Nonparametric methods for *financial models* driven by *Lévy processes*

José Enrique Figueroa-López

Special Colloquium

Purdue University

Department of Statistics

March 1st, 2006

Outline

I. Financial models for stock prices driven by Lévy processes

II. Challenges coming from Lévy-based modeling

III. Statistical properties of Lévy processes

IV. Methodology: approximating linear models, projection estimators, and model selection

V. Numerical illustrations

VI. Overview of theoretical properties: Long-run asymptotics and Adaptivity.

VII. Current and further work

Financial models

- The origins of quantitative mathematical finance is the Black-Scholes model for the pricing of *derivatives* on underlying stocks:

  Feasible objective pricing method based on the principle of the absence of arbitrage opportunities

- Important fundamental assumption: *The return (per dollar invested) of the stock during a small time span dt is normally distributed with constant mean and variance*:

$$\frac{dS_t}{S_t} = \underbrace{\mu}_{\text{Mean rate of return}} dt + \underbrace{\sigma}_{\text{Volatility}} d \underbrace{W_t}_{\text{B.M.}} .$$

  Equivalently,

$$S_t = S_0 e^{(\mu - \sigma^2/2)t + \sigma W_t} .$$

- Implications:
  - *Efficiency*: Future prices depends on the past only through the present value (Markov property). People cannot consistently predict the direction of an individual stock.
  - Log returns in disjoint periods are uncorrelated and normally distributed:

    Return on $[t, t + \Delta t) := \log \frac{S_{t+\Delta t}}{S_t} \sim N((\mu - \sigma^2/2)\Delta t, \sigma^2 \Delta t).$
  - Continuously varying stock prices
- The reality:
  - Distributions with *heavy tails* and *high kurtosis*.
  - Sudden changes in prices due to arrival of information

Financial model driven by Lévy processes

What is a Lévy process?

A random quantity X_t evolving in time $t \geq 0$ so that

- The change $X_{t+\Delta t} - X_t$ during the period $[t, t + \Delta t]$ is both (1) independent of the "past" $\{X_s : s \leq t\}$ and (2) with distribution law depending only on the time span Δt
- The process can exhibit sudden changes in magnitude (jumps), but these occur at unpredictable times (no fixed jump times).

Conclusion: Lévy process is the most natural generalization of Brownian motion, where only continuity of paths is relaxed, while preserving the statistical qualities of the increments.

A natural extension: Exponential Lévy models

$$dS_t = S_{t-}(\mu dt + d \underbrace{Z_t}_{\text{Lévy process}}) \iff S_t = S_0 \exp(\underbrace{X_t}_{\text{Levy process}}),$$

Time series of log returns: $r_i := \log\left(\frac{S_{hi}}{S_{h(i-1)}}\right) = X_{hi} - X_{h(i-1)}$, are independent and identically distributed

Why a Lévy Market?

- More flexible modeling of return distribution; e.g. *heavy-tails*, *asymmetry*, and *high kurtosis*
- Greater consistency with real stock dynamics, which is made up of discrete trades, and is exposed to sudden changes

Challenges coming from Lévy-based modeling

Statistical methodological issues

- Intractable (or not explicit) density functions for the log returns, necessitating approximations based on numerical inversion of the characteristic function
- Massive data coming from intraday (high-frequency) returns
- Too many parametric models around!! Normal, stable, hyperbolic distributions, tempered or truncated stable, etc. Which is better?
- Need for at least four parameters: *location, scale (volatility), decay of tails, and skewness*. Hard to guarantee convergence to Maximum Likelihood Estimators.

Conclusion: Computational demanding and numerically instable estimation

Empirical discrepancies [Cont: 2001]

- Volatility clustering: high-volatility events tend to cluster in time
- Leverage phenomenon: volatility is negatively correlated with returns
- Some sort of long-range memory: Returns do not exhibit significant autocorrelation (market efficiency); however, the autocorrelation of *absolute returns* decays slowly as a function of the time lag.

Conclusion: "Need" for increasingly more complex models

$$\log \frac{S_t}{S_0} = \begin{cases} \int_0^t \mu_u du + \int_0^t \sigma_u dW_u + \sum_{u \leq t} h(\Delta Z_u, u), & \text{or} \\ \mu T_t + \sigma W_{T_t}, & \text{or} \\ Z_{T_t}. & \end{cases}$$

Option pricing and hedging, and portfolio optimization?

Statistical properties of Lévy processes

Characterization and parameters

- The statistical law of the process is determined by distribution of X_t
- Three parameters, two reals σ^2, b, and a measure $\nu(dx)$, so that

$$(1) \quad X_t = bt + W_t + \lim_{\varepsilon \downarrow 0} \{X_t^\varepsilon - t m_\varepsilon\},$$

$$(2) \quad X^\varepsilon = \text{Compound Poisson Process with intensity } m_\varepsilon$$

$$(3) \quad N_t([a,b]) := \sum_{u \le t} \mathbf{1}[\Delta X_u \in [a,b]] \sim \text{Poisson}\,(t\nu([a,b])),$$

- **Lévy Density**: A nonnegative function s such that

$$\nu([a,b]) = \int_a^b s(x)dx,$$

Intuition: $s(x)$ dictates the frequency of jumps with sizes near to x

Important remarks

Necessary and sufficient conditions to be a Lévy density:

$$\int_{-1}^{1} s(x)x^2 dx < \infty \quad \text{and} \quad \int_{|x|>1} s(x)dx < \infty.$$

Consequence:

- A Lévy density is *not* a probability density function:

$$\int_{-\infty}^{\infty} s(x)dx = \infty \Rightarrow s(x) \xrightarrow{|x|\to 0} \infty \iff$$ inf. many jumps of arbitrarily small size

- It is easier to specify a Lévy process via the Lévy density than via the marginal distribution of X_1. Thus, it is easier to model the jump behavior of the process.

Summary

1. Exponential Lévy models are some of the simplest and most practical alternatives to the shortfalls of the geometric Brownian motion.

2. Capture several stylized empirical features of historical returns.

3. Limitations: Lack of stochastic volatility, leverage, quasi-long-memory, etc.

4. As a good “first-order” approximation model to other complex models, should be considered first in developing a successful estimation methodology

5. Statistical issues: computationally expensive and numerically instable to estimate by traditional likelihood-based methods

Nonparametric estimation of Lévy process

Problem: Estimate *directly* the Lévy density s relying only on qualitative assumptions on s.

Objectives

- Avoid model biases: “Let the data speak”.
- Computational efficiency, and suitability to deal with high-frequency

Sampling schemes:

- Continuous observations of the process over $[0, T]$

  Problem $\Longleftrightarrow$ estimation of non-homogeneous Poisson process

- Equally-spaced observations over $[0, T]$ with time span Δt

  ”Non-linear regression” with white noise $\Longleftrightarrow$

$$X_{t+\Delta t} - X_t = \underbrace{\text{Jump-part}(\nu)}_{\text{Jumps on } [t, t+\Delta t)} + \underbrace{\sigma(W_{t+\Delta t} - W_t)}_{\text{White noise}} + \underbrace{b\Delta t}_{\text{Drift}}$$

Why continuous-sampling?

- Results are more powerful, providing benchmarks for what can be accomplished by discrete-time-based methods
- Serve as devises to construct discrete-based procedures by approximating the underlying continuous-time-based statistics

Some related earlier work:

- [Rubin and Tucker, 1959]: Consistent estimation of $\nu\big((a,b)\big)$
- [Kuntoyants 1998, Reynaud-Bouret, 2003]: Consistent estimation of intensity for finite non-homogeneous Poisson processes. Oracle inequalities and adaptivity for smooth intensities
- [Figueroa-Lopez and Houdre, 2004-05] consistency, oracle inequalities, and adaptivity for smooth Lévy densities

Basic ideas of the model selection paradigm

- Approximation by <u>finite-dimensional</u> linear models:

$$s(x) \approx \beta_1 \varphi_1(x) + \cdots + \beta_n \varphi_n,$$

where the φ's are known functions. The space

$$\mathcal{S} := \{\beta_1 \varphi_1(x) + \cdots + \beta_n \varphi_n : \quad \beta_1, \ldots, \beta_n \quad \text{reals}\}$$

is called an (approximating) linear model.

- Typical examples: Histograms, splines, trigonometric polynomials, wavelets.

- Two problems to solve:

  Projection estimation: Aim at estimate an element of $\mathcal{S}$ "close" to s.

  Model selection: Determine a good approximating model from a collection of linear models.

Implementing the model selection approach

- Standing assumption: Existence of a regularizing *reference measure* η such that, on a given *estimation window* $D \subset \mathbb{R}\backslash\{0\}$,

$$(1) \quad \nu((a,b)) = \int_{(a,b)} s(x)\eta(dx),$$

$$(2) \quad s(x) \text{ is bounded on } D$$

$$(3) \quad \int_D s^2(x)\eta(dx) < \infty,$$

  – Typical case:

$$\eta(dx) = w^{-1}(x)dx \quad \text{and} \quad s(x) = (\text{Levy density}) \cdot w(x).$$

  – s *is said to be the Lévy density of the process on* D *with respect to the reference measure* η.

- Implications: The step of projection estimation is simplified

  If $\int_D \varphi_i^2(x)\,\eta(dx) < \infty$, one can always take $\varphi_1, \ldots, \varphi_n$ orthonormal on D with respect to η, and thus, the "closest" member of

$$\mathcal{S} := \{\beta_1\varphi_1(x) + \cdots + \beta_n\varphi_n : \quad \beta_1, \ldots, \beta_n \quad \text{reals}\}$$

  to the function s is when

$$\beta_i = \beta(\varphi_i) := \int_D \varphi_i(x)s(x)\eta(dx).$$

- Natural projection estimator on $\mathcal{S}$:

$$\hat{s}(x) := \hat{\beta}(\varphi_1)\varphi_1(x) + \cdots + \hat{\beta}(\varphi_n)\varphi_n(x),$$

  where

$$\hat{\beta}(\varphi) := \frac{1}{T}\sum_{t \le T} \varphi(\Delta X_t).$$

- Basic Properties
  - Unbiasedness:

    $\hat{s}$ is an unbiased estimator of the orthogonal projection of s on $\mathcal{S}$:

$$s^{\perp} = \beta(\varphi_1)\varphi_1(x) + \cdots + \beta(\varphi_n)\varphi_n(x).$$

  - The integrated-mean square error:

$$\mathbb{E}\|s^{\perp} - \hat{s}\|^2 = \frac{1}{T}\int_D \sum_i \varphi_i^2(x)s(x)dx \xrightarrow{T\to\infty} 0,$$

    Here, $\|f\|^2 := \int_D f^2(x)\eta(dx)$.

  - Standard "mean-variance decomposition":

$$\underbrace{\mathbb{E}\left[\|s - \hat{s}\|^2\right]}_{\text{Risk}} = \underbrace{\|s - s^{\perp}\|^2}_{\text{Bias term}} + \underbrace{\mathbb{E}\|s^{\perp} - \hat{s}\|^2}_{\text{Variance term}},$$

- Data-driven Model Selection:
  - Problem: Choose between two or more finite-dimensional linear models
  - Goal: Accomplish a good trade off between the two terms.
  - A standard solution: Choose the model that minimize an unbiased estimator of the risk. Since

$$\mathbb{E}\left[\|s-\hat{s}\|^2\right] = \|s\|^2 + \mathbb{E}\left[-\|\hat{s}\|^2 + \mathrm{pen}\left(\mathcal{S}\right)\right]$$

where

$$\mathrm{pen}\left(\mathcal{S}\right) := \frac{2}{T^2}\sum_{t\leq T}\bar{\varphi}_{\mathcal{S}}^2\left(\Delta X_t\right), \quad \text{and} \quad \bar{\varphi}_{\mathcal{S}}^2(x) := \sum_i \varphi_i^2(x),$$

a sensible criterion is to choose the model that minimizes the observable statistic

$$-\|\hat{s}\|^2 + \mathrm{pen}\left(\mathcal{S}\right).$$

- Implementation based on discrete observations

Problem: Estimate

$$\hat{\beta}(\varphi) := \frac{1}{T} \sum_{t \leq T} \varphi(\Delta X_t),$$

based on n equally spaced observations on $[0, T]$: $X_{T/n}, \ldots, X_{T}$.

A natural solution: Approximate by

$$\hat{\beta}_n(\varphi) := \frac{1}{T} \sum_{k=1}^{n} \varphi\left(X_{t_k} - X_{t_{k-1}}\right),$$

where $t_k = T\frac{k}{n}$, for $k = 1, \ldots, n$.

A Numerical Illustration

Gamma Lévy process with parameters α and β

- $X(t) \stackrel{\mathfrak{D}}{\sim} Gamma(\alpha t, \beta)$, β - scale parameter and α - shape parameter
- No Brownian part and Lévy density $s(x) = \frac{\alpha}{x} e^{-x/\beta}$, $x > 0$:

  Pure-jump increasing process with infinite jump-activity:

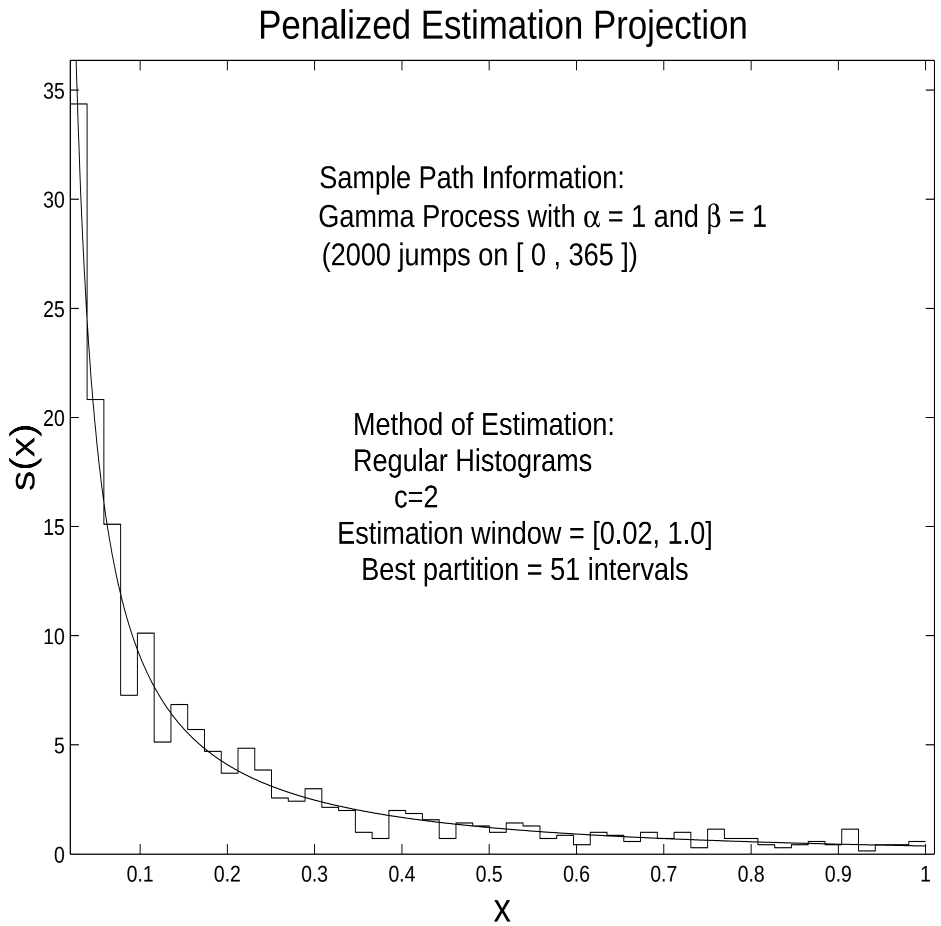
$$X_t = \sum_{k=1}^{\infty} \beta E_k e^{-\frac{1}{\alpha}\Gamma_k} \mathbf{1}_{[U_k \leq t]},$$

- Jump-behavioral interpretation:

  α - overall jump activity and β - frequency of big-jumps
- $\{cX(ht)\}_{t \geq 0}$ is Gamma with parameters αh and βc.

Model selection methods for the Gamma Lévy process

I. Histogram estimators (outside the origin):



- How good is the estimator?

  Fit the model $\frac{\alpha}{x} e^{-x/\beta}$ (using least-squares) to the histogram estimator: $\hat{\alpha}_{PPE} = 0.93$ and $\hat{\beta}_{PPE} = 1.055$ (vs. $\hat{\alpha}_{MLE} = 1.01$ and $\hat{\beta}_{MLE} = 0.94$)

- Estimation based on equally-spaced sampling with time span Δt:

Time span	PPE-LSF		MLE	
1	1.01	1.46	.997	.995
.5	1.03	1.09	.972	.978
.1	.944	.995	1.179	.837
.01	.969	.924	1.01	.98

(Based on 36500 jumps on $[0, 365] \approx$ 100 jumps/ unit time)

- Sampling distribution

Means and standard errors of $\hat{\alpha}_{PPE}$ and $\hat{\beta}_{PPE}$ based on 1000 repetitions

Δt	PPE-LSF		MLE	
.1	0.81 (0.06)	1.40 (0.50)	1.001 (0.01)	0.99 (0.05)
.01	0.92 (0.08)	1.12 (0.31)	1.007 (0.07)	0.99 (0.08)
.001	0.93 (0.08)	1.13 (0.34)	1.007 (0.07)	0.99 (0.08)

(simulations based on equally-spaced observations, with time span Δt, during the time interval $[0, 365]$)

II. Regularization technique around the origin

- Assumption: Levy density s is of the form

$$s(x) = \frac{1}{x} q(x), \quad 0 < x < h,$$

for a bounded nice function q and some $h > 0$.

- Approximate q by the linear model

$$\mathcal{S} := \{\beta_1 \varphi_1(x) + \cdots + \beta_d \varphi_d(x)\},$$

where the φ's are orthonormal functions defined on $D = [0, h]$:

$$\int_D \varphi_i(x)\varphi_j(x)dx = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

- Key observation:

$$\tilde{\varphi}_i(x) := x\varphi_i(x) \quad \Longrightarrow \quad \mathbb{E}\,[\sum_{t\leq T} \tilde{\varphi}_i(\Delta X_t)] = T \underbrace{\int_D \varphi_i(x)q(x)dx}_{\text{inner product}} .$$

- Unbiased estimator of the projection of q on $\mathcal{S}$:

$$\hat{q}(x) = \hat{\beta}(\tilde{\varphi}_1)\varphi_1(x) + \cdots + \hat{\beta}(\tilde{\varphi}_d)\varphi_d(x),$$

where

$$\hat{\beta}\left(\tilde{\varphi}\right) := \frac{1}{T}\sum_{t\leq T} \tilde{\varphi}(\Delta X_t).$$

- Penalized projection estimator can be devised by minimized an unbiased estimator of the risk

- Numerical illustration: Gamma Lévy process. $s(x) = \frac{\alpha}{x} e^{-x/\beta}$
  - Histogram approximations of $q(x) = \alpha e^{-x/\beta}$ on $D = (0, 3]$.
  - Least-square errors fit to the histogram estimator
  - 400 paths - each path generated from 36500 jumps on $[0, 365]$ with $\alpha = \beta = 1$
  - Jumps approximated by equally-spaced increments (time span Δt)

Δt	PPE-LSF	
.1	0.96 (0.14)	1.08 (0.17)
.01	1.05 (0.18)	0.98 (0.16)
.001	1.05 (0.17)	0.97 (0.16)

Table 1: Sampling mean and standard errors

Model selection methods for Tempered Stable Process

Model: [Rosinski 2005] No Brownian part with Lévy density

$$s(x) = \frac{1}{|x|^{\alpha+1}} q(|x|), \quad 0 < \alpha < 2,$$

where $q(x)$ and $q(-x)$ ($x > 0$) are bounded "completely monotone functions" (decreasing, convex; e.g. $\exp(x)$) vanishing for large x.

Appealing Property:

- $h^{-1/\alpha} X_{ht} \stackrel{\mathfrak{D}}{\approx}$ Stable distribution with parameter α, for h small.

  Stable-like high-frequency returns

- $h^{-1/2} X_{ht} \stackrel{\mathfrak{D}}{\approx}$ Normal distribution with parameter α, for h large.

  Normal-like small-frequency returns

Goal: Estimate q.

Projection estimator:

- Approximating linear model for q:

$$\mathcal{S} := \{\beta_1\varphi_1(x) + \cdots + \beta_d\varphi_d(x)\},$$

where the φ's are orthonormal functions defined on $D = [0, h]$.

- Estimator of the projection of q on $\mathcal{S}$:

$$\hat{q}(x) = \hat{\beta}(\tilde{\varphi}_1)\varphi_1(x) + \cdots + \hat{\beta}(\tilde{\varphi}_d)\varphi_d(x),$$

where

$$\hat{\beta}\,(\tilde{\varphi}) := \frac{1}{T}\sum_{t\leq T}\tilde{\varphi}(\Delta X_t) \quad \text{with} \quad \tilde{\varphi}(x) := |x|^{\hat{\alpha}+1}\varphi(x)$$

and $\hat{\alpha}$ being a “good” estimate of α; e.g. using methods for stable processes.

- Numerical illustration: One-tailed TS, $s(x) = \frac{a}{x^{\alpha+1}} e^{-x/b}$, $x > 0$
  - Histogram approximations of $q(x) = ae^{-x/b}$ on $D = [0, 3]$
  - Paths generated from 36500 jumps on $[0, 365]$ with $a = b = 1$ and $\alpha = .1$.
  - Jumps approximated by equally-spaced increments (time span Δt)
  - α estimated by Zolotarev method for stable distributions (valid for $h \approx 0$) based on the moments of $\log |X_t|$.
  - Least-square fit to quantify the quality of the histogram estimator

Δt	Penalized Projection - Least-Squares Fit			Misspecified Gamma MLE	
.01	1.03 (0.15)	0.97 (0.14)	0.09 (0.0002)	1.2 (0.08)	0.89 (0.079)

Table 2: Sampling mean and standard errors (sample size=100 paths)

Overview of theoretical properties

Penalized projection estimators: Suppose $\mathcal{M}$ is a collection of models.

Oracle Inequality.

For each $\mathcal{S} \in \mathcal{M}$, let $\mathcal{B}_{\mathcal{S}}$ be an orthonormal basis for $\mathcal{S}$ and for a fixed time horizon T, let

$$\mathcal{M}_T := \{\mathcal{S} \in \mathcal{M} : \| \sum_{\varphi \in \mathcal{B}_{\mathcal{S}}} \varphi^2(\cdot)\|_\infty \leq T\}.$$

Then, the penalized projection estimator $\tilde{s}_T$ on $\mathcal{M}_T$ satisfies

$$\mathbb{E}\left[\|s - \tilde{s}_T\|^2\right] \leq C \inf_{\mathcal{S} \in \mathcal{M}_T} \mathbb{E}\left[\|s - \hat{s}_{\mathcal{S}}\|^2\right] + \frac{C'}{T},$$

where C and C' are independent of T and the linear models.

Long-run rate of convergence for smooth densities:

Assume that on the estimation window $D := [a, b]$ (not containing 0),

$$|s^{(k)}(x) - s^{(k)}(y)| \leq L|x - y|^{\gamma},$$

for some $k \in \{0, 1, \ldots\}$ and $\gamma \in (0, 1]$. Then, fixing $\alpha := k + \gamma$,

$$\lim_{T \to \infty} T^{2\alpha/(2\alpha+1)} \mathbb{E}\left[\|s - \tilde{s}_T\|^2\right] < \infty,$$

where the projection is on regular splines of degree at most k on $[a, b]$.

Comparison to minimax risks:

$$\liminf_{T \to \infty} T^{2\alpha/(2\alpha+1)} \left\{ \inf_{\hat{s}_T} \sup_{s \in \Theta} \mathbb{E}\left[\|\hat{s}_T - s\|^2\right] \right\} > 0, \qquad (1)$$

where the infimum is over all estimators $\hat{s}_T$ based on the jumps of the Lévy process $\{X(t)\}_{0 \leq t \leq T}$ with sizes falling on $[a, b]$.

Approximate projection estimators based on discrete observations:

Basic Properties:

Suppose that $D := [a, b] \subset \mathbb{R} \backslash \{0\}$

1. $\hat{\beta}_n(\varphi) \quad \xrightarrow{\mathfrak{D}} \quad \hat{\beta}(\varphi)$, as $n \to \infty$

2. $\lim_{n\to\infty} \mathbb{E}\left[\hat{\beta}_n(\varphi)\right] = \beta(\varphi) := \int \varphi(x) s(x) \eta(dx)$

3. $\lim_{n\to\infty} \mathrm{Var}\left[\hat{\beta}_n(\varphi)\right] = \frac{1}{T}\beta\left(\varphi^2\right)$.

4. If $\hat{s}^{(n)}$ is the approximate projection estimator of s on a given linear model $\mathcal{S}$, based on n discrete regular observations on $[0, T]$,

$$\lim_{n\to\infty} \mathbb{E}\left[\|\hat{s}^{(n)} - s\|^2\right] = \mathbb{E}\left[\|\hat{s} - s\|^2\right].$$

Summary

We developed estimation and model selection schemes for the Lévy density of a Lévy process:

- Flexible: it can be used histograms, splines, wavelets, etc.
- Model free
- Easily implementable
- Reliable and robust: Oracle inequality and adaptivity; i.e. asymptotically comparable to minimax estimators on classes of smooth Lévy densities

Current and future work

(1) Investigate asymptotics of non-parametric *discrete-data based estimators*, as well as their comparisons to the long-run minimax risk. These asymptotics will be as both the frequency of the observations and the time horizon increase.

(2) Extend the minimax and rate of convergence results when considering arbitrarily small jumps and when estimating around the origin, where the Lévy density usually blows up.

(3) Confidence bands and test of hypothesis

(4) Apply similar ideas to more realistic models such as time-changed Lévy processes and SDE driven by Lévy processes. For instance, estimation of the functional driving both the random clock and the jump process.

Some references

- [Figueroa-Lopez]
  1. Nonparametric estimation of Lévy processes with a view towards mathematical finance. [Thesis 2004].
  2. Risk bounds for the non-parametric estimation of Lévy processes, submitted 2005 (with C. Houdré)
  3. Non-parametric estimation of tempered stable processes. [Preprint.]
- Methodology described here was motivated by works on the estimation of pdf's by [Birgé & Massart (1994,1997)], and on estimation of intensity functions of Poisson processes by [Reynaud-Bouret (2003)].
- Minimax results were based on similar methods by [Kutoyants (1998)] and Ibraginov & Has'miskii (1981)]