Computational Challenges in Option Pricing

Liuren Wu

Zicklin School of Business, Baruch College

Supercomputing Center, Chinese Academy of Sciences
July 9, 2008, Beijing, China
How to break out in computational finance research?

- Be relevant.
  - Research should be driven by real, relevant economic *issues*, not by the *tools* that you already have.

- Be simple.
  - *Everything should be made as simple as possible, but not simpler.*
    — Albert Einstein.

- Be innovative.
Outline

- How to **model** financial security returns using *time-changed Lévy processes*
  - with an eye on data, economic sense, and computational tractability.

- How to **price** options based on these models
  - with an eye on numerical efficiency.

- How to **estimate** these models
  - with an eye on different applications:
    - market-making,
    - long-term convergence trading,
    - risk-premium taking for systematic risk exposure,
Outline

1. Design economically sensible & computationally feasible option pricing models based on *time-changed Lévy processes*

2. Efficient option pricing via Fourier inversions

3. Estimate option pricing models for different purposes
Why time-changed Lévy processes?

Key advantages:

- **Generality:**
  - Lévy processes can generate almost any return innovation distribution.
  - Applying stochastic time changes randomizes the innovation distribution over time $\Rightarrow$ stochastic volatility, correlation, skewness, ....

- **Explicit economic mapping:**
  - Each Lévy component $\leftrightarrow$ shocks from one economic source.
  - Time change captures the time-varying intensity of its impact.
  $\Rightarrow$ makes model design more intuitive, parsimonious, and economically sensible.

- **Tractability:** A model is tractable for option pricing if we have
  - tractable characteristic exponent for the Lévy components.
  - tractable Laplace transform for the time change.
  $\Rightarrow$ Any combinations of the two generate tractable return dynamics.
A Lévy process is a continuous-time process that generates stationary, independent increments ...

Think of return innovation in discrete time: $R_{t+1} = \mu_t + \sigma_t \epsilon_{t+1}$.

Lévy processes generate iid return innovation distributions via the Lévy triplet $(\mu, \sigma, \pi(x))$. ($\pi(x)$–Lévy density).

The Lévy-Khintchine Theorem:

$$\phi_{X_t}(u) \equiv \mathbb{E}[e^{iuX_t}] = e^{-t\psi(u)}, \quad u \in \mathcal{D} \subseteq \mathbb{C}$$

$$\psi(u) = -iu\mu + \frac{1}{2}u^2\sigma^2 + \int_{\mathbb{R}_0} \left(1 - e^{iux} + iux1_{|x|<1}\right) \pi(x)dx,$$

Innovation distribution $\leftrightarrow$ characteristic exponent $\psi(u) \leftrightarrow$ Lévy triplet

Tractable: The integral can be carried out explicitly.
Tractable examples

1. Brownian motion (BSM) ($\mu t + \sigma W_t$): normal shocks.
2. Compound Poisson jumps (Merton, 76): Large but rare events.

$$\pi(x) = \lambda \frac{1}{\sqrt{2\pi v_J}} \exp \left( -\frac{(x - \mu_J)^2}{2v_J} \right).$$

3. Dampened power law (DPL):

$$\pi(x) = \begin{cases} 
\lambda \exp(-\beta_+ x) x^{-\alpha-1}, & x > 0, \quad \lambda, \beta_+ > 0, \\
\lambda \exp(-\beta_- |x|) |x|^{-\alpha-1}, & x < 0, \quad \alpha \in [-1, 2)
\end{cases}$$

- Finite activity when $\alpha < 0$: $\int_{\mathbb{R}^0} \pi(x) dx < \infty$. Compound Poisson. Large and rare events.
- Infinite activity when $\alpha \geq 0$: Both small and large jumps.
- Infinite variation when $\alpha \geq 1$: many small jumps,

$$\int_{\mathbb{R}^0} (|x| \wedge 1) \pi(x) dx = \infty.$$  

Market movements of all magnitudes, from small movements to market crashes.
Analytical characteristic exponents

- Diffusion: $\psi(u) = -iu\mu + \frac{1}{2}u^2\sigma^2$.

- Merton's compound Poisson jumps:
  $$\psi(u) = \lambda \left(1 - e^{iu\mu} - \frac{1}{2}u^2\nu\right).$$

- Dampened power law: (for $\alpha \neq 0, 1$)
  $$\psi(u) = -\lambda \Gamma(-\alpha) \left[(\beta_+ - iu)^\alpha - \beta_+^\alpha + (\beta_- + iu)^\alpha - \beta_-^\alpha\right] - iuC(h)$$

  - When $\alpha \to 2$, smooth transition to diffusion (quadratic function of $u$).
  - When $\alpha = 0$ (Variance-gamma by Madan et al):
    $$\psi(u) = \lambda \ln(1 - iu/\beta_+) \ln(1 + iu/\beta_-) = \lambda \left(\ln(\beta_+ - iu) - \ln\beta + \ln(\beta_- + iu) - \ln\beta_-\right).$$

  - When $\alpha = 1$ (exponentially dampened Cauchy, Wu 2006):
    $$\psi(u) = -\lambda \left((\beta_+ - iu) \ln(\beta_+ - iu)/\beta_+ + \lambda (\beta_- + iu) \ln(\beta_- + iu)/\beta_-\right) - iuC(h).$$

  - $\beta_{\pm} = 0$ (no dampening): $\alpha$-stable law
Other Lévy examples

- Other examples:
  - The normal inverse Gaussian (NIG) process of Barndorff-Nielsen (1998)
  - The generalized hyperbolic process (Eberlein, Keller, Prause (1998))
  - The Meixner process (Schoutens (2003))
  - ...

- Bottom line:
  - All tractable in terms of analytical characteristic exponents $\psi(u)$.
  - We can use Fourier inversion methods to generate the density function of the innovation (for model estimation).
  - We can also use Fourier inversion methods to compute option values ...

- Reality check: Do we need Lévy jumps to model financial security returns?
  - It is important to look at the data...
Implied volatility smiles & skews on a stock

\[ \text{Moneyness} = \frac{\ln(K/F)}{\sigma \sqrt{\tau}} \]

**Short-term smile**

**Long-term skew**

Maturities: 32  95  186  368  732
Implied volatility skews on a stock index (SPX)

More skews than smiles

Maturities: 32  60  151  242  333  704

Moneyness = \frac{\ln(K/F)}{\sigma \sqrt{\tau}}
Average implied volatility smiles on currencies

Maturities: 1m (solid), 3m (dashed), 1y (dash-dotted)
The role of jumps at very short maturities

- Implied volatility smiles (skews) ↔ non-normality (asymmetry) for the risk-neutral return distribution (Backus, Foresi, Wu (97)):

\[ IV(d) \approx ATMV \left( 1 + \frac{\text{Skew.}}{6} d + \frac{\text{Kurt.}}{24} d^2 \right), \quad d = \frac{\ln K/F}{\sigma \sqrt{\tau}} \]

- Two mechanisms to generate return non-normality:
  - Use Lévy jumps to generate non-normality for the innovation distribution.
  - Use stochastic volatility to generate non-normality through mixing over multiple periods.

- Over very short maturities (1 period), only jumps contribute to return non-normalities.
Time decay of short-term OTM options


- As option maturity ↓ zero, OTM option value ↓ zero.
- The speed of decay is exponential $O(e^{-c/T})$ under pure diffusion, but linear $O(T)$ in the presence of jumps.
- Term decay plot: $\ln(OTM/T) \sim \ln(T)$ at fixed $K$:

  ![Term decay plot](image)

- *In the presence of jumps, the Black-Scholes implied volatility for OTM options $IV(\tau, K)$ explodes as $\tau \downarrow 0$.**
When a company defaults, its stock value jumps to zero.


This default risk generates a steep skew in long-term stock options.

Evidence: Stock option implied volatility skews are correlated with credit default swap (CDS) spreads written on the same company.

Economic implications

- In the Black-Scholes world (one-factor diffusion):
  - The market is complete with a bond and a stock.
  - The world is risk free after delta hedging.
  - Utility-free option pricing. Options are redundant.

- In a pure-diffusion world with stochastic volatility:
  - Market is complete with one (or a few) extra option(s).
  - The world is risk free after delta and vega hedging.

- In a world with jumps of random sizes:
  - The market is inherently incomplete (with stocks alone).
  - Need all options (+ model) to complete the market.
  - **Challenges**: Greeks-based dynamic hedging is no longer risk proof. How to design robust hedges in the presence of jumps?
  - **Challenges**: Options must be used in estimating the risk dynamics.
  - **Opportunities**: Options market is informative/useful:
    - Cross sections \((K, T) \leftrightarrow Q\) dynamics.
    - Time series \((t) \leftrightarrow P\) dynamics.
    - The difference \(Q/P \leftrightarrow \text{market prices of economic risks.}\)
Lévy processes can generate different iid return innovation distributions.

- Any distribution you can think of, we can specify a Lévy process, with the increments of the process matching that distribution.

Yet, return distribution is not iid. It varies over time.

- That’s why I have shown you only cross-sectional plots ...

We need to go beyond Lévy processes to capture the time variation in the return distribution (implied volatility surface):

- Stochastic volatility
- Stochastic risk reversal (skewness)
At-the-money implied volatilities at fixed time-to-maturities from 1 month to 5 years.
Stochastic volatility on currencies

Three-month delta-neutral straddle implied volatility.
Implied volatility spread between 80% and 120% strikes at fixed time-to-maturities from 1 month to 5 years.
Stochastic skewness on currencies

Three-month 10-delta risk reversal (blue lines) and butterfly spread (red lines).
Randomize the time

- Review the Lévy-Khintchine Theorem:

\[
\phi(u) \equiv \mathbb{E}[e^{iuX_t}] = e^{-t\psi(u)},
\]

\[
\psi(u) = -iu\mu + \frac{1}{2}u^2\sigma^2 + \lambda \int_{\mathbb{R}} (1 - e^{iu} + iux1_{|x|<1}) \tilde{\pi}(x)dx,
\]

- The drift \(\mu\), the diffusion variance \(\sigma^2\), and the mean arrival rate \(\lambda\) are all proportional to time \(t\).

- We can directly specify \((\mu_t, \sigma^2_t, \lambda_t)\) as following stochastic processes.

- Or we can randomize time \(t \rightarrow T_t\) for the same result.

- We define \(T_t \equiv \int_0^t \nu_s^- ds\) as the \textit{stochastic time change}, with \(\nu_t\) being the \textit{instantaneous activity rate}.

- Depending on the Lévy specification, the activity rate has the same meaning (up to a scale) as a randomized version of the \textit{instantaneous drift}, \textit{instantaneous variance}, or \textit{instantaneous arrival rate}.
In 1949, Bochner introduced the notion of time change to stochastic processes. In 1973, Clark suggested that time-changed diffusions could be used to accurately describe financial time series.

Ane & Geman (2000) show supporting evidence: Define returns over fixed number of trades, not over fixed calendar time intervals.

Two types of clocks can be used to model business time:

1. Clocks based on increasing jump processes have staircase like paths.
2. Continuous clocks \( \mathcal{T}_t \equiv \int_0^t \nu(s) \, ds \) as we have just defined.

The first type of clock can transform a diffusion into a jump process — All Lévy processes considered earlier can be generated as changing the clock of a diffusion with an increasing jump process (subordinator).

The second type of business clock can be used to describe stochastic volatility (and higher moments).

Monroe (1978): All semimartingales can be generated by applying stochastic time changes (of both types) on Brownian motions.
Economic interpretations

- Treat $t$ as the calendar time, and $I_t \equiv \int_0^t \nu_s \, ds$ as the business time.
  - Business activity accumulates with calendar time, but the speed varies, depending on the business activity.
  - Business activity tends to intensify before earnings announcements, FOMC meeting days...
  - In this sense, $\nu_t$ captures the intensity of business activity at time $t$.
  - In options market making/trading, it is important to build an accurate business calendar.

- Economics shocks (impulse) and financial market responses:
  - Think of each Lévy process (component) as capturing one of the many sources of economic shocks.
  - The stochastic time change on each Lévy component captures the random intensity of the impact of the economic shock on the financial security.

\[
\text{Return} \sim \sum_{i=1}^K X^i_{I_t} \sim \sum_{i=1}^K (\text{Economic shock})^i_{Stochastic \ impact}.
\]
Example: Return on a stock

- Model the return on a stock to reflect shocks from **two sources**:
  - **Credit risk**: In case of corporate default, the stock price falls to zero. Model the impact as a Poisson Lévy jump process with log return jumps to negative infinity upon jump arrival.
  - **Market risk**: Daily market movements (small or large). Model the impact as a diffusion or infinite-activity (infinite variation) Lévy jump process or both.

- Apply separate time changes to the two Lévy components to capture (1) the intensity variation of corporate default, (2) the market risk (volatility) variation, as well as their interactions.

- **Key**: *Each component has a specific economic purpose.*

- **Application**: Cross-market trading.

Example: A CAPM model:

\[ \ln S^j_t / S^j_0 = (r - q)t + \left( \beta^j X^m_{T^m_t} - \varphi_x^m (\beta^j) T^m_t \right) + \left( X^j_{T^j_t} - \varphi_x^j (1) T^j_t \right) . \]

- Estimate \( \beta \) and market prices of return and volatility risk using index and single name options.
- Cross-sectional analysis of the estimates.

- Application: Dispersion trading. Analyze the interactions of the return volatility in both levels and innovations.

An international CAPM:

Example: Return on an exchange rate

- Exchange rate reflects the interaction between two economic forces.
- Use two Lévy processes to model the two economic forces separately.
- Consider a negatively skewed distribution (downside jumps) from each economic source (crash-o-phobia from both sides). Use the difference to model the currency return between the two economies.
- Apply separate time changes to the two Lévy processes to capture the strength variation (tug war) between the two economic forces.
  - Stochastic time changes on the two negatively skewed Lévy processes generate both stochastic volatility and stochastic skew.
- Key: Each component has its specific economic purpose.

Example: Exchange rates and pricing kernels

- Exchange rate reflects the interaction between two economic forces.
- The economic meaning becomes clearer if we model the pricing kernel of each economy.
  - Let $m_{0,t}^{US}$ and $m_{0,t}^{JP}$ denote the pricing kernels of the US and Japan. Then the dollar price of yen $S_t$ is given by
    \[
    \ln \frac{S_t}{S_0} = \ln m_{0,t}^{JP} - \ln m_{0,t}^{US}.
    \]
  - If we model the negative of the logarithm of each pricing kernel ($-\ln m_{0,t}^j$) as a time-changed Levy process, $X_{T_t}^j (j = US, JP)$ with negative skewness. Then, $\ln S_t/S_0 = \ln m_{0,t}^{JP} - \ln m_{0,t}^{US} = X_{T_t}^{US} - X_{T_t}^{JP}$
    - Think of $X$ as consumption growth shocks
    - Think of $T_t$ as time-varying risk premium.

- Application: Consistent and simultaneous modeling of all currency pairs.

Outline

1. Design economically sensible & computationally feasible option pricing models based on *time-changed Lévy processes*

2. Efficient option pricing via Fourier inversions

3. Estimate option pricing models for different purposes
To compute the time-0 price of a European option price with expiry at \( t \), we first compute the Fourier transform of the log return \( s_t \equiv \ln S_t / S_0 \).

The generalized Fourier transform of a time-changed Lévy process:

\[
\phi_Y(u) \equiv \mathbb{E}^Q \left[ e^{iuX_T} \right] = \mathbb{E}^\mathbb{M} \left[ e^{-\psi_x(u)T} \right], \quad u \in D \subseteq \mathbb{C},
\]

where the new measure \( \mathbb{M} \) is defined by the exponential martingale:

\[
\frac{d\mathbb{M}}{d\mathbb{Q}} \bigg|_t = \exp \left( iuX_T + T \psi_x(u) \right).
\]

Without time-change, \( e^{iuX_T + T \psi_x(u)} \) is an exponential martingale by Lévy-Khintchine Theorem.

A continuous time change does not change the martingality.

\( \mathbb{M} \) is complex valued (no longer a probability measure).

Tractability of the transform \( \phi(u) \) depends on the tractability of

- The characteristic exponent of the Lévy process \( \psi_x(u) \).
- The Laplace transform of \( T_t \) under \( \mathbb{M} \).

\((X, T_t)\) can be chosen separately as building blocks to capture the two dimensions: **Moneyness & term structure.**
The Laplace transform of the stochastic time $\mathcal{T}_t$

- We have solved the characteristic exponent of the Lévy process (by the Lévy-Khintchine Theorem).
- Compare the Laplace transform of the stochastic time,
  \[ \mathcal{L}_{\mathcal{T}}(\psi) \equiv \mathbb{E} \left[ e^{-\psi \mathcal{T}_t} \right] = \mathbb{E} \left[ e^{-\psi \int_0^t \nu_s ds} \right] \]  
  (1)
- to the pricing equation for zero-coupon bonds:
  \[ B(0, t) \equiv \mathbb{E}^\mathbb{Q} \left[ e^{-\int_0^t r_s ds} \right] \]  
  (2)
- The two pricing equations look analogous
  - Both $\nu_t$ and $r_t$ need to be positive.
  - If we set $r_t = \psi \nu_t$, $\mathcal{L}_{\mathcal{T}}(\psi)$ is essentially the bond price.
- The analogy allows us to borrow the vast bond pricing literature:
  - **Affine class**: The Laplace transform is exponential affine in the state variable.
  - **Wishart dynamics**: Natural modeling of covariance matrices.
From Fourier transforms to option prices

- With the Fourier transform of the log return ($\phi(u)$), we can compute vanilla option values via Fourier inversion.

- Take a European call option as an example.
  - Perform the following rescaling and change of variables:
    \[
    c(k) = e^{rt} c(K, t)/F_0 = \mathbb{E}_0^Q [(e^{s_t} - e^k)1_{s_t \geq k}],
    \]
    with $s_t = \ln F_t/F_0$ and $k = \ln K/F_0$.
  - $c(k)$: the option forward price in percentage of the underlying forward as a function of moneyness defined as the log strike over forward, $k$ (at a fixed time to maturity).

- Derive the Fourier transform of the scaled option value $c(k)$ ($\chi_c(u)$) in terms of the Fourier transform ($\phi_s(u)$) of the log return $s_t = \ln F_t/F_0$.

- Perform numerical Fourier inversion to obtain option value.

- There are many ways of doing this.
I. The CDF analog

- Treat \( c(k) = E^Q_0 \left[ (e^{st} - e^k) 1_{st \geq k} \right] = \int_{-\infty}^{\infty} (e^{st} - e^k) 1_{st \geq x} dF(s) \) as a CDF.

- The option transform:

\[
\chi_c^I(u) \equiv \int_{-\infty}^{\infty} e^{iku} \, dc(k) = -\frac{\phi_s(u-i)}{iu+1}, \quad u \in \mathbb{R}.
\]

Thus, if we know the CF of the return, \( \phi_s(u) \), we know the transform of the option, \( \chi_c^I(u) \).

- The inversion formula is analogous to the inversion of a CDF:

\[
c(x) = \frac{1}{2} + \frac{1}{2\pi} \int_{0}^{\infty} e^{iux} \chi_c^I(-u) - e^{-iux} \chi_c^I(u) \, du.
\]

- Use quadrature methods for the numerical integration.
  It can work well if done right.

- The literature often writes: \( c(x) = e^{-qt} Q_1(x) - e^{-rt} e^{-x} Q_2(x) \). Then, we must invert twice.


II. The PDF analog

- Treat $c(k)$ analogous to a PDF.

- The option transform:

$$
\chi^\|_c(z) \equiv \int_{-\infty}^{\infty} e^{ik} c(k) \, dk = \frac{\phi_s(z - i)}{(iz)(iz + 1)}
$$

with $z = u - i\alpha$, $\alpha \in \mathcal{D} \subseteq \mathbb{R}^+$ for the option transform to be well defined.

  - The range of $\alpha$ depends on payoff structure and model.
  - The exact value choice of $\alpha$ is a numerical issue.
  - Carr and Madan (1999, Journal of Computational Finance) refer to $\alpha$ as the dampening coefficient.
  - Given the transform on return $\phi_s(u)$, we know the transform on call.

- The inversion is analogous to that for a PDF:

$$
c(k) = \frac{1}{2\pi} \int_{-i\alpha - \infty}^{-i\alpha + \infty} e^{-izk} \chi^\|_c(z) \, dz = \frac{e^{-\alpha k}}{\pi} \int_0^\infty e^{-iuk} \chi^\|_c(u - i\alpha) \, du.
$$

Fast Fourier Transform (FFT)

- FFT is an efficient algorithm for computing discrete Fourier coefficients.

- The discrete Fourier transform is a mapping of \( f = (f_0, \cdots, f_{N-1})^\top \) on the vector of Fourier coefficients \( d = (d_0, \cdots, d_{N-1})^\top \), such that

\[
d_j = \frac{1}{N} \sum_{k=0}^{N-1} f_k e^{-jk \frac{2\pi}{N} i}, \quad j = 0, 1, \cdots, N - 1.
\]

- FFT allows the efficient calculation of \( d \) if \( N \) is an even number, say \( N = 2^n, n \in \mathbb{N} \). The algorithm reduces the number of multiplications in the required \( N \) summations from an order of \( 2^{2n} \) to that of \( n2^{n-1} \), a very considerable reduction.

- By a suitable discretization, we can approximate the inversion of a PDF (also option price) in the above form to take advantage of the computational efficiency of FFT.
Call value inversion

Compare the call inversion (method II) with the FFT form:

\[ c(k) = \frac{e^{-\nu k}}{\pi} \int_0^\infty e^{-iku} \chi^\ll_c(u - i\nu) du. \]

\[ d_j = \frac{1}{N} \sum_{m=0}^{N-1} f_m e^{-jm\frac{2\pi}{N}i} \]

- Discretize the integral using the trapezoid rule:
  \[ c(k) \approx \frac{e^{-\nu k}}{\pi} \sum_{m=0}^{N} \delta_m e^{-imu_0} \chi^\ll_c(u_0 - i\nu) \Delta u. \]
  \[ \delta_k = \frac{1}{2} \text{ when } k = 0 \text{ and } 1 \text{ otherwise.} \]

- Set \( \eta = \Delta u, u_m = \eta m. \)

- Set \( k_j = -b + \lambda j \) with \( \lambda = \frac{2\pi}{(\eta N)} \) being the return grid and \( b \) being a parameter that controls the return range.
  
  - To center return around zero, set \( b = \lambda N/2. \)

- The call value becomes
  \[ c(k_j) \approx \frac{1}{N} \sum_{m=0}^{N-1} f_m e^{jm\frac{2\pi}{N}i}, \quad f_m = \delta_m \frac{N}{\pi} e^{-\nu k_j + iu_m b} \chi^\ll_c(u_m) \eta. \]

with \( j = 0, 1, \cdots, N - 1. \) The summation has the FFT form and can hence be computed efficiently.
To implement the FFT, we need to fix the following parameters:

- $N = 2^n$: The number of summation grids. Setting it to be the power of 2 can speed up the FFT calculation.

- $\eta = \Delta u$: The discrete summation grid width. The smaller the grid, the better the approximation.

However, given $N$, $\eta$ also determines the strike grid $\lambda = 2\pi/(\eta N)$. The finer the summation grid $\eta$, the coarser the strike spacing returned from the FFT calculation. There is a trade off: If we want to have more option value calculated at a finer grid of strikes, we would need to use a coarser summation grid and hence less accuracy.

The lower and upper bound truncation $b = \lambda N/2$ is also determined by the summation grid choice.

FFT generates option values at $N$ strikes simultaneously. However, if the strike grid is larger, many of the returned strikes are out of the relevant region.
III. Fractional FFT

- Fractional FFT (FRFT) separates the integration grid choice from the strike grids. With appropriate control, it can generate more accurate option values given the same amount of calculation.

- The method can efficiently compute,

\[
d_j = \sum_{m=0}^{N-1} f_m e^{-jm\zeta}, \quad j = 0, 1, ..., N - 1,
\]

for any value of the parameter \(\zeta\).

- The standard FFT can be seen as a special case for \(\zeta = \frac{2\pi}{N}\). Therefore, we can use the FRFT method to compute,

\[
c(k_j) \approx \frac{1}{N} \sum_{m=0}^{N-1} f_m e^{j m \eta \lambda}, \quad f_m = \delta_m \frac{N}{\pi} e^{-\nu k_j} e^{iu_m b} \chi_c(u_m) \eta.
\]

without the trade-off between the summation grid \(\eta\) and the strike spacing \(\lambda\).

- We require \(\eta \lambda = \frac{2\pi}{N}\) under standard FFT.
Let \( d = D(f, \zeta) \) denote the FRFT operation, with \( D(f) = D(f, 2\pi/N) \) being the standard FFT as a special case.

An \( N \)-point FRFT can be implemented by invoking three \( 2N \)-point FFT procedures.

Define the following \( 2N \)-point vectors:

\[
y = \left( \left( f_n e^{i\pi n^2 \zeta} \right)_{n=0}^{N-1}, (0)_{n=0}^{N-1} \right),
\]

\[
z = \left( \left( e^{i\pi n^2 \zeta} \right)_{n=0}^{N-1}, \left( e^{i\pi (N-n)^2 \alpha} \right)_{n=0}^{N-1} \right).
\]

The FRFT is given by,

\[
D_k(h, \zeta) = \left( e^{i\pi k^2 \zeta} \right)_{k=0}^{N-1} \odot D^{-1}_k \left( D_j(y) \odot D_j(z) \right),
\]

where \( D^{-1}_k(\cdot) \) denotes the inverse FFT operation and \( \odot \) denotes element-by-element vector multiplication.
Fractional FFT implementation

- Due to the multiple application of the FFT operations, an $N$-point FRFT procedure demands a similar number of elementary operations as a $4N$-point FFT procedure.

- Given the free choices on $\lambda$ and $\eta$, FRFT can be applied more efficiently. Using a smaller $N$ with FRFT can achieve the same option pricing accuracy as using a much larger $N$ with FFT.

- The accuracy improvement is larger when we have a better understanding of the model and model parameters so that we can set the boundaries more tightly.

- Caveat: The more freedom also asks for more discretion and caution in applying this method to generate robust results in all situations. This concern becomes especially important for model estimation, during which the trial model parameters can vary greatly.

- Reference: Chourdakis, 2005, Option pricing using fractional FFT, JCF, 8(2).
IV. Fourier-cosine series expansions

- Given a characteristic function $\phi(u)$, the density function can be numerically obtained via the Fourier-cosine series expansion,

$$ f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \phi(u) du \approx \sum_{j=0}^{N-1} \delta_j \cos ((x - a)u_j) V_j $$

where $u_j = \frac{j\pi}{b-a}, V_j = \frac{2}{b-a} \text{Re} \left[ \phi(u_j) e^{iu_ja} \right]$, and $[a, b]$ denotes a truncation of the return range. Choosing the range to be $\pm 10$ standard deviation away from the mean seems to work well: $b, a = \mu \pm 10\sigma$.

- Applying the expansion to the option valuation, we have

$$ C(K, t) \approx Ke^{-rt} \sum_{j=0}^{N-1} \delta_j \text{Re} \left[ \phi_s(u_j) e^{-iu_j(k+a)} U_j \right] $$

where $U_j = \frac{2}{b-a} \left( \chi_j(0, b) - \psi_j(0, b) \right)$ with

$$ \chi_j(c, d) = \frac{1}{1 + u_j^2} \left[ \cos((d-a)u_j)e^d - \cos((c-a)u_j)e^c + u_j \sin((d-a)u_j)e^d - u_j \sin((c-a)u_j)e^c \right], $$

$$ \psi_j(c, d) = \begin{cases} \frac{[\sin((d-a)u_j) - \sin((c-a)u_j)]}{u_j} & j \neq 0 \\ (d-c) & j = 0 \end{cases} $$

- Works well. Some constraints on how $[a, b]$ are chosen.

1. Design economically sensible & computationally feasible option pricing models based on time-changed Lévy processes

2. Efficient option pricing via Fourier inversions

3. Estimate option pricing models for different purposes
Estimating statistical dynamics

  - Given initial parameters guess, derive the return characteristic function.
  - Apply FFT to generate the probability density at a fine grid of possible return realizations.
  - Interpolate to obtain the density at the observed return values.
  - Numerically maximize the aggregate log likelihood.

- For time-changed Lévy processes with **observable** activity rates, it is still straightforward to apply MLE.

- For time-changed Lévy processes with **hidden** activity rates, some filtering technique is needed to infer the hidden states from the observable.
  - Maximum likelihood with partial filtering: Alireza Javaheri
  - MCMC Bayesian estimation: Eraker, Johannes, Polson (2003, JF), Li, Wells, Yu, (RFS)

- Use more data (and transformation) to turn hidden states into observable quantities. Wu (2007), Aït-Sahalia and Robert Kimmel (2007), Bondarenko (2007)…
Estimating risk-neutral dynamics

- **Daily fitting**: \((\text{Bakshi, Cao, Chen (1997, JF), Carr and Wu (2003, JF)})\)
  - Nonlinear weighted least square to fit models to option prices.
  - Parameters and state variables (activity rates) are treated as the same.
  - *What to hedge*: state variables or parameters or both.
  - Can experience identification issues for sophisticated models.
  - Better applied to Lévy processes without time change.

- **Dynamically consistent estimation**:
  - Parameters are fixed, only activity rates are allowed to vary over time.
  - Numerically more challenging.
  - Better applied to more sophisticated models that perform well over different market conditions.
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).

Different needs for different market participants:

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

- **Long-term convergence traders**:  
  - Dynamic consistency is important.  
  - A good model should generate large (we wish) but highly convergent pricing errors, and provide robust hedging ratios.

A well-designed model (with several time-changed Lévy components) can achieve both dynamic consistency and good performance.
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 \( P \)-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.
Kalman filter (KF) generates efficient forecasts and updates under linear-Gaussian state-space setup:

\[
\begin{align*}
\text{State:} & \quad X_{t+1} = A + \Phi X_t + \sqrt{Q}\epsilon_{t+1}, \\
\text{Measurement:} & \quad y_t = HX_t + \sqrt{\Sigma} e_t
\end{align*}
\]

The ex ante predictions as

\[
\begin{align*}
\bar{X}_t &= A + \Phi \hat{X}_{t-1}; \\
\bar{\Omega}_t &= \Phi \hat{\Omega}_{t-1} \Phi^\top + Q; \\
\bar{y}_t &= H\bar{X}_t; \\
V_t &= HV_tH^\top + \Sigma.
\end{align*}
\]

The ex post filtering updates are,

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

where \( K_{t+1} = \bar{\Omega}_{t+1} H^\top (V_{t+1})^{-1} \) is the Kalman gain.

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

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

  \[
  \begin{align*}
  \text{State:} & \quad X_{t+1} = A + \Phi X_t + \sqrt{Q_t} \epsilon_{t+1}, \\
  \text{Measurement:} & \quad y_t = h(X_t) + \sqrt{\Sigma} e_t
  \end{align*}
  \]

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

- 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 Kalman filter

- Let $k$ 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:

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

with corresponding weights $w_i$ given by

$$
w_0 = \delta/(k + \delta), \quad w_i = 1/[2(k + \delta)].
$$

- 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.

- Caveats:
  - 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.
Given the sigma points, the prediction steps are given by

\[
\bar{X}_{t+1} = A + \sum_{i=0}^{2k} w_i (\Phi \chi_{t,i});
\]

\[
\bar{\Omega}_{t+1} = \sum_{i=0}^{2k} w_i (A + \Phi \chi_{t,i} - \bar{X}_{t+1}) (A + \Phi \chi_{t,i} - \bar{X}_{t+1})^\top;
\]

\[
\bar{y}_{t+1} = \sum_{i=0}^{2k} w_i h (A + \Phi \chi_{t,i});
\]

\[
\bar{V}_{t+1} = \sum_{i=0}^{2k} w_i \left[ h (A + \Phi \chi_{t,i}) - \bar{y}_{t+1} \right] \left[ h (A + \Phi \chi_{t,i}) - \bar{y}_{t+1} \right]^\top + \Sigma,
\]

The filtering updates are given by

\[
\hat{X}_{t+1} = \bar{X}_{t+1} + K_{t+1} (y_{t+1} - \bar{y}_{t+1});
\]

\[
\hat{\Omega}_{t+1} = \bar{\Omega}_{t+1} - K_{t+1} \bar{V}_{t+1} K_{t+1}^\top,
\]

with \(K_{t+1} = S_{t+1} (\bar{V}_{t+1})^{-1}\).
Joint estimation of $\mathbb{P}$ and $\mathbb{Q}$ 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 ($\mathbb{Q}$) information.

  
  - Cast activity rate $\mathbb{P}$-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.
Why are we doing this?

- Understanding the $\mathbb{P}/\mathbb{Q}$ dynamics — Researchers are inherently curious.
- Understanding the human investment behavior: How investors price different sources of risks differently?
- Refine investment decisions.
  - Understand the sources of risks in each contract and the expected return per unit exposure to each risk source.
    - Hedge risk exposures, or
    - Take a controlled exposure to certain risk sources and receive risk premiums accordingly.
  - Exploit violations of no-arbitrage conditions.
    - Perform statistical arbitrage trading on derivative products that profit from short-term market dislocations.
    - Combine short-term prediction with market making.
Concluding remarks

Modeling security returns with time-changed Lévy processes enjoys three key virtues: (1) Generality; (2) explicit economic mapping; (3) tractability.

The framework provides a nice starting point for generating security return dynamics that are parsimonious, tractable, economically sensible, and statistically performing well.

It offers many opportunities... and computational challenges:

- Refine Fourier inversion schemes to generate the density and option values at the relevant region.
- Pricing exotic options
  - Solving multi-dimensional PIDEs in the presence of Lévy jumps.
  - Robust and efficient simulation of Lévy jump processes.
- Embed simulation into model estimation (MCMC, particle filter, pricing under new models, embed exotics into estimation).
Example: A half-analytical, half-simulation based model

- **Background:**
  - All the models I have shown you so far are “sticky-delta” type models. Volatility movements can be correlated with stock price movements, but volatility level does not depend on stock price level.
  - Dupire’s local-volatility model: Volatility is a function of stock price. Popular in the industry, rarely used in academia.

- **Reality:** Somewhere in between.
  - Price level matters over a certain range, ...
  - relative to the company’s business (earnings?) and debt level.

- **Our model:** \( F_t = A_t \chi_t = A_{T_t} X_{T_t} \)
  - \( F_t \) — forward equity level.
  - \( A_t \) — asset price level.
  - \( \chi_t = F_t / A_t \) — leverage.

\[
\begin{align*}
  dA_t / A_t &= -b (dZ_t + dJ_t), \\
  dX_t / X_t &= \delta X_t^{-p} dW_t, \quad \mathbb{E}[dZ_t dW_t] = 0, \\
  T_t &= \int_0^t \alpha_s ds, \quad d\alpha_t = \kappa (1 - \alpha_t) dt + \omega \alpha_t^\pi (dZ_{T_t} + dJ_{T_t})
\end{align*}
\]
Example: A half-analytical, half-simulation based model

- Option pricing:
  - Primer: option pricing under a CEV process \((X_t)\): 
    \[
    CEV \left( X_t, T, K \right) = \mathbb{E} \left[ (X_T - K)^+ \mid X_t \right] = X_t \left( 1 - \chi^2 \left( \zeta_1, 2 + \frac{1}{p}, \zeta_2 \right) \right) - K \chi^2 \left( \zeta_2, \frac{1}{p}, \zeta_1 \right),
    \]
    with \(\zeta_1 = \xi K^{2p}, \zeta_2 = \xi X_t^{2p}, \xi = \frac{1}{\delta^2 p^2 (T-t)}\).
  - Call option on \(F_t\),
    \[
    c \left( F_t, T, K \right) = \mathbb{E} \left[ (F_T - K)^+ \mid F_t, X_t, \alpha_t \right] \\
    = \mathbb{E} \mathbb{E} \left[ (A_T X_T - K)^+ \mid A_T = C, T_T = H, X_t = X_t \mid F_t, X_t, \alpha_t \right] \\
    = \mathbb{E} \mathbb{E} \left[ (CX_H - K)^+ \mid A_T = C, T_T = H, X_t = X_t \mid F_t, X_t, \alpha_t \right] \\
    = \mathbb{E} \mathbb{E} \left[ (X_{HC^{2p}} - K)^+ \mid A_T = C, T_T = H, X_t = CX_t \mid F_t, X_t, \alpha_t \right] \\
    = \mathbb{E} \left[ CEV \left( A_T X_t, T_T A_T^{2p}, K \right) \right]
    \]

- Combine simulation with analytical solution:
  - Simulate \(\alpha_t (W_t, J_t)\).
  - Store the cumulative quantities \(T_T, A_T\) for each simulated path.
  - Compute the option value analytically conditional on each simulated sample path.
  - Average the values over all sample paths.