii} = \begin{cases} 
X_t, & i = 1, \ldots, m; \\
1 + \beta_i^\top X_t, & i = m + 1, \ldots, n.
\end{cases} \]
Model design: A 3-factor Gaussian affine model

The simplest among all $A_m(3)$ models:

- The specification and normalization:
  \[
  dX_t = -\kappa^P X_t dt + dW_t^P \\
  dX_t = -\kappa (\theta - X_t) dt + dW_t^Q, \quad r(X_t) = a_r + b_r^T X_t.
  \]

  - The instantaneous covariance matrix is normalized to $I$.
  - The long-run statistical mean is normalized to zero.

- Zero-coupon bond prices: $B(X_t, \tau) = \exp \left( -a(\tau) - b(\tau)^T X_t \right)$, with $a'(\tau) = a_r + b(\tau)\kappa \theta - \frac{1}{2} b(\tau)^T b(\tau)$ and $b'(\tau) = b_r - \kappa^T b(\tau)$.
  - The coefficients can be solved analytically: $b(\tau) = e^{-\kappa^T \tau} b_r$.

- Properties: Bond yields (spot, forward rates) are affine functions of Gaussian variables.
Estimation: Maximum likelihood with UKF

- State propagation (discretization of the forecasting dynamics):
  \[ X_{t+1} = A + \Phi X_t + \sqrt{Q} \varepsilon_{t+1}. \]
  \[ A = 0, \quad \Phi = e^{-\kappa P \Delta t}, \quad Q = \int_0^{\Delta t} e^{-\kappa s} e^{-\kappa^\top s} ds = \Delta t. \]

- Measurement equation: Assume LIBOR and swap rates are observed with error:
  \[ y_t = \begin{bmatrix} \text{LIBOR}(X_t, i) \\ \text{SWAP}(X_t, j) \end{bmatrix} + \sqrt{\Sigma} e_t, \quad i = 1, 2, 3, 6, 12 \text{ months} \quad j = 2, 3, 5, 7, 10, 15, 30 \text{ years}. \]

- Unscented Kalman Filter (UKF) generates conditional forecasts of the mean and covariance of the state vector and observations.

- Likelihood is built on the forecasting errors:
  \[ l_{t+1} = -\frac{1}{2} \log |V_{t+1}| - \frac{1}{2} \left( (y_{t+1} - \bar{y}_{t+1})^\top (V_{t+1})^{-1} (y_{t+1} - \bar{y}_{t+1}) \right). \]

- Choose model parameters \((\kappa^P, \kappa, \theta, a_r, b_r, \Sigma)\) to maximize the sum of the weekly log likelihood function.
The Classic Kalman filter

- Kalman filter (KF) generates efficient forecasts and updates under linear-Gaussian state-space setup:
  
  \[
  \begin{align*}
  \text{State} : & \quad X_{t+1} = A + \Phi X_t + \sqrt{Q}\varepsilon_{t+1}, \\
  \text{Measurement} : & \quad y_t = HX_t + \sqrt{\Sigma}e_t
  \end{align*}
  \]

- The ex ante predictions as
  \[
  \begin{align*}
  \bar{X}_t & = A + \Phi \hat{X}_{t-1}; \\
  \bar{\Omega}_t & = \Phi \hat{\Omega}_{t-1} \Phi^\top + Q; \\
  \bar{y}_t & = H\bar{X}_t; \\
  \bar{V}_t & = H\bar{V}_t H^\top + \Sigma.
  \end{align*}
  \]

- The ex post filtering updates are,
  \[
  \begin{align*}
  \hat{X}_{t+1} & = \bar{X}_{t+1} + K_{t+1} (y_{t+1} - \bar{y}_{t+1}); \\
  \hat{\Omega}_{t+1} & = \bar{\Omega}_{t+1} - K_{t+1} \bar{V}_{t+1} K_{t+1}^\top,
  \end{align*}
  \]
  where \( K_{t+1} = \bar{\Omega}_{t+1} H^\top (\bar{V}_{t+1})^{-1} \) is the Kalman gain.
In our application, LIBOR and swap rates are NOT linear in states:

State: \[ X_{t+1} = A + \Phi X_t + \sqrt{Q}\varepsilon_{t+1}, \]
Measurement: \[ y_t = h(X_t) + \sqrt{\Sigma} e_t \]

One way to use the Kalman filter is by linear approximating the measurement equation,

\[ y_t \approx H_t X_t + \sqrt{\Sigma} e_t, \quad H_t = \left. \frac{\partial h(X_t)}{\partial X_t} \right|_{X_t = \hat{X}_t} \]

It works well when the nonlinearity in the measurement equation is small.

Numerical issue
- How to compute the gradient?
- The covariance matrix can become negative.
Approximating the distribution

Measurement: \[ y_t = h(X_t) + \sqrt{\Sigma} e_t \]

- The Kalman filter applies Bayesian rules in updating the conditionally normal distributions.
- Instead of approximating the measurement equation \( h(X_t) \), we directly approximate the distribution and then apply Bayesian rules on the approximate distribution.
- There are two ways of approximating the distribution:
  - Draw a large amount of random numbers, and propagate these random numbers — Particle filter. (more generic)
  - Choose “sigma” points deterministically to approximate the distribution (think of binominal tree approximating a normal distribution) — unscented filter. (faster, easier to implement)
The unscented Kalman filter

- Let $k = 3$ be the number of states and $\delta > 0$ be a control parameter. A set of $2k + 1$ sigma vectors $\chi_i$ are generated according to:

$$
\chi_{t,0} = \hat{X}_t, \quad \chi_{t,i} = \hat{X}_t \pm \sqrt{(k + \delta)(\hat{\Omega}_t + Q)}_j
$$

(2)

with corresponding weights $w_i$ given by:

$$
\begin{align*}
    w_0 & = \delta / (k + \delta), \\
    w_i & = 1 / [2(k + \delta)].
\end{align*}
$$

- We can regard these sigma vectors as forming a discrete distribution with $w_i$ as the corresponding probabilities.

- We can verify that the mean, covariance, skewness, and kurtosis of this distribution are $\hat{X}_t$, $\hat{\Omega}_t + Q$, 0, and $k + \delta$, respectively.
The unscented Kalman filter

- Given the sigma points, the prediction steps are given by

\[
\begin{align*}
\overline{X}_{t+1} &= A + \sum_{i=0}^{2k} w_i (\Phi \chi_{t,i}); \\
\overline{\Omega}_{t+1} &= \sum_{i=0}^{2k} w_i (A + \Phi \chi_{t,i} - \overline{X}_{t+1})(A + \Phi \chi_{t,i} - \overline{X}_{t+1})^\top; \\
\overline{y}_{t+1} &= \sum_{i=0}^{2k} w_i h (A + \Phi \chi_{t,i}); \\
\overline{V}_{t+1} &= \sum_{i=0}^{2k} w_i \left[ h (A + \Phi \chi_{t,i}) - \overline{y}_{t+1} \right] \left[ h (A + \Phi \chi_{t,i}) - \overline{y}_{t+1} \right]^\top + \Sigma,
\end{align*}
\]

- The filtering updates are given by

\[
\begin{align*}
\hat{X}_{t+1} &= \overline{X}_{t+1} + K_{t+1} (y_{t+1} - \overline{y}_{t+1}); \\
\hat{\Omega}_{t+1} &= \overline{\Omega}_{t+1} - K_{t+1} \overline{V}_{t+1} K_{t+1}^\top,
\end{align*}
\]
Alternative estimation approaches

- Instead of estimating the model on LIBOR and swap rates, we can estimate the model on stripped continuously compounded spot or forward rates, which are affine in the state vector.
  - Pros: We can use Kalman filter, with no approximation.
  - Cons: Which spot maturities to choose? The measurement errors will depend on the stripping method.
  - For our trading strategy, we actually need to know the pricing error on tradable securities.

- Exact matching: Assume that three of the LIBOR swap rates (or three spot rates) are priced exactly by the 3 factors. Then, we can directly invert the state variables at each date exactly.
  - The likelihood function can be built on the state variables directly, maybe with additional likelihood on pricing errors on other series.
  - Which rate series are priced fair? Which rates have error?
  - Kalman filter amounts to a least square fitting to all rates.
Static v. dynamic consistency

- In many applications, banks choose to use a simpler model (say, one factor interest rate model), but recalibrate the model parameters each day to the market data.
- The “same model” with different parameters are essentially different models.
- When a model is re-calibrated daily, the model parameters also become state variables.
- At a fixed point in time, the model-generated prices satisfy no-arbitrage conditions. They are internally consistent.
- But model-generated prices at different days with different re-calibrated parameters are no longer consistency with each other.
- Static consistency (across different securities) could be sufficient for short-term investors and market makers who do not hold overnight positions.
- Dynamic consistency (over time) becomes important when a firm
Estimated factor dynamics

<table>
<thead>
<tr>
<th>Forecasting dynamics $\kappa^P$</th>
<th>Risk-neutral dynamics $\kappa^*$</th>
</tr>
</thead>
</table>
| \[
\begin{bmatrix}
0.002 & 0 & 0 \\
(0.02) & -- & -- \\
-0.186 & 0.480 & 0 \\
(0.42) & (1.19) & -- \\
-0.749 & -2.628 & 0.586 \\
(1.80) & (3.40) & (2.55)
\end{bmatrix}
| \[
\begin{bmatrix}
0.014 & 0 & 0 \\
(11.6) & -- & -- \\
0.068 & 0.707 & 0 \\
(1.92) & (20.0) & -- \\
-2.418 & -3.544 & 1.110 \\
(10.7) & (12.0) & (20.0)
\end{bmatrix}
|

- The $t$-values are smaller for $\kappa^P$ than for $\kappa$.
- The largest eigenvalue of $\kappa^P$ is 0.586
  $\Rightarrow$ Weekly autocorrelation 0.989, half life 62 weeks.
Summary statistics of the pricing errors (bps)

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>MAE</th>
<th>Std</th>
<th>Max</th>
<th>Auto</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 m</td>
<td>1.82</td>
<td>6.89</td>
<td>10.53</td>
<td>60.50</td>
<td>0.80</td>
<td>99.65</td>
</tr>
<tr>
<td>3 m</td>
<td>0.35</td>
<td>1.87</td>
<td>3.70</td>
<td>31.96</td>
<td>0.73</td>
<td>99.96</td>
</tr>
<tr>
<td>12 m</td>
<td>-9.79</td>
<td>10.91</td>
<td>10.22</td>
<td>55.12</td>
<td>0.79</td>
<td>99.70</td>
</tr>
<tr>
<td>2 y</td>
<td>-0.89</td>
<td>2.93</td>
<td>4.16</td>
<td>23.03</td>
<td>0.87</td>
<td>99.94</td>
</tr>
<tr>
<td>5 y</td>
<td>0.20</td>
<td>1.30</td>
<td>1.80</td>
<td>10.12</td>
<td>0.56</td>
<td>99.98</td>
</tr>
<tr>
<td>10 y</td>
<td>0.07</td>
<td>2.42</td>
<td>3.12</td>
<td>12.34</td>
<td>0.70</td>
<td>99.91</td>
</tr>
<tr>
<td>15 y</td>
<td>2.16</td>
<td>5.79</td>
<td>7.07</td>
<td>22.29</td>
<td>0.85</td>
<td>99.40</td>
</tr>
<tr>
<td>30 y</td>
<td>-0.53</td>
<td>8.74</td>
<td>11.07</td>
<td>34.58</td>
<td>0.90</td>
<td>98.31</td>
</tr>
<tr>
<td>Average</td>
<td>-0.79</td>
<td>4.29</td>
<td>5.48</td>
<td>27.06</td>
<td><strong>0.69</strong></td>
<td>99.71</td>
</tr>
</tbody>
</table>

- The errors are small. The 3 factors explain over 99%.
- The average persistence of the pricing errors (0.69, half life 3 weeks) is much smaller than that of the interest rates (0.991, 1.5 years).
4-week ahead forecasting

Three strategies:
(1) random walk (RW); (2) AR(1) regression (OLS); (3) DTSM.

Explained Variation = 100 × [1 − \(\text{var(Err)/var(ΔR)}\)]

<table>
<thead>
<tr>
<th>Maturity</th>
<th>RW</th>
<th>OLS</th>
<th>DTSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 m</td>
<td>0.00</td>
<td>0.53</td>
<td>-31.71</td>
</tr>
<tr>
<td>2 y</td>
<td>0.00</td>
<td>0.02</td>
<td>-7.87</td>
</tr>
<tr>
<td>3 y</td>
<td>0.00</td>
<td>0.13</td>
<td>-0.88</td>
</tr>
<tr>
<td>5 y</td>
<td>0.00</td>
<td>0.44</td>
<td>0.81</td>
</tr>
<tr>
<td>10 y</td>
<td>0.00</td>
<td>1.07</td>
<td>-3.87</td>
</tr>
<tr>
<td>30 y</td>
<td>0.00</td>
<td>1.53</td>
<td>-36.64</td>
</tr>
</tbody>
</table>

- OLS is not that much better than RW, due to high persistence (max 1.5%).
- **DTSM is the worst!** DTSM can be used to fit the term structure (99%), but not forecast interest rates.
Use DTSM as a decomposition tool

- Linearly decompose the LIBOR/swap rates \((y)\) as
  \[ y_t^i \approx H_i^\top X_t + e_t^i, \quad H_i = \left. \frac{\partial y_t^i}{\partial X_t} \right|_{X_t=0} \]

- Form a portfolio \((m = [m_1, m_2, m_3, m_4]^\top)\) of 4 LIBOR/swap rates so that
  \[ p_t = \sum_{i=1}^{4} m_i y_t^i \approx \sum_{i=1}^{4} m_i H_i^\top X_t + \sum_{i=1}^{4} m_i e_t^i = \sum_{i=1}^{4} m_i e_t^i. \]

- Choose the portfolio weights to hedge away its dependence on the three factors: \(Hm = 0\).
  - 2 rates are needed to cancel out their dependence on 1 factor.
  - 3 rates are needed to cancel out their dependence on 2 factors.
  - 4 rates are needed to cancel out their dependence on 3 factors.
Example: A 4-rate portfolio (2-5-10-30)

Portfolio weights: \( m = [0.0277, -0.4276, 1.0000, -0.6388] \).

Long 10-yr swap, use 2, 5, and 30-yr swaps to hedge.

Hedged 10-yr swap

\( \phi \) (half life): 0.816 (one month) vs. 0.987 (one year).

\[
\Delta R_{t+1} = -0.0849 - 0.2754 R_t + e_{t+1}, \quad R^2 = 0.14, \\
(0.0096) \quad (0.0306)
\]

\( R^2 = 1.07\% \) for the unhedged 10-year swap rate.
12 rates can generate 495 4-instrument portfolios.

Improved predictability for all portfolios (against unhedged single rates)
No guaranteed success for spread (2-rate) and butterfly (3-rate) portfolios.

Predictability improves dramatically after the 3rd factor.
A simple buy and hold investment strategy on interest-rate portfolios

- Form 4-instrument swap portfolios ($m$). Regard each swap contract as a par bond.

- At each time, long the portfolio if the portfolio swap rate is higher than the model value. Short otherwise:

\[ w_t = c \left[ m^\top (y_t - \text{SWAP}(X_t)) \right] \]

- Hold each investment for 4 weeks and liquidate.
  - Need interpolation to obtain swap quotes at off-grid maturities.
  - For discounting, one can use piece-wise constant forward to strip the yield curve.

- Remark: The (over-simplified) strategy is for illustration only; it is not an optimized strategy.
Consider a 4-instrument swap portfolio with fixed portfolio weights $m_i$.

At time $t$, the investment in this portfolio is $w_t$ (based on whether this portfolio value is higher or lower than the model value).

The initial cost of this investment is $C_t = 100w_t \sum_{i=1}^{4} m_i$, regarding each swap contract as a par bond with $100$ par value.

At the end of the holding horizon $T$ (4 weeks later), liquidate the portfolio. The revenue from the liquation is:

$$P_T = w_t \sum_{i=1}^{4} m_i \left( S_t^i \sum_{t_i} D(T, t_i) + 100D(T, T_i) \right)$$

- $D(T, t_i)$ — the time-$T$ discount factor at expiry date $t_i$.
- $t_i$ — the coupon dates of the $i$th bond with coupon rate being the time-$t$ swap rate $S_t^i$.

$$P&L = P_T - C_t e^{r_t(T-t)}$$, with $r_t$ denoting the continuously compounded rate corresponding to the time-$t$ three-month LIBOR, the financing rate of the swap contract.
Profitability of investing in four-instrument swap portfolios

Cumulative Wealth

Annualized Information Ratio

Wu (Baruch) Statistical Arbitrage
The sources of the profitability

- Risk and return characteristics
  - The investment returns are not related to traditional stock and bond market factors (the usual suspects): Rm, HML, SMB, UMD, Credit spread, interest rate volatility,...
  - But are positively related to some swap market liquidity measures.

- Interpretation
  - The first 3 factors relate to systematic economic movements: Inflation rate, output gap, monetary policy, ...
  - What is left is mainly due to short-term liquidity shocks.
  - By providing liquidity to the market, one can earn economically significant returns.
A mean-variance setup

- Let $y_t \in \mathbb{R}^N$ denote a vector of observed derivative prices, let $h(X_t)$ denote the model-implied value as a function of the state vector $X_t \in \mathbb{R}^k$. Let $H_t = \frac{\partial h(X_t)}{\partial X_t} \in \mathbb{R}^{N \times k}$ denote the gradient matrix at time $t$.
- $-e_t = h(X_t) - y_t$ can be regarded as the “alpha” of the asset and $H_t$ its risk exposure.
- We can solve the following quadratic program:

$$\max_{w_t} -w_t^T e_t - \frac{1}{2} \gamma w_t^T \Sigma e w_t$$

subject to factor exposure constraints:

$$H_t^T w_t = c \in \mathbb{R}^k.$$  

- The above equation maximizes the expected return (alpha) of the portfolio subject to factor exposure constraints.
- Setting $c = 0$ maintains (first order) factor-neutrality.
Updated out-of-sample cumulative P&L on US swap rates
After thoughts: Implications for modeling

- Models based on a low-dimensional state vector is important in smoothing and achieving dimension reduction.

- Wherever there is a new instrument, there are new supply-demand shocks.

- Modeling/understanding these supply-demand shocks can be economically important when:
  - making short-term investments / market making,
  - pricing interest-rate derivatives / unspanned volatility,
  - understanding the covariation of different interest-rate series.
No-arbitrage models provide *relative* valuation across assets, and hence can be best used for cross-sectional comparison.

No-arbitrage theory does not tell us how to predict the factors, but it does tell us how each instrument is related to the factor risk (factor loading).

⇒ It is the most useful for hedging:

- *Hedge away the risk, exploit the opportunity.*