Model Estimation and Application

Liuren Wu

Zicklin School of Business, Baruch College

Option Pricing
Outline

1. Statistical dynamics
2. Risk-neutral dynamics
3. Joint estimation
4. Applications
1. Statistical dynamics
2. Risk-neutral dynamics
3. Joint estimation
4. Applications
Estimating statistical dynamics

Constructing likelihood of the Lévy return innovation based on Fourier inversion of the characteristic function.

- If the model is a Lévy process without time change, the maximum likelihood estimation procedure is straightforward.
  - Given initial guesses on model parameters that control the Lévy triplet \((\mu, \sigma, \pi(x))\), derive the characteristic function.
  - Apply FFT to generate the probability density at a fine grid of possible return realizations — Choose a large \(N\) and a large \(\eta\) to generate a fine grid of density values.
  - Interpolate to generate intensity values at the observed return values.
  - Take logs on the densities and sum them.
  - Numerically maximize the aggregate likelihood to determine the parameter estimates.
  - Trick: Do as much pre-calculation and pre-processing as you can to speed up the estimation.
  - Standardizing the data can also be helpful in reducing numerical issues.
The same MLE method can be extended to cases where only the innovation is driven by a Lévy process, while the conditional mean and variance can be predicted by observables:

\[ \frac{dS_t}{S_t} = \mu(Z_t)dt + \sigma(Z_t)dX_t \]

where \( X_t \) denotes a Lévy process, and \( Z_t \) denotes a set of observables that can predict the mean and variance.

- Perform Euler approximation:

\[ R_{t+\Delta t} = \frac{S_{t+\Delta t} - S_t}{S_t} = \mu(Z_t)\Delta t + \sigma(Z_t)\sqrt{\Delta t}(X_{t+\Delta t} - X_t) \]

- From the observed return series \( R_{t+\Delta t} \), derive a standardized return series,

\[ SR_{t+\Delta t} = (X_{t+\Delta t} - X_t) = \frac{R_{t+\Delta t} - \mu(Z_t)\Delta t}{\sigma(Z_t)\sqrt{\Delta t}} \]

- Since \( SR_{t+\Delta t} \) is generated by the increment of a pure Lévy process, we can build the likelihood just like before.

- Given the Euler approximation, the exact forms of \( \mu(Z) \) and \( \sigma(Z) \) do not matter as much.
Estimating statistical dynamics

\[ \frac{dS_t}{S_t} = \mu(Z_t)dt + \sigma(Z_t)dX_t \]

- When \( Z \) is unobservable (such as stochastic volatility, activity rates), the estimation becomes more difficult.
- One normally needs some filtering technique to infer the hidden variables \( Z \) from the observables.
- GARCH: Use observables (return) to predict un-observable (volatility).
- Constructing variance swap rates from options and realized variance from high-frequency returns to make activity rates more observable.
  - Wu, Variance Dynamics: Joint Evidence from Options and High-Frequency Returns.
Estimating statistical dynamics

Wu, Variance Dynamics: Joint Evidence from Options and High-Frequency Returns.

- Use index options to replicate variance swap rates, VIX.
  - Under affine specifications, \( \text{VIX}_t = \frac{1}{T} \mathbb{E}^Q[\int_{t}^{t+h} \nu_s ds] = a(h) + b(h)v_t \), where \( (a(h), b(h)) \) are functions of risk-neutral \( \nu \)-dynamics.
  - Solve for \( \nu_t \) from VIX: \( \nu_t = (\text{VIX}_t^2 - a(h))/b(h) \).
  - Build the likelihood on \( \nu_t \) as an observable:
    \[
    dv_t = \mu(v_t)dt + \sigma(v_t)dX_t
    \]
    - Use Euler approximation to solve for the Lévy component \( X_{t+\Delta t} - X_t \) from \( \nu_t \).
    - Build the likelihood on the Lévy component based on FFT inversion of the characteristic function.

- Use high-frequency returns to construct daily realized variance (RV).
  - Treat RV as noisy estimators of \( \nu_t \): \( \text{RV}_t = \nu_t \Delta t + \text{error} \).
  - Given \( \nu_t \), build quasi-likelihood function on the realized variance error.

- Future research: Incorporate more observables.
Outline

1. Statistical dynamics
2. Risk-neutral dynamics
3. Joint estimation
4. Applications
Estimating the risk-neutral dynamics

- Nonlinear weighted least square to fit Lévy models to option prices. Daily calibration (Bakshi, Cao, Chen (1997, JF), Carr and Wu (2003, JF))

- The key issue is how to define the pricing error and how to build the weight:
  - In-the-money is dominated by the intrinsic value, not by the model. At each strike, use the out-of-the-money option: Call when $K > F$ and put when $K \leq F$.
  - Pricing errors can either be absolute errors (market minus model), or percentage errors (log (market/model)).
    - Using absolute errors favors options with higher values (longer maturity, near the money).
    - Using percentage errors put more uniform weight across options, but may put too much weight on illiquid options (far out of money).
  - Errors can be either in dollar prices or implied volatilities.
  - My current choice: Use out of money option prices to define absolute errors, use the inverse of vega as weights.
Estimating the risk-neutral dynamics

- Sometimes separate calibration per maturity is needed for a simple Lévy model (e.g., VG, MJD)
  - Lévy processes with finite variance implies that non-normality dies away quickly with time aggregation.
  - Model-generated implied volatility smile/smirk flattens out at long maturities.
  - Separate calibration is necessary to capture smiles at long maturities.
- Adding a persistent stochastic volatility process (time change) helps improve the fitting along the maturity dimension.
  - Daily calibration: activity rates and model parameters are treated the same as free parameters.
  - Dynamically consistent estimation: Parameters are fixed, only activity rates are allowed to vary over time.
Static v. dynamic consistency

- **Static cross-sectional consistency**: Option values across different strikes/maturities are generated from the same model (same parameters) at a point in time.

- **Dynamic consistency**: Option values over time are also generated from the same no-arbitrage model (same parameters).

While most academic & practitioners appreciate the importance of being both cross-sectionally and dynamically consistent, it can be difficult to achieve while generating good pricing performance. So it comes to compromises.

- **Market makers**:
  - Achieving static consistency is sufficient.
  - Matching market prices is important to provide two-sided quotes.

- **Long-term convergence traders**:
  - Pricing errors represent trading opportunities.
  - Dynamic consistency is important for long-term convergence trading.

A well-designed model (with several time changed Lévy components) can achieve both dynamic consistency and good performance.

*Fewer parameters (parsimony), more activity rates.*
Dynamically consistent estimation

- Nested nonlinear least square (Huang and Wu (2004)): Often has convergence issues.

- Cast the model into state-space form and use MLE.
  - Define state propagation equation based on the $\mathbb{P}$-dynamics of the activity rates. (Need to specify market price on activity rates, but not on return risks).
  - Define the measurement equation based on option prices (out-of-money values, weighted by vega,...)
  - Use an extended version of Kalman filter (EKF, UKF, PKF) to predict/filter the distribution of the states and measurements.
  - Define the likelihood function based on forecasting errors on the measurement equations.
  - Estimate model parameters by maximizing the likelihood.
The Classic Kalman filter

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

  \[ \begin{align*}
  \text{State:} & \quad X_{t+1} = FX_t + \sqrt{\Sigma_x} \varepsilon_{t+1}, \\
  \text{Measurement:} & \quad y_t = HX_t + \sqrt{\Sigma_y} e_t 
  \end{align*} \]

- The ex ante predictions as

  \[ \begin{align*}
  \overline{X}_t & = F \hat{X}_{t-1}; \\
  \overline{V}_{x,t} & = F \hat{V}_{x,t-1} F^\top + \Sigma_x; \\
  \overline{y}_t & = H \overline{X}_t; \\
  \overline{V}_{y,t} & = H \overline{V}_{x,t} H^\top + \Sigma_y. 
  \end{align*} \]

- The ex post filtering updates are,

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

- The log likelihood is built on the forecasting errors of the measurements,

  \[ l_t = -\frac{1}{2} \log |\overline{V}_{y,t}| - \frac{1}{2} \left( (y_t - \overline{y}_t)^\top (\overline{V}_{y,t})^{-1} (y_t - \overline{y}_t) \right). \]
Numerical twists

- Kalman filter with fading memory: Replace $\bar{V}_{x,t}$ with

$$
\bar{V}_{x,t} = \alpha^2 F \bar{V}_{x,t-1} F^\top + \Sigma_x,
$$

with $\alpha > 1$ (e.g., $\alpha \approx 1.01$). This slight raise on $\bar{V}_{x,t}$ increases the Kalman gain and hence increases the responsiveness of the states to the new observations (versus old observations). This twist allows the filtering to forget about old observations gradually.

- Application: Recursive least square estimation of the coefficient $X$:

$$
y_t = H_t X + e_t.
$$

In this case, the coefficient $X$ is the hidden state. In this case, we can estimate the coefficient recursively and update the coefficient estimate after each new observation $y_t$:

$$
\hat{X}_t = \hat{X}_{t-1} + K_t (y_t - H_t \hat{X}_{t-1}),
$$

with

$$
\begin{align*}
\bar{V}_{x,t} &= \bar{V}_{x,t-1} \alpha^2, \\
K_t &= \bar{V}_{x,t} H_t^\top (H_t \bar{V}_{y,t} H_t^\top + \Sigma_y), \\
V_{x,t} &= (I - K_t H_t) \bar{V}_{x,t}.
\end{align*}
$$
The Extended Kalman filter: Linearly approximating the measurement equation

- If we specify affine-diffusion dynamics for the activity rates, the state dynamics ($X$) can be regarded as Gaussian linear, but option prices ($y$) are not linear in the states:

  State: \( X_{t+1} = FX_t + \sqrt{\Sigma_x} \epsilon_{t+1} \),
  Measurement: \( y_t = h(X_t) + \sqrt{\Sigma_y} e_t \)

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

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

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

- Numerical 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.
- If the state vector does not follow diffusion dynamics and hence can no longer be approximated by Gaussian, the sigma points may not be enough. Particle filtering is needed.
The unscented Kalman filter

- State prediction:

\[ \chi_{t-1} = \begin{bmatrix} \hat{X}_{t-1} \\ \hat{X}_{t-1} \pm \sqrt{(k + \delta)\hat{V}_{x,t-1}} \end{bmatrix}, \quad \text{(draw sigma points)} \]

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

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

\[ \bar{V}_{x,t} = \sum_{i=0}^{2k} w_i^c (\bar{X}_{t,i} - \bar{X}_t)(\bar{X}_{t,i} - \bar{X}_t)^\top + \Sigma_x. \]

- Measurement prediction:

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

\[ \bar{\zeta}_{t,i} = h(\bar{\chi}_{t,i}), \]

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

- Redrawing the sigma points in (3) is to incorporate the effect of process noise \( \Sigma_x \).

- Since the state propagation equation is linear in our specification, 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:**

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

- One can also do square root UKF to increase the numerical precision and to maintain the positivity definite property of the covariance matrix.
Let $S_t$ denote the Cholesky factor of $V_t$ such that $V_t = S_t S_t^T$.

State prediction:

\[
\chi_{t-1} = \left[ \hat{X}_{t-1}, \hat{X}_{t-1} \pm \sqrt{(k + \delta)\hat{S}_{x,t-1}} \right], \quad \text{(draw sigma points)}
\]

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

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

\[
\overline{S}_{x,t} = \operatorname{qr}\left\{ \left[ \sqrt{w_1^c} (\overline{\chi}_{t,1:2^k} - \overline{X}_t) \right] \sqrt{\Sigma_x} \right\},
\]

\[
\overline{S}_{x,t} = \operatorname{cholupdate}\left\{ \overline{S}_{x,t}, \overline{\chi}_{t,0} - \overline{X}_t, w_0^c \right\}.
\]

Measurement prediction:

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

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

\[
\bar{S}_{y,t} = \text{qr}\{ \sqrt{w_1^c (\zeta_{t,1:2k} - \bar{y}_t)} \sqrt{\Sigma_y} \},
\]

\[
\tilde{S}_{y,t} = \text{cholupdate}\{ \bar{S}_{y,t}, \zeta_{t,0} - \bar{y}_t, w_0^c \},
\]

\[
V_{xy,t} = \sum_{i=0}^{2k} w_i^c [\bar{X}_{t,i} - \bar{X}_t] [\zeta_{t,i} - \bar{y}_t]^T,
\]

\[
K_t = \left( V_{xy,t}/\tilde{S}_{y,t}^\top \right) / \tilde{S}_{y,t},
\]

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

\[
U = K_t \tilde{S}_{y,t},
\]

\[
\hat{S}_{x,t} = \text{cholupdate}\{ \bar{S}_{x,t}, U, -1 \}.
\]
1. Statistical dynamics
2. Risk-neutral dynamics
3. Joint estimation
4. Applications
Joint estimation of $P$ and $Q$ dynamics

- Pan (2002, JFE): GMM. Choosing moment conditions becomes increasing difficult with increasing number of parameters.

- Eraker (2004, JF): Bayesian with MCMC. Choose 2-3 options per day. Throw away lots of cross-sectional ($Q$) information.

- Bakshi & Wu (2010, MS), “The Behavior of Risk and Market Prices of Risk over the Nasdaq Bubble Period”

  MLE with filtering

  - Cast activity rate $P$-dynamics into state equation, cast option prices into measurement equation.
  - Use UKF to filter out the mean and covariance of the states and measurement.
  - Construct the likelihood function of options based on forecasting errors (from UKF) on the measurement equations.
  - Given the filtered activity rates, construct the conditional likelihood on the returns by FFT inversion of the conditional characteristic function.

  The joint log likelihood equals the sum of the log likelihood of option pricing errors and the conditional log likelihood of stock returns.
Existing issues

- When the state dynamics are discontinuous, the performance of UKF can deteriorate. Using particle filter instead increases the computational burden dramatically.

- Inconsistency regarding the current state of the state vector.
  - When we price options, we derive the option values (Fourier transforms) as a function of the state vector.
  - In model estimation, the value of the state vector is never known, we only know its distribution (characterized by the sigma points).
  - The two components (pricing and model estimation) are inherently inconsistent with each other.
  - Several papers try to use incomplete information to explain the correlation between credit events and their arrival intensity.
  - Conditional on incomplete information about the state vector (the current state of the state vector is not known in full), how the option valuation should adjust?
Why are we doing this?

Go back to the overview slides:

1. Inferring expectations and risk preferences. The literature on this line of research is very long... You can tell a more detailed story with options than with the primary securities alone.

2. Interpolation and extrapolation:
   - Market makers, normally perform frequent calibration of a simpler model. The calibration could even be on each maturity separately.
   - Data vendors (such as Bloomberg, Markit): This is a rapidly growing business — Populating quotes using models for markets without valid quotes.

3. Alpha generation: *Statistical arbitrage based on no-arbitrage models*
   - Normally needs a dynamically consistent way of estimating the model. The model can be more rigid but needs to be stable.
   - I’ll give some examples on this line of application.
Example I: Statistical arbitrage on interest rate swaps based on no-arbitrage dynamic term structure models.


This is not on options or an option pricing model, but it is one of the first papers that provide detailed steps of such an application.

Basic idea:

- Interest rates across different maturities are related.
- A dynamic term structure model provides a (smooth) 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.
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 dynamic term structure models

- 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.
  - 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.
Static arbitrage trading based on no-arbitrage dynamic term structure models

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

Hedged (left) v. unhedged (right)

It is much easier to predict the hedged portfolio (left panel) than the unhedged swap contract (right panel).
Back-testing results from a simple investment strategy

95-00: In sample. 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.
  - The model not only needs to generate mean-reverting pricing errors, but also needs to generate stable risk sensitivities — The swap portfolio constructed based on the model must be really free of market risk.
  - 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.
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

Liuren Wu (Baruch)
A no-arbitrage model that prices 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
A more generic 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 (by assuming that it will go to zero down the road) and \( H_t \) its risk exposure.

- We can solve the following quadratic program:
  \[
  \max_{w_t} -w^\top e_t - \frac{1}{2} \gamma w^\top V_e w \\
  \text{subject to factor exposure constraints:} \\
  H_t^\top w_t = c \in \mathbb{R}^k.
  \]

  - The above equation maximizes the expected return (alpha) of the portfolio subject to factor exposure constraints.
  - Setting \( c = 0 \) maintains (first order) factor-neutrality.
  - One can also take on factor exposures to obtain factor risk premium.
  - Caveat: Special care is needed for factor neutrality when the state vector can jump.
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, or
- Take a controlled exposure to certain risk sources and receive risk premiums accordingly.
  - Understanding risk premium behavior is useful not just for academic papers.
- Perform statistical arbitrage trading on derivative products that profit from short-term market dislocations.

Caveat: When hedging is off, risk can overwhelm profit opportunities.