Modeling Yield Curve Dynamics

Liuren Wu

Zicklin School, Baruch College, CUNY
1. A general framework

2. Bond pricing: the general setting

3. Dynamic term structure models

4. Model design

5. Model estimation

6. Using analysts forecasts as measurements
Overview of questions

- Here is a list of questions for which you want answers:
  - How to forecast the yield curve (LIBOR, swap, Treasury)? How to link the forecasts to economic cycle?
  - What are the factors that drive the yield curve? What are the factors that drive the spread between curves?
  - How to perform stress analysis on the curve?
  - How Fed policies (Fed Fund Target, QE) affect the yield curve?
  - How do your analytics compare to Bloomberg, bluechip?

- It is hopeless to try to provide an answer to each question, or to think that all you need is an answer to a simple question.
  
  *Give a man a fish and you feed him for a day; teach a man to fish and you feed him for a lifetime.* —Maimonedes

- Instead, I will try to explain to you a general framework, with which you can analyze all these questions.
  - Once you have learned the right approach to analyze the questions, you can come up with your own answers.
A general framework: Analyze rates behavior via DTSMs

- Within each class (e.g., Treasury, LIBOR/swap, OIS), interest rates vary across maturity.
  - This is what we normally refer to as the \textit{yield curve}.

- \textit{Dynamic term structure models} (DTSM) specify the short rate dynamics, and derive the value for the whole yield curve.

- By constructing and estimating the appropriate DTSM, one can consolidate information sources (e.g., rates, occasional bluechip (and other) rate/macro forecasts) to
  - generate frequent prediction updates on the whole curve and across all forecasting horizons
  - Generate stress scenarios
  - Link to macroeconomic activities and policies

- By modeling the spreads across different classes (due to credit, liquidity), one can carry the yield curve prediction (and risk analysis) from one class to another.
The analytics behind DTSMs

- Not only to understand particular models in the literature, but also to construct a whole classes of models of your desire.

Design DTSMs to match findings from statistical data analysis and structural economic analysis

Estimate DTSMs with different sources/types of informations

Applications

- Yield curve prediction
- Scenario analysis
- Policy linkages
1. A general framework

2. Bond pricing: the general setting

3. Dynamic term structure models

4. Model design

5. Model estimation

6. Using analysts forecasts as measurements
The starting point: The concept of a *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 value process associated with any trading strategy is a martingale:

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

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

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

- $M_t$ 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.

- *Bond valuation* reduces to taking expected value of the pricing kernel:

$$B(t, T) = \mathbb{E}^P_t [M_{t,T}].$$
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,

\[ M_{t,t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)} = \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma}, \quad \gamma > 0 - \text{relative risk aversion} \]

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

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

\( X \) denotes the risk sources in the economy, \( r \) is the instantaneous riskfree rate (determined as a function of \( X \)), and \( \gamma \) is the market price of the risk \( X \).

  - \( \mathcal{E} \) is the stochastic exponential martingale operator.
  - In a continuous time version of the representative agent example, \( dX_s = d\ln c_t \) and \( \gamma \) is relative risk aversion.
Bond pricing

- Given the multiplicative decomposition of the pricing kernel:

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

- The value of the zero-coupon bond can be written as

\[ B(t, T) = \mathbb{E}_{t}^{\mathbb{P}} [ M_{t,T} ] \]

\[ = \mathbb{E}_{t}^{\mathbb{P}} \left[ \exp \left( - \int_t^T r(X_s) ds \right) \mathcal{E} \left( - \int_t^T \gamma_s^T dX_s \right) \right] \]

\[ = \mathbb{E}_{t}^{\mathbb{Q}} \left[ \exp \left( - \int_t^T r(X_s) ds \right) \right] \]

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

- Different researchers have different starting points:
  - Economists start with utility function \( u(c) \)
  - Financial economists start with a pricing kernel \( M_{t,t+1} \)
  - Finance professors start with statistical short-rate \( r_t \) dynamics and market price of risk \( \gamma_t \) specification
  - Math finance professors start with the risk-neutral short-rate dynamics
The yield curve

- The yield on the zero-coupon bond is defined as

\[
y_t(\tau) \equiv -\frac{1}{T-t} \ln B(t, T), \quad \tau \equiv T - t
\]

The **yield curve** refers to the yield pattern across time to maturity \(\tau = T - t\) at a certain date.

- The yield can be decomposed generically into three components:

\[
y_t(\tau) = \frac{1}{\tau} \mathbb{E}_t^P \left[ \int_t^T r_u du \right] \quad \text{(Expectation)}
\]
\[
+ \frac{1}{\tau} \mathbb{E}_t^P \left[ \left( \frac{dQ}{dP} - 1 \right) \int_t^T r_u du \right] \quad \text{(Risk premium)}
\]
\[
- \frac{1}{\tau} \ln \mathbb{E}_t^Q \left[ \exp \left( -\left( \int_t^T r_u du - \mathbb{E}_t^Q \int_t^T r_u du \right) \right) \right] \quad \text{(Convexity)}
\]

- **Relative to expectation, risk premium tends to raise the yield curve whereas convexity lowers it.**
Yield curve decomposition

\[ y_t(\tau) = \frac{1}{\tau} \mathbb{E}_t \left[ \int_t^T r_u du \right] \]  
\[ + \frac{1}{\tau} \mathbb{E}_t \left[ \left( \frac{dQ}{dP} - 1 \right) \int_t^T r_u du \right] \]  
\[ - \frac{1}{\tau} \ln \mathbb{E}_t^Q \left[ \exp \left( -\left( \int_t^T r_u du - \mathbb{E}_t^Q \int_t^T r_u du \right) \right) \right] \]

(Expectation)  
(Risk premium)  
(Convexity)

1. The 1st term captures investor *expectation* of average future short rates \( \frac{1}{T-t} \int_t^T r_u du \).

2. The 2nd term measures the *covariance* between \( \frac{dQ}{dP} \) and the average future short rate, capturing the risk premium on the interest rate risk.

3. The 3rd term measures convexity induced by short rate *variation*.
   - If \( \int_t^T r_u du \) is normally distributed with variance \( V \), the log of the expectation in the convexity term is simply half of the variance \( -\frac{1}{2} V \).
1. A general framework
2. Bond pricing: the general setting
3. Dynamic term structure models
4. Model design
5. Model estimation
6. Using analysts forecasts as measurements
Dynamic term structure models

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

- **Classic 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**
  - 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:
  - Make assumptions on factor dynamics ($X$), market prices ($\gamma$), and how interest rates are related to the factors $r(X)$
  - Derive the fair valuation of bonds based on these dynamics and market price specifications.

- The back-filling (reverse engineering) procedure:
  - State the form of solution that one desires for bond prices.
  - Figure out what dynamics specifications generate the desired solutions.

- 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-$t$ value of a zero-coupon bond with maturity $\tau = T - t$, we have

$$B(X_t, \tau) = \exp \left( -a(\tau) - b(\tau)^\top X_t \right)$$

(2)

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.

Questions to be answered:
- What is the short rate function $r(X_t)$?
- What’s the dynamics of $X_t$ under measure $\mathbb{Q}$?

that give us the desired solution form in (2).
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,

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

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

  \[
  f_t + \mathcal{L}f = rf,
  \]

  \( \mathcal{L}f \) — infinitesimal generator

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

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

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

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

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

- Set \( \mu(X) = a_m + b_m X + c_m XX^\top + \cdots \) and
  \[ [\sigma(X)\sigma(X)^\top]_i = \alpha_i + \beta_i^\top X + \eta_i XX^\top + \cdots, \]
  and collect terms:

  constant
  \[ a'(\tau) - b(\tau)^\top a_m + \frac{1}{2} \sum b(\tau)b(\tau)^\top \cdot \alpha_i = a_r \]
  \[ b'(\tau)^\top - b(\tau)^\top b_m + \frac{1}{2} \sum b(\tau)b(\tau)^\top \cdot \beta_i^\top = b_r^\top \]
  \[ XX^\top - b(\tau)^\top c_m + \frac{1}{2} \sum b(\tau)b(\tau)^\top \cdot \eta_i = 0 \]

- The quadratic and higher-order terms are almost surely zero.
- We thus have the conditions to have exponential-affine bond prices:

  \[ \mu(X) = a_m + b_m X, \quad [\sigma(X)\sigma(X)^\top]_i = \alpha_i + \beta_i^\top X, \quad r(X) = a_r + b_r^\top X. \]

- We can solve the coefficients \([a(\tau), b(\tau)]\) via the following ODEs:

  \[ a'(\tau) = a_r + b(\tau)^\top a_m - \frac{1}{2} \sum b(\tau)b(\tau)^\top \cdot \alpha_i \]
  \[ b'(\tau) = b_r + b_m^\top b(\tau) - \frac{1}{2} \sum b(\tau)b(\tau)^\top \cdot \beta_i \]

starting at \( a(0) = 0 \) and \( b(0) = 0 \).
From $Q$ to $P$, not from $P$ to $Q$

- The affine conditions,

\[ \mu(X) = a_m + b_m X, \quad \left[ \sigma(X)\sigma(X)^\top \right]_i = \alpha_i + \beta_i^\top X, \quad r(X) = a_r + b_r^\top X. \]

are about the $Q$-dynamics, not about the $P$-dynamics.

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

- The back-filling exercise shows that the real requirement for tractability (e.g., affine) is on $Q$-dynamics.

- Hence, we should also back fill the specification:
  - Start with $Q$-dynamics specification to generate tractable derivative pricing solutions.
  - Fill in any market price of risk specification that one desires.
  - It does not matter if the $P$-dynamics is complex/tractable or not.
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)

- Statistical: Principal component analysis can be useful.
- Structural: Economic knowledge can be embedded in model design.

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.
Example: Data analysis

- Take US dollar LIBOR and swap rates as example, 1995.1.3-2016.5.11, 5378 business days
- Swaps are based on 6-month LIBOR, maturities from 2 to 30 years
- Quoting conventions: actual/360 for LIBOR; 30/360 with semi-annual payment for swaps.

\[
LIBOR(X_t, \tau) = \frac{100}{\tau} \left( \frac{1}{B(X_t, \tau)} - 1 \right), \text{(simple compounding rates)}
\]
\[
SWAP(X_t, \tau) = 200 \times \frac{1 - B(X_t, \tau)}{\sum_{i=1}^{2\tau} B(X_t, i/2)}, \text{(par bond coupon rates)}.
\]
Data: Time series and term structure

- Time series of selected series on the left.
- Term structure shapes on the right, with the solid line being the average, and the two dash-dotted lines being the one-standard deviation band.
Principal component analysis

- Compute the covariance matrix (\( \Omega \)) of the daily changes in swap rates.
- Perform eigenvalue/eigenvector decomposition on the covariance matrix. \([V,D]=\text{eigs}(\Omega)\).
- The eigenvalues (after normalization) represents the percentage explained variation.
- The eigenvector defines the loading on each rate to mimic the corresponding principal component.
The first principal component explains over 90% of the daily variation. It has positive loadings on all rates, representing the “level” factor.

The second principal component explains about 5% of the variation. It has negative loadings on short-term rates, positive loadings on long-term rates, generating a “slope” effect.

The third principal component captures the “curvature” movement of the yield curve.
Model design: Three-factor affine DTSMs

- Affine specifications:
  - Risk-neutral ($Q$) factor dynamics:
    \[
    dX_t = \kappa (\theta - X_t) \, dt + \sqrt{\Omega_t} dW^Q_t,
    \]
    \[
    [\Omega_t]_{ii} = \alpha_i + \beta_i^\top X_t.
    \]
  - 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 ($P$):
  \[
  dX_t = \kappa^P (\theta^P - X_t) \, dt + \sqrt{\Omega_t} dW^P_t
  \]
  - Does not matter for bond pricing.
  - Specification is up to identification.

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

\[
[\Omega_t]_{ii} = \begin{cases} 
X_{t,i} & i = 1, \ldots, m; \\
1 + \beta_i^\top X_t & i = m + 1, \ldots, n.
\end{cases}
\]
The simplest among all $A_m(3)$ models:

- The specification and normalization:

$$dX_t = -\kappa^P X_t dt + dW^P_t$$

$$\gamma(X_t) = \gamma_0 + \gamma_1 X_t \Rightarrow dX_t = \kappa (\theta - X_t) dt + dW^Q_t,$$

$$r(X_t) = a_r + b_r^\top X_t.$$

- $\kappa\theta = -\gamma_0$, $\kappa = \kappa^P + \gamma_1$
- $\kappa^P$, $\kappa$ are constrained to be lower triangular matrix, with positive eigenvalues.

- Zero-coupon bond prices: $B(X_t, \tau) = \exp \left( -a(\tau) - b(\tau)^\top X_t \right)$, with

$a'(\tau) = a_r + b(\tau)\kappa\theta - \frac{1}{2} b(\tau)^\top b(\tau)$ and $b'(\tau) = b_r - \kappa^\top b(\tau)$.

- The coefficients can be solved analytically: $b(\tau) = e^{-\kappa^\top \tau} b_r$.

- Properties: Bond yields (spot, forward rates) are affine functions of Gaussian variables.

Examples: Bali, Heidari, Wu (2009), Kim&Wright (2005)
Cochrane & Piazzesi (2004) find that bond excess returns are all proportional to a single risk factor, formed by a portfolio of forward rates, with tent-shaped weights.

We can propose a model (GA3FCP) capturing this feature:

- The market price of all risk factors are proportional to the same combination of states:

\[ \gamma_{j,t} = \gamma_{0,j} + (\gamma_1 x_{1t} + \gamma_2 x_{2t} + \gamma_3 x_{3t}) s_j, \quad \text{for } j = 1, 2, 3 \]

- Even though \( \kappa^P \) is lower triangular, \( \kappa = \kappa^P + \gamma s^\top \) is not.

- The model has one fewer parameter, but performs better. The identified risk premium indeed has a tent shape.
We also consider a more parsimonious specification by imposing a cascade structure on the state dynamics (CA3FCP):

\[
\begin{align*}
    dX_{3,t} &= \kappa_3 (X_{2,t} - X_{3,t}) \, dt + dW_{3,t}, \\
    dX_{2,t} &= \kappa_2 (X_{1,t} - X_{2,t}) \, dt + dW_{2,t}, \\
    dX_{1,t} &= \kappa_1 (0 - X_{1,t}) \, dt + dW_{1,t} \\
    r_t &= \theta_r + \sigma_r X_{3,t}
\end{align*}
\]

The model has 5 fewer parameters than GA3FCP and can thus be better identified while generating similar performance.

- The short rate mean reverts to a middle rate \(X_2\), which mean reverts to a long rate \(X_1\), which mean reverts to a long run mean \(\theta_r\).
- The cascade structure can be used to add as many factors as performance needs, but always with only 5 parameters via a power scaling specification: Calvet, Fisher, Wu (JFQA, 2018).
We can treat the short rate \( r \) as the policy rate and link it to economic variables as in a Taylor rule (1993):

\[
    r_t = \theta r + b_1 (\pi_t - \bar{\pi}_t) + b_2 x_t + s_t
\]

- \( \pi_t - \bar{\pi}_t \) — deviation of inflation expectation from target
- \( x_t \) — output gap
- \( s_t \) — policy surprise (inertia)

The derived term structure would be a function of the inflation/output gap dynamics and their market prices of risk.

The effects of various macroeconomic announcements on the yield curve can be analyzed within this framework by further assuming that these announcements are noisy observations of the inflation/output:

\[
    m_t = H[\pi_t, x_t] + \varepsilon_t
\]

Example: Lu & Wu (JME, 2009), Wu & Zhang (Management Science, 2008)
Summary: Bond pricing under affine structures

- When the dynamics \( \mu(X_t) = \kappa^P(\theta^P - X_t), \Sigma(X_t) \), market prices \( \gamma(X_t) \) and short rate function \( r(X_t) \) are all affine functions of the state \( X_t \), model values for zero-coupon bonds are exponential affine in the states:

\[
P(X_t, \tau) = \exp(-A(\tau) - B(\tau)^T X_t).
\]

- Continuously compounded zero rates are affine in the states:

\[
z(X_t, \tau) = a(\tau) + b(\tau)^T X_t,
\]

with \( a(\tau) = A(\tau)/\tau \) and \( b(\tau) = B(\tau)/\tau \).

- Forecasts on zero rates are also affine in states:

\[
\mathbb{E}_t[z(X_{t+h}, \tau)] = a(\tau) + b(\tau)^T \mathbb{E}_t[X_{t+h}]
= a(\tau) + b(\tau)^T (I - e^{-\kappa^P h})\theta + b(\tau)^T e^{-\kappa^P h}X_t.
\]

- One can consider various model designs within this tractable class, for different purposes.
A general framework

Bond pricing: the general setting

Dynamic term structure models

Model design

Model estimation

Using analysts forecasts as measurements
Purpose of estimating a dynamic term structure model

1. Determine the model parameters that govern the dynamics, risk premium, and short rate response function.
   - Gain understanding on dynamics, risk premium, monetary policy.
   - Generate risk sensitivities and simulate risk/stress scenarios.

2. Extract the state variables that describe the economy, from the observed prices/rates/forecasts.
   - Model provides a way to reduce dimension by summarizing many observations with a few states.
   - Well-specified states can be treated as economic signals.
   - Global macro trades can be based on identified risk premium variations with economic factors.

3. Identify relative value for statistical arbitrage trading
   - When market prices deviate from model valuation, market prices tend to revert back to model in the future (Bali, Heidari, Wu, 2009).

4. Interpolation/extrapolation:
   - Provide a consistent, intuitive, and parsimonious structural form for stripping curves (Calvet, Fisher, Wu, 2018).
The state space setting includes a pair of state propagation equation and measurement equation:

1. **State propagation** can be built on a discretization of the $\mathbb{P}$ state dynamics,
   - In Euler approximation, $\hat{f}(X_t) = X_t + \mu(X_t)\Delta t$ and $\Sigma_x = \Sigma(X_t)\Delta t$.
   - One can be exact in certain simple cases (e.g., Gaussian linear).
   - The time step $\Delta t$ can be fixed or vary over time.

2. **Measurement equations** can be built on observed prices/rates/forecasts: $y_t = h(X_t) + \sqrt{\Sigma_y} e_t$, where $e_t$ denoting an additive observation error.
   - The number of available observations can vary over time.
   - $\Sigma_y$ determines the relative accuracy of the observation.

One can regard state as “signals” and measurements as noisy observations.

- State propagation describes the trajectory of the signal movement (direction, magnitude of variation).
- Measurement equation describes the relative accuracy (signal/noise ratio) of each observation.
The classic Kalman filter for Gaussian-linear cases

- Kalman filter (KF) generates efficient forecasts and updates under linear-Gaussian state-space setup:

  \[
  \text{State:} \quad X_{t+1} = FX_t + \sqrt{\Sigma_x} \varepsilon_{t+1},
  \]

  \[
  \text{Measurement:} \quad y_t = HX_t + \sqrt{\Sigma_y} e_t.
  \]

  We can interpret \((t + 1)\) as the “next step,” with state innovation \(\Sigma_x\) increasing with the size of the time step.

- The ex ante predictions as

  \[
  \bar{X}_t = F\hat{X}_{t-1}; \quad \bar{V}_{x,t} = F\hat{V}_{x,t-1} F^\top + \Sigma_x; \\
  \bar{y}_t = H\hat{X}_t; \quad \bar{V}_{y,t} = H\bar{V}_{x,t} H^\top + \Sigma_y.
  \]

- The ex post filtering updates on the state variables are,

  \[
  K_t = \bar{V}_{x,t} H^\top (\bar{V}_{y,t})^{-1} = \bar{V}_{xy,t} (\bar{V}_{y,t})^{-1}, \quad \rightarrow \text{Kalman gain} \\
  \hat{X}_t = \bar{X}_t + K_t (y_t - \bar{y}_t), \quad \hat{V}_{x,t} = (I - K_t H) \bar{V}_{x,t}
  \]

- Model parameters can be estimated by maximizing the log likelihood,

  \[
  l_t = -\frac{1}{2} \log |\bar{V}_{y,t}| - \frac{1}{2} \left( (y_t - \bar{y}_t)^\top (\bar{V}_{y,t})^{-1} (y_t - \bar{y}_t) \right).
  \]
Intuition and control behind the Kalman filter

- Kalman filter estimates the states as a weighted average of old and new info:

\[
\hat{X}_t = X_t + K_t (y_t - \bar{y}_t) = (I - K_tH)F\hat{X}_{t-1} + K_ty_t
\]

- The weighting is given by the Kalman gain: 

\[
K_t = \frac{\bar{V}_{x,t}H^\top}{\bar{V}_{y,t}} \left( \bar{V}_{y,t}^{-1} \right)
\]

defined by the ratio of state variation (\(V_x\)) with measurement noise (\(V_y\)).

- Higher signal variation (\(V_x\)) and more accurate observation (small \(V_y\)) lead to more aggressive weights that better match the new observation.

- In case of multiple observations, \(K_t\) also provides a weighted average of the new observations, with weight proportional to \(V_y^{-1}\).

- While the dimension of the state is fixed, the dimension of measurement (number of observations) can change over time.

- Set the element of \(\Sigma_y\) (or directly \(V_y\)) to a very small number if one wants to fit one particular observation accurately.

- In general settings, \(V_x\) increases with time step (\(\sim \Sigma(X_t)\Delta t\)) — The less frequent the update (the older the prior-step estimate \(\hat{X}_{t-1}\) is), the higher is \(V_x\), and hence the less weight is given to the old.
In most applications, the measurement equations may not be linear in states: $y_t = h(X_t) + \sqrt{\Sigma} \varepsilon_t$

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

$$y_t \approx H_t X_t + \sqrt{\Sigma} \varepsilon_t, \quad H_t = \frac{\partial h(X_t)}{\partial X_t} \bigg|_{X_t=\hat{X}_t}$$

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

Numerical issues (some are well addressed in the engineering literature)

- How to compute the gradient?
- How to keep the covariance matrix positive definite.
Approximating the distribution

\[ y_t = h(X_t) + \sqrt{\Sigma_y} e_t \]

- The Kalman filter applies Bayesian rules in updating the conditionally normal distributions.

- Instead of linearly 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 binomial tree approximating a normal distribution) — unscented filter. (faster, easier to implement, and works reasonably well when \( X \) follow pure diffusion dynamics)
The unscented transformation

- Let $k$ be the number of states. A set of $2k + 1$ sigma vectors $\chi_i$ are generated according to:

$$
\chi_{t,0} = \tilde{X}_t, \quad \chi_{t,i} = \tilde{X}_t \pm \sqrt{(k + \delta)(\tilde{V}_{x,t})_j}
$$

with corresponding weights $w_i$ given by

$$
w_{0}^m = \frac{\delta}{(k + \delta)}, \quad w_{0}^c = \frac{\delta}{(k + \delta)} + (1 - \alpha^2 + \beta), \quad w_i = 1/[2(k + \delta)],
$$

where $\delta = \alpha^2(k + \kappa) - k$ is a scaling parameter, $\alpha$ (usually between $10^{-4}$ and 1) determines the spread of the sigma points, $\kappa$ is a secondary scaling parameter usually set to zero, and $\beta$ is used to incorporate prior knowledge of the distribution of $x$. It is optimal to set $\beta = 2$ if $x$ is Gaussian.

- We can regard these sigma vectors as forming a discrete distribution with $w_i$ as the corresponding probabilities.
  - Think of sigma points as a trinomial tree v. particle filtering as simulation.
The unscented Kalman filter

- **State prediction:**

\[
\chi_{t-1} = \left[ \hat{X}_{t-1}, \hat{X}_{t-1} \pm \sqrt{(k + \delta)\hat{V}_{X,t-1}} \right],
\]

\[
\bar{\chi}_{t,i} = F\chi_{t-1,i},
\]

\[
\bar{X}_t = \sum_{i=0}^{2k} w_i^{m} \bar{\chi}_{t,i},
\]

\[
\bar{V}_{X,t} = \sum_{i=0}^{2k} w_c^i(\bar{\chi}_{t,i} - \bar{X}_t)(\bar{\chi}_{t,i} - \bar{X}_t)^\top + \Sigma_x. \tag{4}
\]

- **Measurement prediction:**

\[
\bar{\chi}_t = \left[ \bar{X}_t, \bar{X}_t \pm \sqrt{(k + \delta)\bar{V}_{X,t}} \right], \quad \text{(re-draw sigma points)}
\]

\[
\bar{\zeta}_{t,i} = h(\bar{\chi}_{t,i}), \quad \bar{y}_t = \sum_{i=0}^{2k} w_i^{m} \bar{\zeta}_{t,i}.
\]

- Redrawing the sigma points in (5) is to incorporate the effect of process noise \( \Sigma_x \).

- If the state propagation equation is linear, we can replace the state prediction step in (4) by the Kalman filter and only draw sigma points in (5).
The unscented Kalman filter

- **Measurement update:**

\[
\bar{V}_{y,t} = \sum_{i=0}^{2k} w^c_i \left[ \bar{\zeta}_{t,i} - \bar{y}_t \right] \left[ \bar{\zeta}_{t,i} - \bar{y}_t \right]^\top + \Sigma_y,
\]

\[
\bar{V}_{xy,t} = \sum_{i=0}^{2k} w^c_i \left[ \bar{x}_{t,i} - \bar{X}_t \right] \left[ \bar{\zeta}_{t,i} - \bar{y}_t \right]^\top,
\]

\[
K_t = \bar{V}_{xy,t} \left( \bar{V}_{y,t} \right)^{-1},
\]

\[
\hat{X}_t = \bar{X}_t + K_t \left( y_t - \bar{y}_t \right),
\]

\[
\hat{V}_{x,t} = \bar{V}_{x,t} - K_t \bar{V}_{y,t} K_t^\top.
\]

- One can also do square root UKF to increase the numerical precision and to maintain the positivity definite property of the covariance matrix.
Outline

1. A general framework
2. Bond pricing: the general setting
3. Dynamic term structure models
4. Model design
5. Model estimation
6. Using analysts forecasts as measurements
Estimating term structure models with Kalman filter

- State propagation equations are essentially predictive regressions on highly persistent autoregressive dynamics.
  - Predictive regressions on persistent series tend to generate unstable parameter estimates, leading to bad out-of-sample performance.
  - For persistent series such as interest rates, inflation rates, exchange rates, predictive regressions can rarely beat random walk out of sample.

- When estimating term structure models with interest rate data
  - The shape of the term structure determines the risk-neutral dynamics.
    - Since we can learn the shape of the term structure fairly accurately, the risk-neutral dynamics can in general be estimated with accuracy.
  - The propagation equation determines the statistical dynamics.
    - Since there is little power in the predictive regression for such highly persistent series, the identified statistical dynamics are not trustworthy and cannot be used to predict future interest rates out of sample.
    - Most estimated term structure models cannot beat random walk in forecasting interest rates (Bali, Heidari, Wu, 2009; Duffee, 2011).
  - Estimated models are good for relative value trading, but not for predicting systematic movements.
To avoid the issues of predictive regressions, one can think of ways of estimating predictive relations without relying on predictive regressions.

Use economists’ forecasts: Estimate the predictive relation (state dynamics) via a contemporaneous relation between forecasts and current states.

- Add blue-chip forecasts on Treasury rates to the measurement equation to help identify the forecasting dynamics.
- With enhanced identification on both statistical and risk-neutral dynamics, we can gain a better understanding of bond risk premium, and understand better on what predict bond excess returns.
- Kim & Orphanides (2005) is an example.
Forecasts as measurements

- State dynamics as usual:
  \[ X_{t+1} = FX_t + \sqrt{\sum x\varepsilon_{t+1}}. \]

- Measurements include both spot rates and blue chip forecasts:
  \[ y_t = \begin{bmatrix} z(X_t, \tau) \\ BC(X_t, \tau_{BC}, h_{BC}) \end{bmatrix} + \begin{bmatrix} \sqrt{v_e^Z e_t^Z} \\ \sqrt{v_{BC}^e e_t^{BC}} \end{bmatrix} \]

- 8 Treasury zero maturities: \( \tau = 0.25, .5, 1, 2, 3, 5, 10, 30. \)

- 8 Maturities on blue chip forecasts: \( \tau_{BC} = FF, 0.25, 0.5, 1, 2, 5, 10, 30. \)
  Forecasting horizons \( h_{BC} = 1-18 \) months, updated monthly
  \( h_{BC} = 1, 2, 3, 4, 5, 6, 8, 9 \) years at long term, updated every half year.
  \( 6-10, \) and \( 7-11 \) year forecasts are approximated with \( h_{BC} = 8 \) and \( 9. \)

- Internal or other sources of forecasts \( (OS(X_t, \tau_{OS}, h_{OS})) \) can also be added as measurements, with a quality scale \( qs \) to control the weighting, \( v_e^{OS} = v_e^{BC}/qs. \)