Lecture 4

Disentangling Diffusion from Jumps

Yacine Aït-Sahalia
Princeton University

based in part on joint work with Jean Jacod

References

- Disentangling Diffusion from Jumps, *Journal of Financial Economics*, 2004, 74, 487-528.

- Volatility Estimators for Discretely Sampled Lévy Processes, with Jean Jacod, *Annals of Statistics*, 2007, 35, 355-392.

- Fisher's Information for Discretely Sampled Lévy Processes, with Jean Jacod, *Econometrica*, 2008, 76, 727-761.

1. Introduction

Basic question asked in the paper: how does the presence of jumps impact our ability to estimate the diffusion parameter σ^2?

- I start by *presenting* some intuition that seems to suggest that the identification of σ^2 is hampered by the presence of the jumps...

- But, surprisingly, maximum-likelihood can actually perfectly disentangle Brownian noise from jumps provided one samples frequently enough.

- I first show this result in the context of a compound Poisson process, i.e., a jump-diffusion as in Merton (1976).

- One may wonder whether this result is driven by the fact that Poisson jumps share the dual characteristic of being large and infrequent.

- Is it possible to perturb the Brownian noise by a Lévy pure jump process other than Poisson, and still recover the parameter σ^2 as if no jumps were present?

- The reason one might expect this not to be possible is the fact that, among Lévy pure jump processes, the Poisson process is the only one with a finite number of jumps in a finite time interval.

- All other pure jump processes exhibit an infinite number of small jumps in any finite time interval.

- Intuitively, these tiny jumps ought to be harder to distinguish from Brownian noise, which it is also made up of many small moves.

- Perhaps more surprisingly then, I find that maximum likelihood can still perfectly discriminate between Brownian noise and a Cauchy process.

- Every Lévy process can be uniquely expressed as the sum of three independent canonical Lévy processes:

  1. A continuous component: Brownian motion (with drift);
  2. A "big jumps" component in the form of a compound Poisson process having only jumps of size greater than one;
  3. A "small jumps" component in the form of a pure jump martingale having only jumps of size smaller than one;

- So the two examples considered in this paper represent the prototypical cases of:

  1. Distinguishing the Brownian component from the "big jumps" component;

  2. Distinguishing the Brownian component from an example of the class of "small jumps" components.

- I show later how this generalizes to all Lévy pure jump processes.

- I also look at the extent to which GMM estimators using absolute moments of various non-integer orders can recover the efficiency of maximum-likelihood

- The answer is no, but they do better than traditional moments such as the variance and kurtosis.

- Beyond the econometrics, why should one care about being able to decompose the noise in the first place?

  1. In option pricing, the two types of noise have different hedging requirements and possibilities;

  2. In portfolio allocation, the demand for assets subject to both types of risk can be optimized further if a decomposition of the total risk into a Brownian and a jump part is available;

  3. In risk management, such a decomposition makes it possible over short horizons to manage the Brownian risk using Gaussian tools while assessing VaR and other tail statistics based on the identified jump component.

2. The Model and Setup

Example: the simple Merton (1976) Poisson jump-diffusion model

$$dX_t = \mu dt + \sigma dW_t + J_t dN_t$$

- X_t denotes the log-return derived from an asset, W_t a Brownian motion and N_t a Poisson process with arrival rate λ.

- The log-jump size J_t is $N(\beta, \eta)$.

- The density exhibits skewness (if the jumps are asymmetric) and excess kurtosis

3. Intuition

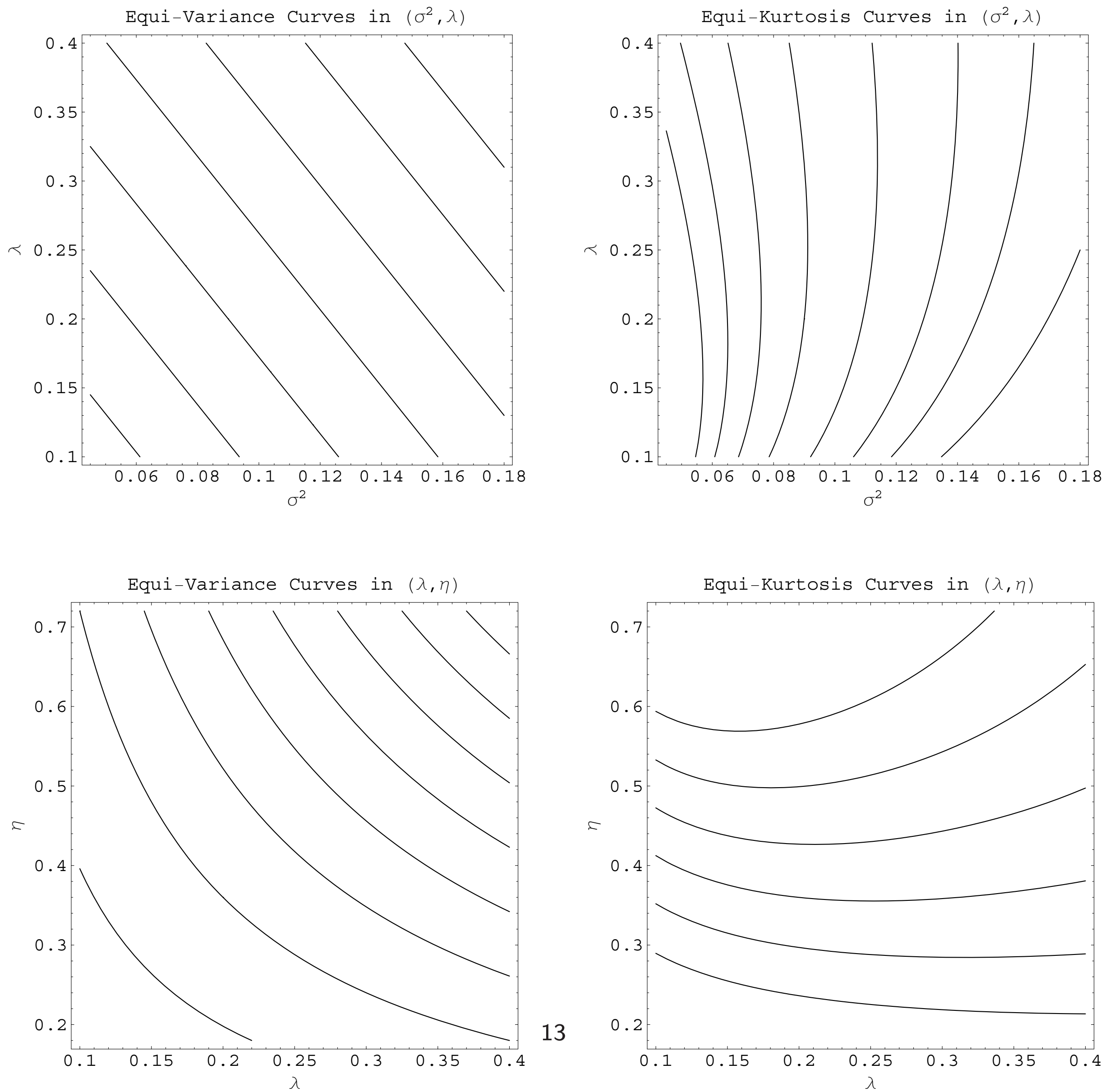
3.1. Isonoise Curves

- The first intuition I provide is based on the traditional method of moments, combined with non-linear least squares.

- In NLLS, the asymptotic variance of the estimator is proportional to the inverse of the partial derivative of the moment function (or conditional mean) with respect to the parameter.

- Consider what can be called isonoise curves. These are combinations of parameters of the process that result in the same observable conditional variance of the log returns; excess kurtosis is also included.

- Intuitively, any two combinations of parameters on the same isonoise curve cannot be distinguished by the method of moments using these moments.


Equi-Variance Curves in (σ^2, λ)
λ
σ^2
0.1 0.15 0.2 0.25 0.3 0.35 0.4
0.06 0.08 0.1 0.12 0.14 0.16 0.18
Equi-Kurtosis Curves in (σ^2, λ)
λ
σ^2
0.1 0.15 0.2 0.25 0.3 0.35 0.4
0.06 0.08 0.1 0.12 0.14 0.16 0.18
Equi-Variance Curves in (λ, η)
η
λ
0.2 0.3 0.4 0.5 0.6 0.7
0.1 0.15 0.2 0.25 0.3 0.35 0.4
Equi-Kurtosis Curves in (λ, η)
η
λ
0.2 0.3 0.4 0.5 0.6 0.7
0.1 0.15 0.2 0.25 0.3 0.35 0.4

3.2. Inferring Jumps from Large Realized Returns

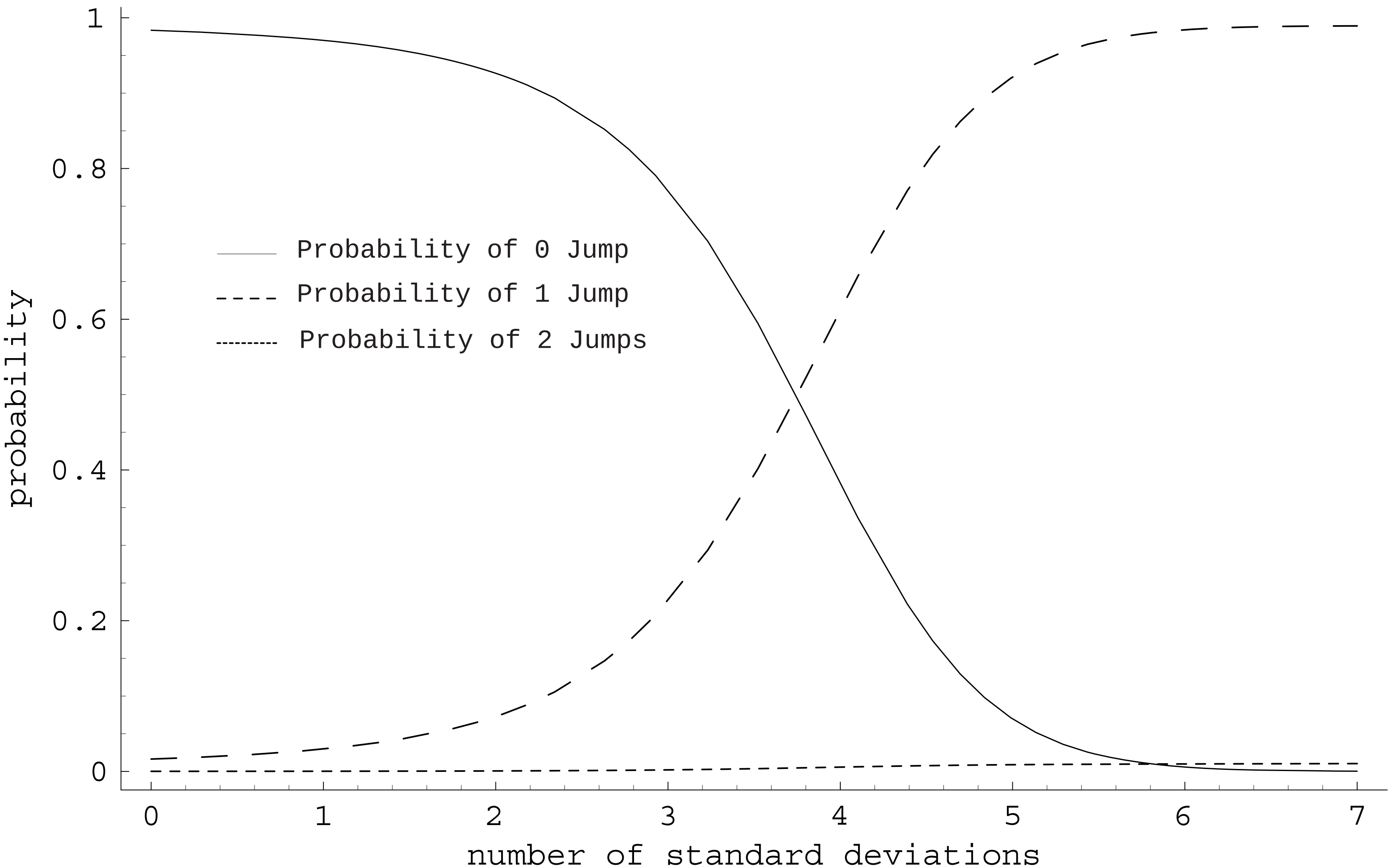
- In discretely sampled data, every change in the value of the variable is by nature a discrete jump

- Given that we observe in discrete data a change in the asset return of a given magnitude z or larger, what does that tell us about the likelihood that such a change involved a jump (as opposed to just a large realization of the Brownian noise)?

- To investigate that question, let's use Bayes' Rule:

$$\Pr\left(B_\Delta = 1 \mid Z_\Delta \geq z; \theta\right) = \Pr\left(Z_\Delta \geq z \mid B_\Delta = 1; \theta\right) \frac{\Pr\left(B_\Delta = 1; \theta\right)}{\Pr\left(Z_\Delta \geq z; \theta\right)}$$

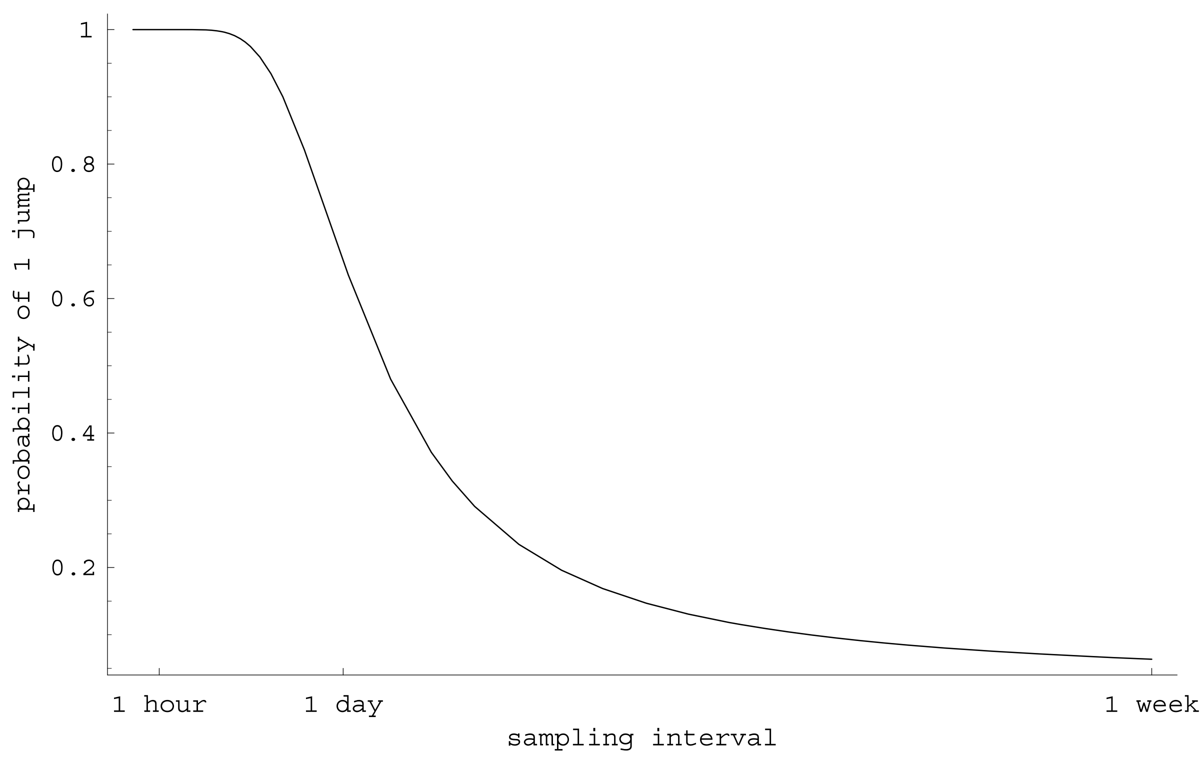
$$= \frac{e^{-\lambda\Delta}\lambda\Delta\left(1 - \Phi\left(\frac{z-\mu\Delta-\beta}{2\left(\eta+\Delta\sigma^2\right)^{1/2}}\right)\right)}{\sum_{n=0}^{+\infty}\frac{e^{-\lambda\Delta}(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{z-\mu\Delta-n\beta}{2\left(n\eta+\Delta\sigma^2\right)^{1/2}}\right)\right)}$$


1
0.8
0.6
0.4
0.2
0
probability
Probability of 0 Jump
Probability of 1 Jump
Probability of 2 Jumps
0
1
2
3
4
5
6
7
number of standard deviations

- The figure shows that as far into the tail as 4 standard deviations, it is still more likely that a large observed log-return was produced by Brownian noise only.

- Since these moves are unlikely to begin with (and hence few of them will be observed in any given series of finite length), this underscores the difficulty of relying on large observed returns as a means of identifying jumps.

- This said, our ability to visually pick out the jumps from the sample path increases with the sampling frequency:



- But our ability to infer the provenance of the large move tails off very quickly as we move from Δ equal to 1 minute to 1 hour to 1 day.

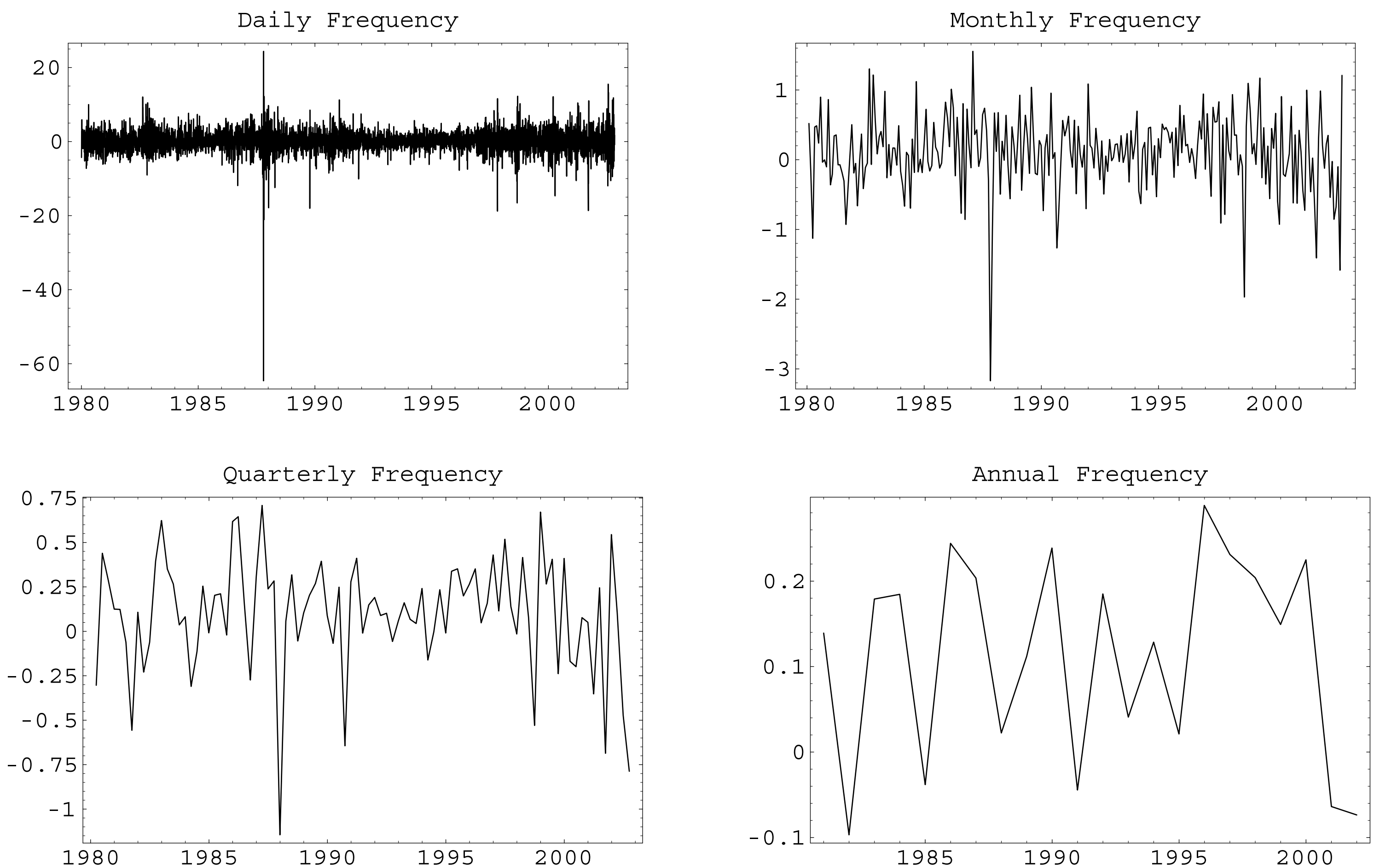
- At some point, enough time has elapsed that the 10% move could very well have come from the sum over the time interval $(0, \Delta)$ of all the tiny Brownian motion moves.

3.3. The Time-Smoothing Effect

- The final intuition for the difficulty in telling Brownian noise apart from jumps lies in the effect of time aggregation, which in the present case takes the form of time smoothing.

- Just like a moving average is smoother than the original series, log returns observed over longer time periods are smoother than those observed over shorter horizons. In particular, jumps get averaged out.

- This effect can be severe enough to make jumps visually disappear from the observed time series of log returns.


Daily Frequency
20
0
-20
-40
-60
1980 1985 1990 1995 2000
Monthly Frequency
1
0
-1
-2
-3
1980 1985 1990 1995 2000
Quarterly Frequency
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
1980 1985 1990 1995 2000
Annual Frequency
0.2
0.1
0
-0.1
1985 1990 1995 2000

4. Disentangling Diffusion from Jumps Using the Likelihood

- The time-smoothing effect suggests that our best chances of disentangling the Brownian noise from the jumps lie in high frequency data.

Theorem 1: When the Brownian motion is contaminated by Poisson jumps,

$$\text{AVAR}_{\text{MLE}}\left(\sigma^2\right) = 2\sigma^4\Delta + o(\Delta)$$

so that in the limit where sampling occurs infinitely often ($\Delta \longrightarrow 0$), the MLE estimator of σ^2 has the same asymptotic distribution as if no jumps were present.

- Theorem 1 says that maximum-likelihood can in theory perfectly disentangle σ^2 from the presence of the jumps, when using high frequency data.

- The presence of the jumps imposes no cost on our ability to estimate σ^2 : the variance is σ^2, not the total variance $\sigma^2 + (\beta^2 + \eta)\lambda$.

- This can be contrasted with what would happen if, say, we contaminated the Brownian motion with another Brownian motion with known variance s^2. In that case, we could also estimate σ^2, but the asymptotic variance of the MLE would be $2\left(\sigma^2 + s^2\right)^2 \Delta$.

- In light of the Cramer Rao lower bound, Theorem 1 establishes $2\sigma^4\Delta$ as the benchmark for alternative methods (based on the quadratic variation, absolute variation, GMM, etc.)

5. How Close Does GMM Come to MLE?

- I form moment functions of the type $h(y, \delta, \theta) = y^r - E[Y^r]$ and/or $h(y, \delta, \theta) = |y|^r - E[|Y|^r]$ for various values of r.

- By construction, these moment functions are unbiased and all the GMM estimators considered will be consistent.

- The question becomes one of comparing their asymptotic variances among themselves, and to that of MLE.

5.1. Moments of the Process

The first four conditional moments of the process X are $E\left[Y_{\Delta}\right] = \Delta\left(\mu + \beta\lambda\right)$ and, with

$$M\left(\Delta, \theta, r\right) \equiv E\left[\left(Y_{\Delta} - \Delta\left(\mu + \beta\lambda\right)\right)^{r}\right]$$

we have

$$M\left(\Delta, \theta, 2\right) = \Delta\left(\sigma^{2} + (\beta^{2} + \eta)\lambda\right)$$
$$M\left(\Delta, \theta, 3\right) = \Delta\lambda\beta\left(\beta^{2} + 3\eta\right)$$
$$M\left(\Delta, \theta, 4\right) = \Delta\left(\beta^{4}\lambda + 6\beta^{2}\eta\lambda + 3\eta^{2}\lambda\right) + 3\Delta^{2}\left(\sigma^{2} + (\beta^{2} + \eta)\lambda\right)^{2}$$

5.2. Absolute Moments of Non-Integer Order

- Consider the quadratic variation of the X process

$$[X,X]_t = \text{plim}_{n\to\infty} \sum_{i=1}^{n} \left(X_{t_i} - X_{t_{i-1}}\right)^2$$

- We have

$$\begin{aligned}[X,X]_t &= [X,X]_t^c + \sum_{0\leq s\leq t} (X_s - X_{s-})^2 \\ &= \sigma^2 t + \sum_{i=1}^{N_t} J_{s_i}^2\end{aligned}$$

- Not surprisingly, the quadratic variation no longer estimates σ^2 in the presence of jumps.

- However, Lepingle (1976) studied the behavior of the power variation of the process, i.e., the quantity

$$_r\left[X,X\right]_t = \text{plim}_{n\to\infty} \sum_{i=1}^{n} \left| X_{t_i} - X_{t_{i-1}} \right|^r$$

and showed that the contribution of the jump part to $_r\left[X,X\right]_t$ is, after normalization, zero when $r \in (0,2)$, $\sum_{i=1}^{N_t} J_{s_i}^2$ when $r = 2$ and infinity when $r > 2$.

- These results suggest that it will be useful to consider absolute moments (i.e., the plims of the power variations) when forming GMM moment conditions.

Proposition 1: For any $r \geq 0$, the centered absolute moment of order r is:

$$M_a(\Delta, \theta, r) \equiv E\left[|Y_\Delta - \Delta(\mu + \beta\lambda)|^r\right]$$

$$= \sum_{n=0}^{\infty} \frac{1}{\pi^{1/2} n!} e^{-\lambda\Delta - \frac{(n\beta - \Delta\beta\lambda)^2}{2\left(\Delta\sigma^2 + n\eta\right)}} \left(n\eta + \sigma^2\Delta\right)^{r/2} (\lambda\Delta)^n$$

$$\times\, 2^{r/2}\Gamma\left(\frac{1+r}{2}\right) F\left(\frac{1+r}{2}, \frac{1}{2}, \frac{\beta^2(n-\Delta\lambda)^2}{2\left(n\eta + \sigma^2\Delta\right)}\right)$$

where Γ denote the gamma function and F denotes the Kummer confluent hypergeometric function ${}_1F_1(a, b, \omega)$.

In particular, when $\beta = 0$, $F\left(\frac{1+r}{2}, \frac{1}{2}, 0\right) = 1$.

5.3. Estimating σ^2 Alone

I find that, although it does not restore full maximum likelihood efficiency, using absolute moments in GMM helps.

1. When σ^2 is estimated using classical moments, then $\text{AVAR}_{\text{GMM}}\left(\sigma^2\right) = O(1)$, a full order of magnitude bigger than achieved by MLE.

2. When absolute moments of order $r \in (0,1)$ are used, $\text{AVAR}_{\text{GMM}}\left(\sigma^2\right) = O(\Delta)$, i.e., the same order as MLE, although the constant of proportionality is always greater than $2\sigma^4$.

3. When σ^2 is estimated based on absolute moments of order $r \in (1,2]$, $\text{AVAR}_{\text{GMM}}\left(\sigma^2\right) = O(\Delta^{2-r})$.

Proposition 2: AVAR of GMM Estimators of σ^2

Moment(s)	$\text{AVAR}_{\text{GMM}}\left(\sigma^2\right)$ with jumps	$\text{AVAR}_{\text{GMM}}\left(\sigma^2\right)$ no jumps
$M(\Delta,\theta,2)$	$3\eta^2\lambda+2\Delta\left(\sigma^2+\eta\lambda\right)^2$	$2\Delta\sigma^4$
$\begin{pmatrix} M(\Delta,\theta,2) \\ M(\Delta,\theta,4) \end{pmatrix}$	$\frac{6\eta^2\lambda}{7}+\Delta\left(2\sigma^4+\frac{44\eta^2\lambda^2}{7}+\frac{100\eta\lambda\sigma^2}{49}\right)+o(\Delta)$	$2\Delta\sigma^4$
$M_a(\Delta,\theta,r),\ r\in(0,1)$	$\Delta\frac{4\sigma^4}{r^2}\left(\frac{\pi^{1/2}\Gamma\left(\frac{1}{2}+r\right)}{\Gamma\left(\frac{1+r}{2}\right)^2}-1\right)+o(\Delta)$	$\Delta\frac{4\sigma^4}{r^2}\left(\frac{\pi^{1/2}\Gamma\left(\frac{1}{2}+r\right)}{\Gamma\left(\frac{1+r}{2}\right)^2}-1\right)+o(\Delta)$
$M_a(\Delta,\theta,1)$	$2\Delta\sigma^2\left((\pi-2)\sigma^2+\pi\eta\lambda\right)$	$2(\pi-2)\Delta\sigma^4$
$M_a(\Delta,\theta,r),\ r\in(1,2]$	$\Delta^{2-r}\frac{4\pi^{1/2}\eta^r\lambda\sigma^{2(2-r)}\Gamma\left(\frac{1}{2}+r\right)}{r^2\Gamma\left(\frac{1+r}{2}\right)^2}+o(\Delta^{2-r})$	$\Delta\frac{4\sigma^4}{r^2}\left(\frac{\pi^{1/2}\Gamma\left(\frac{1}{2}+r\right)}{\Gamma\left(\frac{1+r}{2}\right)^2}-1\right)+o(\Delta)$
$\begin{pmatrix} M(\Delta,\theta,2) \\ M_a(\Delta,\theta,1) \end{pmatrix}$	$2\Delta\sigma^2\left((\pi-2)\sigma^2+\frac{(3\pi-8)}{3}\eta\lambda\right)+o(\Delta)$	$2\Delta\sigma^4$

6. Other Jump Processes: The Cauchy Case

- The result so far has been the ability of maximum-likelihood to fully distinguish the diffusive component from the Poisson jump component, as shown in Theorem 1.

- I now examine whether this phenomenon is specific to the fact that the jump process considered so far was a compound Poisson process, or whether it extends to other types of jump processes.

6.1. The Cauchy Pure Jump Process

- A process is a Lévy process if it has stationary and independent increments and is continuous in probability.

- The log-characteristic function of a Lévy process is given by the Lévy-Khintchine formula:

$$\psi(u) = i\gamma u - \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{+\infty} \left(e^{iuz} - 1 - iuzc(z)\right) \nu(dz).$$

- γ is the drift rate of the process, σ its volatility from its continuous (Brownian) component

- The Lévy measure $\nu(\cdot)$ describes the pure jump component: $\nu(E)$ for any subset $E \subset \mathbb{R}$ is the rate at which the process takes jumps of size $x \in E$, i.e., the number of jumps of size falling in E per unit of time. $\nu(\cdot)$ satisfies $\int_{-\infty}^{+\infty} Min\left(1, z^2\right) \nu(dz) < \infty$.

- Is it possible to perturb the Brownian noise by a Lévy pure jump process other than Poisson, and still recover the parameter σ^2 as if no jumps were present?

- The reason one might expect this not to be possible is the fact that, among Lévy pure jump processes, the Poisson process is the only one with a finite $\nu([-\infty, +\infty])$, i.e., a finite number of jumps in a finite time interval.

- All other pure jump processes are such that $\nu([-\varepsilon, +\varepsilon]) = \infty$ for any $\varepsilon > 0$, so that the process exhibits an infinite number of small jumps in any finite time interval.

- Intuitively, these tiny jumps ought to be harder to distinguish from Brownian noise, which it is also made up of many small moves.

- I will consider as an example the Cauchy process, which is the pure jump process with Lévy measure

$$\nu(dx) = \frac{\xi}{x^2} dx$$

- This is an example of a symmetric stable distribution of index $\alpha \leq 2$ and rate $\xi > 0$. The Cauchy process corresponds to $\alpha = 1$, while the limit $\alpha \to 2$ produces a Gaussian distribution.

6.2. Mixing Cauchy Jumps with Brownian Noise

- So I now look at the situation where

$$dX_t = \mu dt + \sigma dW_t + dC_t$$

  where C_t is a Cauchy process independent of the Brownian motion W_t.

- Theorem 2: When the Brownian motion is contaminated by Cauchy jumps, it still remains the case that

$$\text{AVAR}_{MLE}\left(\sigma^2\right) = 2\sigma^4\Delta + o(\Delta).$$

6.3. How Small are the Small Jumps?

- Theorem 2 shows that Cauchy jumps do not come close enough to mimicking the behavior of the Brownian motion to reduce the accuracy of the MLE estimator of σ^2.

- The intuition behind this is the following:

  – While there is an infinite number of small jumps in a Cauchy process, this "infinity" remains relatively small (just like the cardinality of the set of integers is smaller than the cardinality of the set of reals)

  – And while the jumps are infinitesimally small, they remain relatively bigger than the increments of a Brownian motion during the same time interval Δ.

Formally:

- If Y_Δ is the log-return from a pure Brownian motion, then

$$\Pr(|Y_\Delta| > \varepsilon) = \frac{\Delta^{1/2}\sigma}{\varepsilon}\left(\frac{2}{\pi}\right)^{1/2}\exp\left(-\frac{\varepsilon^2}{2\Delta\sigma^2}\right)(1+o(1))$$

is exponentially small as $\Delta \to 0$.

- However, if Y_Δ results from a Lévy pure jump process with jump measure $v(dz)$, then

$$\Pr(|Y_\Delta| > \varepsilon) = \Delta \times \int_{|y|>\varepsilon} v(dy) + o(\Delta)$$

which decreases only linearly in Δ.

- In other words, Lévy pure jump processes will always produce moves of size greater than ε at a rate that is small but nevertheless far greater than Brownian motion.

- Do jumps always have to behave that way? Yes, because of Ray's Theorem (1956): the sample paths of a Markov process are almost surely continuous iff, for every $\varepsilon > 0$,

$$\Pr\left(|Y_\Delta| > \varepsilon\right) = o(\Delta)$$

- We have seen that: Brownian has $\Pr\left(|Y_\Delta| > \varepsilon\right) = o(e^{-1/\Delta})$ and Lévy pure jump processes $\Pr\left(|Y_\Delta| > \varepsilon\right) = O(\Delta)$.

7. Extension: All Lévy Processes Which Can Be Disentangled from Diffusion

- Intuitively, there must be a limit to how many small jumps can occur in a finite amount of time for this to continue to hold. This question is examined in joint work with Jean Jacod.

- Let the Fisher information at stage n for σ^2 be $nI(\sigma^2, \Delta, G)$, where G denotes the law of the pure jump process. G can be arbitrary in the set $\mathcal{G}$ of all infinitely divisible law with vanishing Gaussian part.

- The closure of $\mathcal{G}$ (for the weak convergence) contains all Gaussian laws:
  - So if the convergence were uniform in $G \in \mathcal{G}$ it would hold as well when the Lévy process is a Browian motion with variance, say, s^2.
  - Then the best one can do is to estimate $\sigma^2 + s^2$.

- So the idea now is to restrict the set $\mathcal{G}$ to a subset which lies at a positive distance of all Gaussian laws ($\alpha = 2$):
  - For any index $\alpha \in [0, 2]$ define $\mathcal{G}(\alpha)$ as the family of jump measures such that for some $K > 0$:
$$\nu([-\varepsilon, \varepsilon]^c) \leq K \left(1 \vee \frac{1}{\varepsilon^\alpha}\right) \quad \forall\, \varepsilon > 0.$$
  - A stable law of index$\alpha < 2$ belongs to $\mathcal{G}(\alpha)$.
  - Any infinitely divisible law without Gaussian part belongs to $\mathcal{G}(2)$.
  - If $G \in \mathcal{G}(0)$ then Y is a compound Poisson process plus a drift.

- The following two results give respectively the uniformity of the convergence on the set $\mathcal{G}(\alpha)$, $\alpha < 2$, and the lack of uniformity otherwise:

1. For all K and $\alpha \in [0,2)$ and $\sigma^2 > 0$ we have

$$\sup_{G \in \mathcal{G}(\alpha)} \left(\frac{1}{2\sigma^4} - I(\sigma^2, \Delta, G) \right) \to 0 \qquad \text{as} \ \ \Delta \to 0.$$

2. For each n let G^n be the symmetric stable law with index $\alpha_n \in (0,2)$ and scale parameter $s^2/2$. Then if $\alpha_n \to 2$, for all sequences $\Delta_n \to 0$ satisfying $(2-\alpha_n)\log\Delta_n \to 0$ we have

$$I(\sigma^2, \Delta_n, G^n) \to \frac{1}{2(\sigma^2 + s^2)^2}.$$

- Therefore the result continues to hold for all Lévy pure jump processes, no matter how close the fixed value of α is to 2.

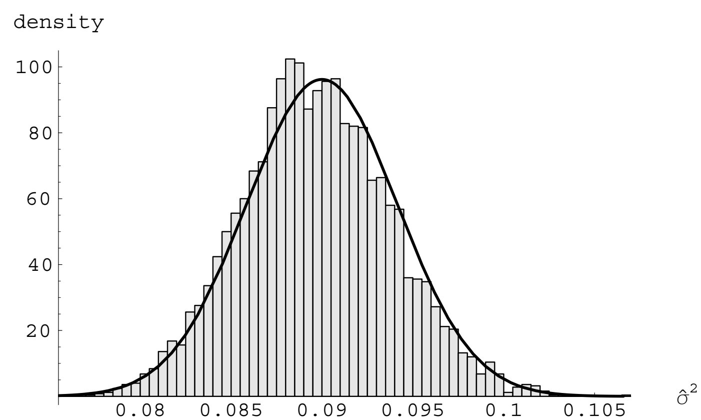
- But, if one lets $\alpha \longrightarrow 2$ at the same time as $\Delta \longrightarrow 0$ then it is no longer possible to perfectly distinguish the Brownian and jump components:
  - This is to be expected since the pure jump component converges to Brownian motion.
  - The information behaves in the limit like the superposition of two Gaussian processes: you cannot identify the variance of one if you only observe the sum of the two.

8. Monte Carlo Simulations

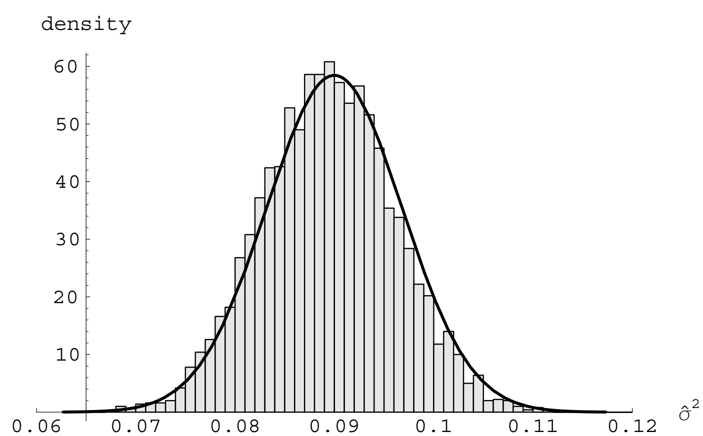
- $5,000$ simulations of the jump-diffusion, each of length $n = 1,000$ at the daily frequency.

- I then estimate the parameters using MLE.

- Small sample and asymptotic distributions for σ^2 in the Poisson case



- Small sample and asymptotic distributions for σ^2 in the Cauchy case



9. Conclusions

- MLE can perfectly disentangle Brownian noise from jumps provided one samples frequently enough.

- True for a compound Poisson process, i.e., a jump-diffusion. But also for Cauchy jumps.

- Extends to all Lévy jump processes that stay at a finite distance from $\alpha = 2$.

- GMM estimators using absolute moments of various non-integer orders do better than traditional moments such as the variance and kurtosis.