State Space Estimation of Dynamic Term Structure Models with Forecasts

Liuren Wu

November 19, 2015
Outline

1 General setting

2 State space estimation

3 Using analysts forecasts as measurements

4 Model designs

5 Application I: Predicting bond excess returns

6 Application II: Statistical arbitrage trading

7 Shadow rate modeling

8 Going forward
Dynamic term structure models have the following generic structure:

Model specification contains three components:

1. **State dynamics**, often with autoregressive structures:
   \[ dX_t = \mu(X_t) + \sqrt{\Sigma(X_t)}dW_t \]

2. **Risk premium specification** \( \gamma(X_t) \) that translates the dynamics to the risk-neutral measure (\( Q \)-dynamics).

3. **Short rate function** that links the short rate \( r \) to the state vector \( X_t \).

The three components combine to determine the term structure:
\[ y_t = h(X_t). \]

- \( y_t \) can be zero rates, swap rates, bond prices at different maturities...
- It can also include forecasts of future rates at different maturities for different horizons.
- Bond prices are determined by the \( Q \)-dynamics and the short-rate link. Forecasts are driven by the \( P \) state dynamics. Bond excess returns (risk premium) are dictated by the \( (Q - P) \) difference, or \( \gamma(X_t) \) specification.
When the dynamics \( \mu(X_t) = \kappa(\theta - X_t), \Sigma(X_t) \) 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:

\[
\mathbb{E}_t[z(X_{t+h}, \tau)] = a(\tau) + b(\tau)^T \mathbb{E}_t[X_{t+h}]
\]

\[
= a(\tau) + b(\tau)^T (I - e^{-\kappa h})\theta + b(\tau)^T e^{-\kappa h} X_t.
\]

We consider various model designs within this tractable class.
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 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, 2014).
Outline

1. General setting
2. State space estimation
3. Using analysts forecasts as measurements
4. Model designs
5. Application I: Predicting bond excess returns
6. Application II: Statistical arbitrage trading
7. Shadow rate modeling
8. Going forward
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 state dynamics,
   - In Euler approximation, \( f(X_t) = X_t + \mu(X_t)\Delta t \) 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:
   \[ y_t = h(X_t) + \sqrt{\Sigma_y} e_t, \]
   - 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
  
  \[
  \bar{X}_t = F \hat{X}_{t-1}; \quad \bar{V}_{x,t} = F \hat{V}_{x,t-1} F^\top + \Sigma_x; \\
  \bar{y}_t = H \bar{X}_t; \quad \bar{V}_{y,t} = H \bar{V}_{x,t} H^\top + \Sigma_y.
  \]

- 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:

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

- The weighting is given by the Kalman gain: 
\[
K_t = \overline{V}_{x,t} H^\top (\overline{V}_{y,t})^{-1}
\]
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.
The Extended Kalman filter: Linearly approximating the measurement equation

- In most applications, the measurement equations may not be linear in states: \( y_t = h(X_t) + \sqrt{\Sigma_y} e_t \)

- One way to use the Kalman filter is by linear approximating the measurement equation,
  \[
  y_t \approx H_t X_t + \sqrt{\Sigma_y} e_t, \quad H_t = \frac{\partial h(X_t)}{\partial X_t} \bigg|_{X_t = \hat{X}_t}
  \]

- 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

**Measurement:** \( y_t = h(X_t) + \sqrt{\Sigma_y} e_t \)

- 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 binominal 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:

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

with corresponding weights $w_i$ given by

$$
w_0^m = \delta/(k + \delta), \quad w_0^c = \delta/(k + \delta) + (1 - \alpha^2 + \beta), \quad w_i = 1/[2(k + \delta)],
$$

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:
  \[
  \chi_{t-1} = \left[ \hat{X}_{t-1}, \hat{X}_{t-1} \pm \sqrt{(k + \delta)}V_{x,t-1} \right], \quad \text{(draw sigma points)}
  \]
  \[
  \bar{X}_{t,i} = F\chi_{t-1,i},
  \]
  \[
  \bar{X}_t = \sum_{i=0}^{2k} w_i m \bar{X}_{t,i},
  \]
  \[
  \bar{V}_{x,t} = \sum_{i=0}^{2k} w_i^c (\bar{X}_{t,i} - \bar{X}_t)(\bar{X}_{t,i} - \bar{X}_t)^\top + \Sigma_x.
  \] (2)

- Measurement prediction:
  \[
  \bar{\chi}_t = \left[ \bar{X}_t, \bar{X}_t \pm \sqrt{(k + \delta)}\bar{V}_{x,t} \right], \quad \text{(re-draw sigma points)}
  \]
  \[
  \bar{\zeta}_{t,i} = h(\bar{X}_{t,i}),
  \]
  \[
  \bar{y}_t = \sum_{i=0}^{2k} w_i^m \bar{\zeta}_{t,i}.
  \] (3)

Redrawing the sigma points in (3) 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 (2) by the Kalman filter and only draw sigma points in (3).
The unscented Kalman filter

- Measurement update:

\[
\begin{align*}
\overline{V}_{y,t} &= \sum_{i=0}^{2k} w_i^c \left[ \overline{\zeta}_{t,i} - \overline{y}_t \right] \left[ \overline{\zeta}_{t,i} - \overline{y}_t \right]^\top + \Sigma_y, \\
\overline{V}_{x,y,t} &= \sum_{i=0}^{2k} w_i^c \left[ \overline{x}_{t,i} - \overline{X}_t \right] \left[ \overline{\zeta}_{t,i} - \overline{y}_t \right]^\top, \\
K_t &= \left( \overline{V}_{x,y,t} \right)^{-1} \overline{V}_{x,y,t}, \\
\hat{X}_t &= \overline{X}_t + K_t \left( y_t - \overline{y}_t \right), \\
\hat{V}_{x,t} &= \overline{V}_{x,t} - K_t \overline{V}_{y,t} K_t^\top.
\end{align*}
\]

- 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. General setting
2. State space estimation
3. **Using analysts forecasts as measurements**
4. Model designs
5. Application I: Predicting bond excess returns
6. Application II: Statistical arbitrage trading
7. Shadow rate modeling
8. Going forward
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.

Hua & Wu (2015) on inflation forecasting: Relate inflation to interest rate, and use forward rate to predict interest rate without predictive regression.

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.
- We try to do something similar, and possibly better.
Forecasts as measurements

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

- Measurements include both spot rates and blue chip forecasts:
  \[ y_t = \left[ \begin{array}{c} z(X_t, \tau) \\ BC(X_t, \tau_{BC}, h_{BC}) \end{array} \right] + \left[ \begin{array}{c} \sqrt{v^z_e} e^z_t \\ \sqrt{v^BC_e} e^{BC}_t \end{array} \right] \]

- 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 \( h_{BC} = 1\text{-}18 \) 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 forecasts \( (TD(X_t, \tau_{TD}, h_{TD})) \) can also be added as measurements, with a quality scale \( qs \) to control the weighting, \( v^{TD}_e = v^{BC}_e / qs. \)
Outline

1. General setting
2. State space estimation
3. Using analysts forecasts as measurements
4. Model designs
5. Application I: Predicting bond excess returns
6. Application II: Statistical arbitrage trading
7. Shadow rate modeling
8. Going forward
Kim & Wright (2005) estimate a general three-factor Gaussian affine model (GA3F):

- The state is simply Gaussian VAR process, which can be standardized as
  \[ dX_t = -\kappa X_t dt + dW_t, \]
  where we standardize the state to have zero long-run mean and identity covariance matrix, and constrain the mean-reverting matrix \( \kappa \) to be lower triangular with positive diagonal values.

- Flexible market price of risk: \( \gamma(X_t) = \gamma_0 + \gamma_1 X_t \), so that the risk-neutral dynamics have a similar structure,
  \[ dX_t = (-\gamma_0 - (\kappa + \gamma_1)X_t) dt + dW_t, \]
  We also constraint \( \kappa^* = \kappa + \gamma_1 \) to be lower triangular with positive diagonal values.

- An affine short rate function: \( r(X_t) = a_r + b_r^T X_t \), where we constrain all elements of \( b_r \) to be positive to limit factor rotation.

- We estimate this model as a benchmark.
II. General Gaussian affine with common risk premium

- 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 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 \( \kappa \) is lower triangular, \( \kappa^* = \kappa + \gamma s^\top \) 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):

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

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 \(\theta_r\).
Outline

1. General setting
2. State space estimation
3. Using analysts forecasts as measurements
4. Model designs
5. Application I: Predicting bond excess returns
6. Application II: Statistical arbitrage trading
7. Shadow rate modeling
8. Going forward
Model values for zero-coupon bonds are exponential affine in the state variables,

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

and with \( X_t \) normalized to have zero mean, and \( \mathbb{E}_t^P[X_{t+h}] = e^{-\kappa h} X_t. \)

Annualized expected excess returns (risk premium) on zero bonds are:

\[
EER(X_t, \tau, h) = h^{-1} \mathbb{E}_t^P \left[ \ln \frac{P(X_{t+h}, \tau-h)}{P(X_t, \tau)} \right] - z(X_t, h) \\
= h^{-1} (A(\tau) - A(\tau-h) - A(h)) \\
+ h^{-1} \left( B(\tau)^\top - B(\tau-h)^\top e^{-\kappa h} - B(h)^\top \right) X_t \\
= \text{constant} + h^{-1} B(\tau-h) \left( e^{-\kappa^* h} - e^{-\kappa h} \right) X_t.
\]

Bond risk premium prediction depends on market price of specification, which determines \((\kappa^* - \kappa)\).

Realized excess return: \( RER(t, t+h, \tau) = h^{-1} \left[ \ln \frac{P(X_{t+h}, \tau-h)}{P(X_t, \tau)} \right] - z(X_t, h). \)
(Graphs)

- The common risk premium specification (GA3FCP) generates the best fitting and risk premium prediction.

- Parameters that were hard to identify using interest rates alone can now be identified with strong statistical significance with forecasts.

- Factor loading shows that \((X_3, X_2, X_1)\) can be regarded as short, intermediate, and long-term rate, respectively.

- The single predictive factor has a shape similar to Cochrane and Piazzesi’s finding: positive on \(X_1\) and \(X_3\), negative on \(X_2\).

- This factor can be proxyed by a tent-shaped portfolio of forward rates.
Bond excess return prediction: Out-of-sample analysis

- Model is re-estimated once a year with data up to the year. Risk premium is computed with model estimated the year before.

- Bond realized excess returns (RER) and the model-estimated risk premium (EER) show positive correlation.
  - The correlation peaks around 12-month investment horizon, and reaches about 30% for 10-year zero.
  - Expected risk premiums are mostly positive.
  - Current bond risk premium prediction is negative (starting early 2015).

- If we long the 10-year zero whenever the risk premium is positive and short 1 unit if the risk premium estimate is negative, we can generate a Sharpe ratio of 0.71, relative to a Sharpe of 0.64 for long all the time.

- Since all bond excess returns are predicted by the same factor, there is no diversification effect among bonds at different maturities.
  - Need to expand the analysis to global markets and different asset classes to increase width of investment for better diversification.
1. General setting
2. State space estimation
3. Using analysts forecasts as measurements
4. Model designs
5. Application I: Predicting bond excess returns
6. Application II: Statistical arbitrage trading
7. Shadow rate modeling
8. Going forward
Statistical arbitrage based on DTSMs


- Model provides a decomposition of observed interest rate series $y_t = h(X_t) + e_t$
  - The model value $h(X_t)$ captures the persistent (hard to predict) component of the interest rate series
  - The residuals $e_t$ capture more transient supply-demand shocks.
- Form interest rate portfolios that are neutral to the persistent factors ($X_t$) but are appropriately exposed to the residuals
  - Even though each interest rate series is very persistent, the factor-neutral portfolios are much more mean reverting.
- One can regard $-e_t = h(X_t) - y_t$ as the “alpha” of each series and $H_t = \partial h(X_t) / \partial X_t$ as its risk exposure.
- Choose portfolio weights to maximize “alpha,” minimize risk, while targeting factor neutrality (or targeted exposure):
  \[
  \max_{w_t} -w^T e_t - \frac{1}{2} \gamma w^T \Sigma_e w
  \]
  subject to factor exposure constraints: $H_t^T w_t = c \in \mathbb{R}^k$. 
Since the investment horizon for the stat arb strategy is shorter (around a month), returns computed from the stripped zeros are not reliable.

We propose to implement the strategy on swap rates.

- Re-estimate the models on libor/swap rates once a year.
- In forming portfolios, treat the swaps as par bonds, financed by 3-month LIBOR.
- At each date, form par bond portfolios based on model estimated the year before.
- Compute excess returns on the portfolio over 1-month investment horizon.
Statistical arbitrage: Some old results

Cumulative Wealth
US: Mean–variance

Cumulative Wealth
BP: Mean–variance
Outline

1. General setting
2. State space estimation
3. Using analysts forecasts as measurements
4. Model designs
5. Application I: Predicting bond excess returns
6. Application II: Statistical arbitrage trading
7. Shadow rate modeling
8. Going forward
Shadow rate modeling

- Since the financial crisis in 2008, the Fed Fund Rate has been dropped and kept close to zero.
- The short-to-mid range of the yield curve reflects market expectation on how long the short rate will be trapped at zero.
- Standard Gaussian affine dynamics no longer adequately describe the future projection of the interest rate path, and cannot readily accommodate the “S”-shaped term structure pattern.
- One approach to deal with the zero bound is Black (1995)’s shadow rate model, which allows a shadow rate to go below zero and sets the actual short rate as a maximum of a lower bound (e.g., zero) and the shadow rate.
- Krippner provides a simplification/approximation that makes the pricing more tractable.
  - In a series of papers, documents, and a book, Krippner describe his methodology and implementation for various specifications.
  - He has also been publishing key statistics from one of his estimated models (K-ANSM(2)) at: http://www.rbnz.govt.nz/research_and_publications/
Bond pricing under K-ANSM(2)

- The model has a simple two-factor structure, with \( r_t = L_t + S_t \).
- Under \( \mathbb{Q} \), the level \( (L_t) \) follows a random walk and the slope \( (S_t) \) follows a mean-reverting process,

\[
dL_t = \sigma_1 dW^1_t, \quad dS_t = -\phi S_t dt + \sigma_2 dW^2_t, \quad \rho dt = \mathbb{E}[dW^1_t dW^2_t].
\]

- The shadow forward rate can be derived in standard methods,

\[
f(t, \tau) = L_t + S_t e^{-\phi \tau} - c(\tau),
\]

where \( L_t + S_t e^{-\phi \tau} = \mathbb{E}_t^\mathbb{Q}[r_{t+\tau}] \), and \( c(\tau) \) captures a convexity effect,

\[
c(\tau) = \frac{1}{2} \left[ \sigma_1^2 \tau^2 + \sigma_2^2 G(\phi, \tau)^2 + 2\rho \sigma_1 \sigma_2 \tau G(\phi, \tau) \right], \quad G(\phi, \tau) = \frac{1}{\phi} (1 - e^{-\phi \tau}).
\]

- The shadow forward rate can go negative. The actual forward rate is taken as an option on the shadow rate with strike equal to a lower round \( r_L \),

\[
\bar{f}(t, \tau) = r_L + (f(t, \tau) - r_L) N(x(t, \tau) + S(\tau) n(x(t, \tau))),
\]

where \( N(\cdot) \) and \( n(\cdot) \) are the normal cumulative/probability density function

\[
S(\tau) = \sqrt{\sigma_1^2 u + \sigma_2^2 G(2\phi, \tau) + 2\rho \sigma_1 \sigma_2 G(\phi, \tau)}
\]
is conditional vol of \( r(t + \tau) \),

\[
x(t, \tau) = (f(t, \tau) - r_L) / S(\tau)
\]
denotes the standardized variable.
Identification under K-ANSM(2)

- The two-factor structure generates a rigid term structure for the shadow rate, allowing the model to contribute the S-shape fully to optionality.
  
  - A three-factor structure can partially accommodate the S-shape, thus making the optionality effect identification harder.

- Rates are no longer affine in states, thus necessitating the use of unscented Kalman filter.

- Estimating the model using zero rates at 3-month to 30 years generate similar patterns for SSR, ETZ, and EMS.
  
  - SSR: Shadow short rate
  - ETZ: Expected recovery time if the shadow rate is currently negative
  - EMS: A term structure slope measure meant to capture the monetary stimulus, defined by Krippner.
1. General setting
2. State space estimation
3. Using analysts forecasts as measurements
4. Model designs
5. Application I: Predicting bond excess returns
6. Application II: Statistical arbitrage trading
7. Shadow rate modeling
8. Going forward
There are many ways to model the zero-bound behavior. As an alternative,

- One can think of the economy as in two alternative states: a normal Gaussian affine state and a zero-trapped state.
- Once trapped at zero, economic recovery can be modeled as a random Poisson jump (out of recession) with arrival rate $\lambda$.
- Bond price is given by a weighted average of affine pricing and $1$.
- The extracted $\lambda$ captures the expected recovery.
- Bond pricing is more tractable in this alternative structure.

In reduced form, these models all reveal similar behaviors.
II. Link to monetary policy

- The shadow rate model makes the most economic sense when linked to a monetary policy rule:
  \[ r_t = \theta_r + \beta_{\pi}(\pi_t - \bar{\pi}_t) + \beta_x x_t + s_t \]
  - \((\pi_t - \bar{\pi}_t)\) — expected inflation deviation from target,
  - \(x_t\) — output gap,
  - \(s_t\) — policy surprise.

- The policy rule will imply a negative policy rate when inflation and output are low (or negative).

- While the actual short rate is bounded from below at zero, the Fed resorts to QE to carry the policy rule via the long end of the curve.

- The shadow rate model can use a negative shadow rate to accommodate QE at the long end.

- Estimation needs forecasts on inflation and output
  - The model can be used to link the term structure to a long list of economic announcements/forecasts (Lu & Wu, 2009).
  - Better identification of global macroeconomic movements.
One objective of designing a structural model is to connect the dots ("data").

The more dense the dots are, the less structures are required on the model.

With forecasts on 8 bond instruments over 26 horizons, and with forecasts on both inflation and real growth, we have dense observations on forecasts and the zero curve.

How can we build a flexible model structure to accommodate these dense observations while extracting the part we do not observe (e.g., risk premiums)?

Sequential steps to separately identify/model different curves
Data-driven short rate model

- Use a simple AR(1) structure to fit a smooth curve on forecasts on the short rate (FF rate, Three-Month T Bill), inflation rate (CPI/GPI), real growth (GDP/GNP):

\[
\begin{align*}
  r(t, h) &\equiv \mathbb{E}_t^p[r_{t+h}] = e^{-\kappa_r h} r_t + (1 - e^{-\kappa_r h}) \theta_r, \\
  \pi(t, h) &\equiv \mathbb{E}_t^p[\pi_{t+h}] = e^{-\kappa_\pi h} \pi_t + (1 - e^{-\kappa_\pi h}) \theta_\pi, \\
  g(t, h) &\equiv \mathbb{E}_t^p[g_{t+h}] = e^{-\kappa_g h} g_t + (1 - e^{-\kappa_g h}) \theta_g.
\end{align*}
\]

- Decompose forward rates into expectation \((r(t, \tau))\), risk premium \((\eta(t, \tau))\), and convexity \((\zeta(t, \tau))\).
  - The expectation curve \(r(t, \tau)\) is obtained from the forecasts.
  - With the AR(1) short rate structure, the convexity effect is given by
    \[
    \zeta(t, \tau) = \frac{1}{2\kappa_r^2} \left(1 - e^{-\kappa_r \tau}\right)^2 \sigma_r^2
    \]
  - What's left is risk premium, which can be regarded as a residual of the forward curve once accounting for expectation and convexity.
  - The inflation and real growth expectation curve can be brought in via a monetary policy rule.