NBER WORKING PAPER SERIES

DISENTANGLING VOLATILITY FROM JUMPS

Yacine Aït-Sahalia

Working Paper 9915
http://www.nber.org/papers/w9915

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2003

Financial support from the NSF under grant SES-0111140 is gratefully acknowledged. I am also grateful to Arnaud Godard and Jialin Yu for excellent research assistance. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Disentangling Volatility from Jumps
Yacine Aït-Sahalia
NBER Working Paper No. 9915
August 2003
JEL No. G12, C22

ABSTRACT

Realistic models for financial asset prices used in portfolio choice, option pricing or risk management include both a continuous Brownian and a jump components. This paper studies our ability to distinguish one from the other. I find that, surprisingly, it is possible to perfectly disentangle Brownian noise from jumps. This is true even if, unlike the usual Poisson jumps, the jump process exhibits an infinite number of small jumps in any finite time interval, which ought to be harder to distinguish from Brownian noise, itself made up of many small moves.

Yacine Aït-Sahalia
Department of Economics
and Bendheim Center for Finance
26 Prospect Avenue
Princeton, NJ 08540-5296
and NBER
yacine@princeton.edu

1. Introduction

From an asset pricing perspective, being able to decompose the total amount of noise into a continuous Brownian part and a discontinuous jump part is useful in a number of contexts: for instance, in option pricing, the two types of noise have different hedging requirements and possibilities; 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; 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. In fact, the ability to disentangle jumps from volatility is the essence of risk management, which should focus on controlling large risks leaving aside the day-to-day Brownian fluctuations. This paper shows that it is possible to use likelihood-based statistical methods to distinguish volatility from jumps with (asymptotically) perfect accuracy, thereby focusing on the part of the overall risk that should be the object of concern in risk management or asset allocation.

The fact that jumps play an important role in many variables in finance, such as asset returns, interest rates or currencies, as well as a sense of diminishing marginal returns in studies of the "simple" diffusive case, has led to a flurry of recent activity dealing with jump processes. This activity has developed in three broad directions: estimating ever more complex and realistic financial models incorporating jumps (see e.g., Schaumburg (2001) using maximum-likelihood, Eraker et al. (2003) using MCMC, Chernov et al. (2002) using EMM, and the references therein), testing from discrete data whether jumps are present (see Aït-Sahalia (2002b) using a characterization of the transition function of a diffusion and Carr and Wu (2003b) using short dated options) and studying the behavior of interesting statistics, such as the quadratic variation and related quantities, in the presence of jumps (see Barndorff-Nielsen and Shephard (2002)).

The present paper asks a different yet basic question, which, despite its importance and apparent simplicity, appears to have been overlooked in the literature: 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, before showing that 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 model 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 is itself 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, a canonical example of such processes.

Indeed, while the early use on jumps in finance has focused exclusively on Poisson jumps (see Press (1967), Merton (1976), Beckers (1981) and Ball and Torous (1983)), the literature is rapidly moving towards incorporating other types of Lévy processes, such as the Cauchy jumps which are considered here. This is the case either for theoretical option pricing (see e.g., Madan et al. (1998), Chan (1999) and Carr and Wu (2003a)), risk management (see e.g., Eberlein et al. (1998)), or as a means of providing more accurate description of asset returns data (see e.g., Carr et al. (2002)). In term structure modelling, different Central Bank policies can give rise to different types of jumps and recent models do also allow for Lévy jumps other than Poisson (see e.g., Eberlein and Raible (1999)). Given this literature, the results I present provide statistical support for the use of non-Poisson jump processes: in the various contexts that matter in finance, it is possible to mix such jump processes with the usual Brownian volatility and still distinguish one from the other.

The paper is organized as follows: in Section 2, I briefly present the basic Poisson model, before giving in Section 3 different types of intuition which all suggest that it would be difficult to distinguish the volatility from the jumps. In Section 4, I show that the intuition is actually misleading, at least in the Poisson case. I then look in Section 5 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). The next question is whether any of this is specific to Poisson jumps. I show that this is not the case by studying more general Lévy pure jump processes in Section 6. Finally, I present in Section 7 Monte Carlo evidence to show that the asymptotic results in the theorems of the previous sections provide a close approximation to the behaviors that we are likely to encounter in daily data. Section 8 concludes. All proofs are in the Appendix.

2. The Model and Setup

Most of the points made in this paper are already apparent in a Poisson-based jump-diffusion model. So, for clarity, I will start with the simple Merton (1976) model where the jump term is Poisson-driven; in this section, I collect a number of useful results about this basic model. Later, I will turn to the more complex situation where the jump term is Cauchy-driven. Consider for now the jump-diffusion specification

$$dX_t = \mu dt + \sigma dW_t + J_t dN_t \tag{2.1}$$

where X_t denotes the log-return derived from an asset. W_t denotes a standard Brownian motion and N_t a Poisson process with arrival rate λ. The log-jump size J_t is a Gaussian random variable with mean β and variance η. By Itô's Lemma, the corresponding model for the asset price $S_t = S_0 \exp(X_t)$ is

$$\frac{dS_t}{S_t} = \left(\mu + \sigma^2/2\right) dt + \sigma dW_t + \left(\exp\left(J_t\right) - 1\right) dN_t \tag{2.2}$$

For further simplicity, assume that W_t, N_t and J_t are independent stochastic processes. As noted above, extensions to dependent drift, diffusion, and jump arrival intensity functions, as well as to other distributions of the jump size J_t, pose no conceptual difficulties but are notationally more cumbersome with little associated gain. The objective is to study our ability to estimate the parameter vector $\theta = (\mu, \sigma^2, \lambda, \beta, \eta)'$, where μ is the drift of the Brownian process, σ the volatility of the Brownian process, λ the arrival rate of the Poisson process, β the average size of the jumps J and η their variance. θ is an unknown parameter in a bounded set $\Theta \subset \mathbb{R}^5$. I focus in particular on our ability to distinguish information about the diffusive part (σ^2) from information about the jump part (λ, η), the respective means (μ, β) being largely inconsequential.

2.1. *The Transition Density*

The transition density for the model under consideration has a known form which I briefly review. The solution of the stochastic differential equation (2.1) is

$$Z_\Delta = X_\Delta - X_0 = \mu\Delta + \sigma W_\Delta + \int_0^\Delta J_s dN_s \tag{2.3}$$

which implies in particular that, for this simple model, the log-returns are i.i.d. This is a consequence of the assumptions made that the parameters and distribution of the jump term are state-independent. Since the distribution of the Poisson process is discrete, and

$$\Pr(N_\Delta = n; \theta) = \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!}.$$

Conditioning on the number of possible jumps between 0 and Δ and applying Bayes' Rule, we have

$$\Pr(X_\Delta \le x | X_0 = x_0, \Delta; \theta) = \sum_{n=0}^{+\infty} \Pr(X_\Delta \le x \,|\, X_0 = x_0, \Delta, N_\Delta = n; \theta) \times \Pr(N_\Delta = n; \theta).$$

Conditioned upon the event $N_\Delta = n$, there must have been exactly n times, say τ_i, $i = 1, ..., n$, between 0 and Δ such that $dN_{\tau_i} = 1$. Thus

$$\int_0^\Delta J_s dN_s = \sum_{i=1}^n J_{\tau_i}$$

is the sum of n independent jump terms. Under the assumption that each one has the distribution $J \sim N(\beta, \eta)$, it follows that the transition density of X_Δ given X_0 is given by

$$\begin{aligned} p(x|x_0, \Delta; \theta) &= \sum_{n=0}^{+\infty} p(x|x_0, \Delta, N_\Delta = n; \theta) \Pr(N_\Delta = n; \theta) \\ &= \sum_{n=0}^{+\infty} \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{\sqrt{2\pi}\sqrt{n\eta + \Delta\sigma^2} n!} \exp\left(-\frac{(x - x_0 - \mu\Delta - n\beta)^2}{2(n\eta + \Delta\sigma^2)}\right). \end{aligned} \tag{2.4}$$

As expected in the presence of jumps, the density exhibits excess kurtosis: see Figures 1 and 2 (at $\Delta = 1/12$ with parameters $\mu = \beta = 0$, $\sigma = 0.3, \lambda = 0.2$ and $\eta^{1/2} = 0.6$). Early examples of the use of this or similar formulae for maximum likelihood in finance are contained in Press (1967), Beckers (1981) and Ball and Torous

(1983). A non-zero value of the mean jump size β would add skewness. Note that, for purposes of maximum-likelihood estimation, care must be taken to ensure that the mixture of normals remains bounded by properly restricting the admissible parameters. Otherwise, setting the mean of one of the elements to be exactly equal to the observations, the variance parameter of that element can be driven to zero thereby increasing the likelihood to arbitrarily high levels (see Kiefer (1978) and Honoré (1998) for further discussion).

2.2. *Moments of the Process*

The first four conditional moments of the process X were calculated by Press (1967) using the transition density. They 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

$$\begin{aligned} M\left(\Delta, \theta, 2\right) &= \Delta\left(\sigma^2 + (\beta^2 + \eta)\lambda\right) \\ M\left(\Delta, \theta, 3\right) &= \Delta\lambda\beta\left(\beta^2 + 3\eta\right) \\ M\left(\Delta, \theta, 4\right) &= \Delta\left(\beta^4\lambda + 6\beta^2\eta\lambda + 3\eta^2\lambda\right) + 3\Delta^2\left(\sigma^2 + (\beta^2 + \eta)\lambda\right)^2 \end{aligned} \tag{2.5}$$

More generally, to evaluate moments of the process, for (2.1) and more complex stochastic differential equations, let A denote the infinitesimal generator of the process X, defined by its action on functions $f\left(\delta, x, x_0\right)$ in its domain:

$$\begin{aligned} A \cdot f\left(\Delta, x, x_0\right) &= \frac{\partial f\left(\Delta, x, x_0\right)}{\partial \Delta} + \mu\frac{\partial f\left(\Delta, x, x_0\right)}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f\left(\Delta, x, x_0\right)}{\partial x^2} \\ &\quad + \lambda E_J\left[f(\Delta, x + J, x_0) - f(\Delta, x, x_0)\right]. \end{aligned} \tag{2.6}$$

To evaluate a conditional expectation, I use the Taylor expansion

$$E\left[f(\Delta, X_\Delta, X_0) | X_0 = x_0\right] = \sum_{k=0}^{K} \frac{\Delta^k}{k!} A^k \cdot f(\delta, x, x_0)_{|\, x = x_0, \delta = 0} + O\left(\Delta^{K+1}\right) \tag{2.7}$$

In all cases, this expression is a proper Taylor series (as in Aït-Sahalia (2002a)); whether the series is analytic at $\Delta = 0$ is not guaranteed. In the present case, the moments of the process of integer order all lead to a finite series, which is therefore exact: applying (2.7) to $f(\delta, x, x_0) = (x - x_0)^i$, $i = 1, ..., 4$, yields exact expressions.

2.3. *Absolute Moments of Non-Integer Order*

It turns out that the absolute value of the log returns is less sensitive than the quadratic variation to large deviations, which makes them suitable in the context of high frequency data with the possibility of jumps. This has been noted by e.g., Ding et al. (1993) and Andersen and Bollerslev (1997). Consider the quadratic variation of the X process

$$[X,X]_t = \text{plim}_{n\to\infty} \sum_{i=1}^{n} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \tag{2.8}$$

for any increasing sequence $0 = t_0, ..., t_n = t$. We have

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

where s_i, $i = 1, ..., N_t$ denote the jump dates of the process, with the continuous part of the quadratic variation given by $[X,X]_t^c = \sigma^2 t$ and $X_s - X_{s-} = J_s (N_s - N_{s-})$ (i.e., X only jumps when N jumps; when N jumps, it jumps by one unit).

Not surprisingly, the quadratic variation in this case no longer estimates σ^2 (see e.g., the related discussion in Andersen et al. (2003)). However, Lepingle (1976) studied the behavior of the power variation of the process, i.e., the quantity

$${}_r[X,X]_t = \text{plim}_{n\to\infty} \sum_{i=1}^{n} \left|X_{t_i} - X_{t_{i-1}}\right|^r \tag{2.9}$$

and showed that the contribution of the jump part to ${}_r[X,X]_t$ is, after normalization, zero when $r \in (0,2)$, $\sum_{i=1}^{N_t} J_{s_i}^2$ when $r = 2$ and infinity when $r > 2$. Barndorff-Nielsen and Shephard (2002) use this result to show that the full ${}_r[X,X]_t$ depends only on the diffusive component when $r \in (0,2)$. They also compute the asymptotic distribution of the sample analog of ${}_r[X,X]_t$ constructed from discrete approximations to the continuous-time process.

I will use these insights when forming GMM moment conditions to estimate the parameters in the presence of jumps, with the objective of studying their ability to reproduce the efficiency of MLE. I will consider in particular absolute moments of order r (i.e., the plims of the power variations). To form unbiased moment conditions, I will need an exact expression for these moments, which is given in the following result:

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

$$\begin{aligned} M_a(\Delta,\theta,r) &\equiv E\left[\left|Y_\Delta - \Delta(\mu+\beta\lambda)\right|^r\right] \\ &= \sum_{n=0}^{\infty} \frac{2^{r/2}\Gamma\left(\frac{1+r}{2}\right) F\left(\frac{1+r}{2}, \frac{1}{2}, \frac{\beta^2(n-\Delta\lambda)^2}{2(n\eta+\sigma^2\Delta)}\right)\left(n\eta+\sigma^2\Delta\right)^{r/2}(\lambda\Delta)^n}{\pi^{1/2} n!} e^{-\lambda\Delta - \frac{(n\beta-\Delta\beta\lambda)^2}{2(\Delta\sigma^2+n\eta)}} \end{aligned}$$

where Γ denote the gamma function and F denotes the Kummer confluent hypergeometric function ${}_1F_1(a,b,\omega)$.[1]

[1] See Chapter 13 in Abramowitz and Stegun (1972) for definitions and properties of ${}_1F_1$.

In particular, when $\beta = 0$, $F\left(\frac{1+r}{2}, \frac{1}{2}, 0\right) = 1$. *The expansion of* $M_a(\Delta, X_0, r)$ *in* Δ *is, at the leading order,*

$$M_a(\Delta, \theta, r) = \begin{cases} \pi^{-1/2} 2^{r/2} \Gamma\left(\frac{1+r}{2}\right) \sigma^r \Delta^{r/2} + o(\Delta^{r/2}) & \text{if } r < 2 \\ \left(\sigma^2 + \left(\beta^2 + \eta\right) \lambda\right) \Delta & \text{if } r = 2 \\ \pi^{-1/2} 2^{r/2} \eta^{r/2} \lambda \Gamma\left(\frac{1+r}{2}\right) H\left(\frac{1+r}{2}, \frac{1}{2}, \frac{\beta^2}{2\eta}\right) e^{-\frac{\beta^2}{2\eta}} \Delta + o(\Delta) & \text{if } r > 2 \end{cases}$$

3. Intuition for the Difficulty in Identifying the Parameters

Before turning to the formal study of estimators in the context of this model, I describe intuitively in this section why distinguishing the volatility parameter from the jump component could be expected to be difficult.

3.1. *Isonoise Curves*

The first intuition I provide is based on the traditional method of moments, combined with non-linear least squares. We know that, in the nonlinear least squares context, 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. In other words, if small changes in the parameter value result in large changes in the moment function then the parameter will be estimated precisely. If on the other hand large changes in the parameter result in small changes in the moment function, then the parameter will not be estimated precisely.

I plot in Figure 3 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. These are the curves $E\left[(X_\Delta - X_0)^2 |X_0\right] =$ constant and $E\left[(X_\Delta - X_0)^4 |X_0\right] =$ constant, with the other parameters fixed. Intuitively, any two combinations of parameters on the same isonoise curve cannot be distinguished by the method of moments using these moments. (An additional issue is that in practice kurtosis is estimated with little precision.) The top row of the figure looks at distinguishing σ^2 from λ, the bottom one at distinguishing λ from η (in the figure, $\Delta = 1/12$ and the other parameters are $\mu = \beta = 0$, $\eta^{1/2} = 0.6$ in the top row and $\sigma = 0.3$ in the bottom one). Combinations of the two parameters (λ, η) that are on the same isonoise curve result in the same amount of "jumpiness" from the perspective of these two moments. This analysis provides further arguments for including moments other than the variance and kurtosis in an a GMM-type setting (see Section 5 below).

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, yet we wish to estimate jointly from these data the underlying continuous-time parameters driving the Brownian and jump terms. So the next question I examine, still with the objective of providing some intuition, is the following: given that we observe in discrete data an 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 to calculate

$$\begin{aligned}
\Pr\left(B_\Delta = 1 \,|\, Z_\Delta \geq z; \theta\right) &= \frac{\Pr\left(Z_\Delta \geq z,\, B_\Delta = 1; \theta\right)}{\Pr\left(Z_\Delta \geq z; \theta\right)} \\
&= \Pr\left(Z_\Delta \geq z \,|\, B_\Delta = 1; \theta\right) \frac{\Pr\left(B_\Delta = 1; \theta\right)}{\Pr\left(Z_\Delta \geq z; \theta\right)} \\
&= \frac{\exp\left(-\lambda\Delta\right)\lambda\Delta\left(1 - \Phi\left(\frac{z-\mu\Delta-\beta}{2(\eta+\Delta\sigma^2)^{1/2}}\right)\right)}{\sum_{n=0}^{+\infty}\frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{z-\mu\Delta-n\beta}{2(n\eta+\Delta\sigma^2)^{1/2}}\right)\right)}
\end{aligned}$$

since

$$\begin{aligned}
\Pr\left(B_\Delta = 1; \theta\right) &= \exp\left(-\lambda\Delta\right)\lambda\Delta \\
\Pr\left(Z_\Delta \geq z; \theta\right) &= \int_z^{+\infty} q(z, \Delta; \theta) dz = \sum_{n=0}^{+\infty} \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{z - \mu\Delta - n\beta}{2\left(n\eta + \Delta\sigma^2\right)^{1/2}}\right)\right) \\
\Pr\left(Z_\Delta \geq z \,|\, B_\Delta = 1; \theta\right) &= 1 - \Phi\left(\frac{z - \mu\Delta - \beta}{2\left(\eta + \Delta\sigma^2\right)^{1/2}}\right)
\end{aligned}$$

where Φ denotes the Normal cdf and

$$q\left(y, \Delta; \theta\right) \equiv p(x_0 + y | x_0, \Delta; \theta). \tag{3.1}$$

The probability of seeing more than one jump is:

$$\begin{aligned}
\Pr\left(B_\Delta \geq 1 \,|\, Z_\Delta \geq z; \theta\right) &= \Pr\left(Z_\Delta \geq z \,|\, B_\Delta \geq 1; \theta\right) \frac{\Pr\left(B_\Delta \geq 1; \theta\right)}{\Pr\left(Z_\Delta \geq z; \theta\right)} \\
&= \frac{\sum_{n=1}^{+\infty}\frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{z-\mu\Delta-n\beta}{2(n\eta+\Delta\sigma^2)^{1/2}}\right)\right)}{\sum_{n=0}^{+\infty}\frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{z-\mu\Delta-n\beta}{2(n\eta+\Delta\sigma^2)^{1/2}}\right)\right)}
\end{aligned}$$

since

$$\begin{aligned}
\Pr\left(Z_\Delta \geq z \,|\, B_\Delta \geq 1; \theta\right) &= \frac{\Pr\left(Z_\Delta \geq z\,,\, B_\Delta \geq 1; \theta\right)}{\Pr\left(B_\Delta \geq 1; \theta\right)} \\
&= \frac{\sum_{n=1}^{+\infty}\Pr\left(Z_\Delta \geq z\,,\, B_\Delta = n; \theta\right)}{\Pr\left(B_\Delta \geq 1; \theta\right)} \\
&= \sum_{n=1}^{+\infty}\Pr\left(Z_\Delta \geq z \,|\, B_\Delta = n; \theta\right)\frac{\Pr\left(B_\Delta = n; \theta\right)}{\Pr\left(B_\Delta \geq 1; \theta\right)}.
\end{aligned}$$

Then

$$\Pr\left(B_\Delta = 0 \,|\, Z_\Delta \geq z; \theta\right) = 1 - \Pr\left(B_\Delta \geq 1 \,|\, Z_\Delta \geq z; \theta\right).$$

If we look at jumps of a given size, irrespective of the direction, then:

$$\begin{aligned}\Pr\left(B_\Delta = 1 \,|\, |Z_\Delta| \geq z; \theta\right) &= \frac{\Pr\left(|Z_\Delta| \geq z,\, B_\Delta = 1; \theta\right)}{\Pr\left(|Z_\Delta| \geq z; \theta\right)} \\ &= \frac{\Pr\left(Z_\Delta \geq z,\, B_\Delta = 1; \theta\right) + \Pr\left(Z_\Delta \leq -z,\, B_\Delta = 1; \theta\right)}{\Pr\left(Z_\Delta \geq z; \theta\right) + \Pr\left(Z_\Delta \leq -z; \theta\right)} \\ &= \frac{\left(\Pr\left(Z_\Delta \geq z \,|\, B_\Delta = 1; \theta\right) + \Pr\left(Z_\Delta \leq -z \,|\, B_\Delta = 1; \theta\right)\right)\Pr\left(B_\Delta = 1; \theta\right)}{\Pr\left(Z_\Delta \geq z; \theta\right) + \Pr\left(Z_\Delta \leq -z; \theta\right)} \\ &= \frac{\left(1 - \Phi\left(\frac{z - \mu\Delta - \beta}{2(\eta + \Delta\sigma^2)^{1/2}}\right) + \Phi\left(\frac{-z - \mu\Delta - \beta}{2(\eta + \Delta\sigma^2)^{1/2}}\right)\right)\exp\left(-\lambda\Delta\right)\lambda\Delta}{\sum_{n=0}^{+\infty} \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{z - \mu\Delta - n\beta}{2(n\eta + \Delta\sigma^2)^{1/2}}\right) + \Phi\left(\frac{-z - \mu\Delta - n\beta}{2(n\eta + \Delta\sigma^2)^{1/2}}\right)\right)}\end{aligned}$$

If the process is symmetric ($\mu = \beta = 0$), then

$$\Phi\left(\frac{-z}{2\left(n\eta + \Delta\sigma^2\right)^{1/2}}\right) = 1 - \Phi\left(\frac{z}{2\left(n\eta + \Delta\sigma^2\right)^{1/2}}\right)$$

and it makes no difference whether we condition on $|Z_\Delta| \geq z$ or $Z_\Delta \geq z$, in which case:

$$\begin{aligned}\Pr\left(B_\Delta = 1 \,|\, |Z_\Delta| \geq z; \theta\right) &= \frac{2\left(1 - \Phi\left(\frac{z}{2(\eta + \Delta\sigma^2)^{1/2}}\right)\right)\exp\left(-\lambda\Delta\right)\lambda\Delta}{\sum_{n=0}^{+\infty} \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!} 2\left(1 - \Phi\left(\frac{z}{2(n\eta + \Delta\sigma^2)^{1/2}}\right)\right)} \\ &= \frac{\left(1 - \Phi\left(\frac{z}{2(\eta + \Delta\sigma^2)^{1/2}}\right)\right)\exp\left(-\lambda\Delta\right)\lambda\Delta}{\sum_{n=0}^{+\infty} \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{z}{2(n\eta + \Delta\sigma^2)^{1/2}}\right)\right)}\end{aligned}$$

and similarly for $\Pr\left(B_\Delta \geq 1 \,|\, |Z_\Delta| \geq z; \theta\right)$.

Figure 4 plots the functions $\Pr\left(B_\Delta = 1 \,|\, Z_\Delta \geq z; \theta\right)$ (as well as the matching probabilities of zero and two jumps) evaluated at $z = u\Delta^{1/2}\left(\sigma^2 + (\beta^2 + \eta)\lambda\right)^{1/2}$, so u measures the size of the log-return observed in terms of number of standard deviations away from the mean, at the same parameter values as above. The figure shows that as far into the tail as 3.5 standard deviations, it is still more likely that a large observed log-return was produced by Brownian noise only (since the probability of zero jump is higher than that of one jump). Since 3.5 standard deviation 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. Implicit in these calculations is also the interaction between the unconditional arrival rate of the jumps and our ability to properly identify a large move as having been generated by a jump: if the unconditional jump probability is low, then it takes an even bigger observed log-return before we can assign its origin to a jump.

This said, it is intuitively clear that our ability to visually pick out the jumps from the sample path increases when the time interval Δ between successive observations on the path decreases. Figure 5 shows this effect by plotting the dependence of $\Pr\left(B_\Delta = 1 \,|\, |Z_\Delta| \geq z; \theta\right)$, evaluated at $z = 10\%$, as a function of the sampling interval Δ. The smaller Δ, the higher the probability that an observed log-return of fixed magnitude 10% of greater was caused by a jump. Note however from the figure that 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. As an example, albeit extreme but real world, of this phenomenon, consider the effect of the 1987 crash on the Dow Jones Industrials Average. As Figure 6 shows, there was no 1987 crash as far as the annual data were concerned. Of course, the crash is quite visible at higher frequencies, the more so the higher the frequency.

4. Disentangling the Diffusion from the Jumps Using the Likelihood

Armed with these various intuitions, I now turn to the question of determining formally what is the effect of the presence of the jumps on our ability to estimate the value of σ^2. The ability to pick out jumps from the sample path as well as the time-smoothing effect suggest that our best chance of disentangling the Brownian noise from the jumps lies in high frequency data. I will show that it is actually possible to recover the value of σ^2 with the *same degree of precision* as if there were no jumps and the only source of noise were the Brownian motion. In other words, the various intuitions suggesting otherwise are misleading, at least in the limit of infinitely frequent sampling.

4.1. *Asymptotics*

Since I will consider both likelihood and non-likelihood types of estimators below, I embed both types into the GMM framework. Let $Y_{n\Delta} = X_{n\Delta} - X_{(n-1)\Delta}$ denote the first differences of the process X. They are i.i.d. under this simple model. To estimate the d-dimensional parameter vector θ, consider a vector of m moment conditions $h(y, \delta, \theta)$, $m \geq d$,continuously differentiable in θ ($\dot{h}$ denotes the gradient of h with respect to θ). Then form the sample average

$$m_T(\theta) \equiv N^{-1} \sum_{n=1}^{N} h(Y_{n\Delta}, \Delta, \theta) \tag{4.1}$$

and obtain $\hat{\theta}$ by minimizing the quadratic form

$$Q_T(\theta) \equiv m_T(\theta)' \, G_T \, m_T(\theta) \tag{4.2}$$

where G_T is an $m \times m$ positive definite weight matrix assumed to converge in probability to a positive definite limit G. If the system is exactly identified, $m = d$, the choice of G_T is irrelevant and minimizing (4.2) amounts to setting $m_T(\theta)$ to 0.

To insure consistency of $\hat{\theta}$, h is assumed to satisfy

$$E\left[h(Y_\Delta \Delta, \theta_0)\right] = 0. \tag{4.3}$$

It follows from standard arguments, subject to regularity conditions (see Hansen (1982)) that $\sqrt{T}(\hat{\theta} - \theta_0)$ converges in law to $N(0, \Omega)$, with

$$\Omega^{-1} = \Delta^{-1} D' G D (D' G S G D)^{-1} D' G D \tag{4.4}$$

where

$$D \equiv E\left[\dot{h}(Y_\Delta, \Delta, \theta_0)\right] \tag{4.5}$$

is $m \times d$, and

$$S \equiv E\left[h(Y_\Delta, \Delta, \theta_0) h(Y_\Delta, \Delta, \theta_0)'\right] \tag{4.6}$$

is $m \times m$. The weight matrix G_T can be chosen optimally to minimize the asymptotic variance Ω, by taking it to be any consistent estimator of S^{-1}. A consistent first-step estimator of θ, needed to compute the optimal weight matrix, can be obtained by minimizing (4.2) with $G_T = Id$. When G_T is then chosen optimally, $G = S^{-1}$ and as a result equation (4.4) reduces to

$$\Omega^{-1} = \Delta^{-1} D' S^{-1} D. \tag{4.7}$$

In particular, (4.7) applies when the system is exactly identified ($r = d$) since the choice of the weight matrix is irrelevant in that case.

4.2. *Fisher's Information in the Presence of Jumps*

Of course, by choosing h to be the score vector, this class of estimator encompasses maximum likelihood. If we let

$$l(y, \delta, \theta) \equiv \ln p(x_0 + y | x_0, \delta; \theta)$$

denote the log-likelihood function, this corresponds to $h(y, \delta, \theta) = -\dot{l}(y, \delta, \theta)$. Then $S = E[\dot{l}\dot{l}']$, $D = -E[\ddot{l}]$ and

$$S = D \tag{4.8}$$

is Fisher's Information matrix. The asymptotic variance of $\hat{\theta}_{\text{MLE}}$ takes the form

$$\text{AVAR}_{\text{MLE}}(\theta) = \Delta(DS^{-1}D)^{-1} = \Delta D^{-1}. \tag{4.9}$$

The following theorem shows that, despite the difficulties described earlier, it is still possible, using maximum likelihood, to identify σ^2 with the same degree of precision as if there were no jumps:

Theorem 1. *When the Brownian motion is contaminated by Poisson jumps, it remains the case that*

$$\text{AVAR}_{\text{MLE}}\left(\sigma^2\right) = 2\sigma^4\Delta + o(\Delta) \tag{4.10}$$

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

Theorem 1 says that maximum-likelihood can theoretically perfectly disentangle σ^2 from the presence of the jumps, when using high frequency data. I will show in Section 7 below Monte Carlo evidence that suggests that this holds true in practice, too.

Note also that the result of Theorem 1 states that the presence of the jumps imposes no cost on our ability to estimate σ^2 : the variance which is squared in the leading term is only the diffusive variance 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 independent 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 $2\left(\sigma^2 + s^2\right)^2\Delta$.

What is happening here is that, as Δ gets smaller, our ability to identify price discontinuities improves (recall Figure 5).This is because these Poisson discontinuities are, by construction, discrete, and there are few of them relative to the diffusive moves. Then if we can see them, we can exclude them, and do as if they did not happen in the first place. More challenging therefore will be the case where the jumps occur infinitely often and are infinitely frequent (see Section 6 below). But before examining that question, I will investigate the ability of a large class of GMM estimators to approach the efficiency of MLE. Indeed, in light of the Cramer Rao lower bound, Theorem 1 establishes $2\sigma^4\Delta$ as the benchmark for alternative methods that attempt to estimate σ^2 (based on the quadratic variation, absolute variation, absolute power variation, etc.) and it is interesting to know how closely GMM using such moment functions can approximate MLE.

5. Using Moments: How Close Does GMM Come to MLE?

The first question I now address is whether the identification of σ^2 achieved by the likelihood, despite the presence of jumps, can be reproduced by conditional moments of the process of integer or non-integer type, and which moments or combinations of moments come closest to achieving maximum likelihood efficiency .While it is clear that MLE is the preferred method, and as discussed above has been used extensively in that context, it is nevertheless instructive to determine which specific choices of moment functions do best in terms of approximating its efficiency. So, in GMM estimation, I form moment functions of the type $h(y, \delta, \theta) = y^r - M(\delta, \theta, r)$ and/or $h(y, \delta, \theta) = |y|^r - M_a(\delta, \theta, r)$ 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.

I will refer to different GMM estimators of θ by listing the moments $M(\Delta, \theta, r)$ and/or $M_a(\Delta, \theta, r)$ that are used for that particular estimator. For example, the estimator of σ^2 obtained by using the single moment $M(\Delta, \theta, 2)$ corresponds to the discrete approximation to the quadratic variation of the process. Estimators based on the single moment $M_a(\delta, \theta, r)$ correspond to the power variation, etc. By using Taylor expansions in Δ, I characterize in closed form the properties of these different GMM estimators.

5.1. *Estimating σ^2 Alone*

I start with the case where only σ^2 is to be estimated. The jump term, while present, has known parameters, or one could think of it as being a nuisance process that is of no interest. For simplicity, I will assume here and in the rest of the paper that the drift and mean jump are centered at zero ($\mu = \beta = 0$). This assumption is largely inconsequential, except that it greatly simplifies the expressions below. It also makes the standard moments of odd order zero.

The technique I use to obtain tractable closed form expressions for the asymptotic variances of the different estimators under consideration is to Taylor-expand them in Δ around $\Delta = 0$ (see Aït-Sahalia and Mykland (2003a) for another use of this technique in a different context). Indeed, computation of AVAR_{GMM} requires the separate computation of the matrices D and S in (4.5)-(4.6). These matrices are expected values of functionals of the moment vector h, taken with respect to the law of the observed process Y_Δ. In the present example, this law has density (2.4). With polynomial moment functions in h (including possibly absolute values and non integer powers), the functionals $\dot{h}$ and hh' retain the polynomial form. Thus D and S can be computed explicitly using the moments calculated in Proposition 1. If non-polynomial moment functions were to be used, the corresponding calculations would involve Taylor-expanding using the infinitesimal generator of the process, as described in (2.7). The results for D and S can then be combined to form the matrix Ω in (4.7), which in turn has a natural expansion in powers of Δ.(again, possibly non-integer when non-integer moments are used).

With this method, it becomes possible to compare different estimators by looking at the Taylor expansions of their respective asymptotic variances. I find that, although it does not restore full maximum likelihood efficiency, using absolute moments in GMM helps greatly. In particular, the next proposition shows that when σ^2 is estimated using exclusively moments of the form $M(\Delta, \theta, r)$, then $\text{AVAR}_{\text{GMM}}\left(\sigma^2\right) = O(1)$, a full order of magnitude bigger than achieved by MLE. When moments of the form $M_a(\Delta, \theta, r)$ with $r \in (0, 1)$ are used, however, $\text{AVAR}_{\text{GMM}}\left(\sigma^2\right) = O(\Delta)$, i.e., the same order as achieved by MLE, although the constant of proportionality is always greater than $2\Delta\sigma^4$ as should be the case in light of the Cramer-Rao lower bound.

When σ^2 is estimated based on the moment $M_a(\Delta, \theta, r)$ with $r \in (1, 2]$ are used, $\text{AVAR}_{\text{GMM}}(\sigma^2) = O(\Delta^{2-r})$. Specifically:

Proposition 2. *The following table gives the asymptotic variance of the GMM estimator of σ^2 using different combinations of moment functions:*

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

When the moments $(M_a(\Delta, \theta, r)\,,\ M_a(\Delta, \theta, q))'$ are used jointly, $\text{AVAR}_{\text{GMM}}(\sigma^2)$ is given in the no jumps case by

$$\text{AVAR}_{\text{GMM}}(\sigma^2) = 2\Delta\sigma^4\frac{2\pi^{1/2}A(r, q)}{B(r, q)} + o(\Delta)$$

where

$$\begin{aligned}
A(r, q) &= \Gamma\left(\frac{1}{2} + q\right)\left(\pi^{1/2}\Gamma\left(\frac{1}{2} + r\right) - \Gamma\left(\frac{1+r}{2}\right)^2\right) + 2\Gamma\left(\frac{1+q}{2}\right)\Gamma\left(\frac{1+r}{2}\right)\Gamma\left(\frac{1+q+r}{2}\right) \\
&\quad -\pi^{1/2}\Gamma\left(\frac{1+q+r}{2}\right)^2 - \Gamma\left(\frac{1+q}{2}\right)^2\Gamma\left(\frac{1}{2} + r\right) \\
B(r, q) &= \pi^{1/2}r^2\Gamma\left(\frac{1}{2} + q\right)\Gamma\left(\frac{1+r}{2}\right)^2 + \Gamma\left(\frac{1+q}{2}\right)^2\left(\pi^{1/2}q^2\Gamma\left(\frac{1}{2} + r\right) - (q - r)^2\Gamma\left(\frac{1+r}{2}\right)^2\right) \\
&\quad -2\pi^{1/2}qr\Gamma\left(\frac{1+q}{2}\right)\Gamma\left(\frac{1+r}{2}\right)\Gamma\left(\frac{1+q+r}{2}\right)
\end{aligned}$$

Figure 7 plots the efficiency of the GMM estimator of σ^2 using $.M_a(\Delta, \theta, r)$, relative to MLE, as a function of r. In light of the table in Proposition 2, this is given by the function

$$r \mapsto \frac{2}{r^2}\left(\frac{\pi^{1/2}\Gamma\left(\frac{1}{2}+r\right)}{\Gamma\left(\frac{1+r}{2}\right)^2} - 1\right).$$

In the absence of jumps, the minimum is achieved by selecting $r = 2$, which reproduces the MLE's asymptotic variance of $2\Delta\sigma^4$. This is not surprising since the MLE for σ^2 in the absence of jumps is simply the quadratic variation of the process (at frequency Δ^{-1}). When jumps are present, however, absolute moments taken individually (even though they do better than regular moments) are no longer capable of attaining the efficiency of MLE. Figure 8 plots the corresponding picture, at the weekly frequency and the same parameters as above.

However, taking such absolute moments of different orders in combination improves upon any single one. Figure 9 plots the relative efficiency, as a function of (r, q) that results from estimating σ^2 using the overidentified GMM system based on the vector of moment conditions $(M_a(\Delta, \theta, r), \ M_a(\Delta, \theta, q))'$. Given Proposition 2, this is the surface

$$(r, q) \mapsto \frac{2\pi^{1/2} A(r, q)}{B(r, q)}$$

In the no jumps case, using two functions improves upon one when $r \neq 2$, and achieves MLE efficiency provided one of the two is the quadratic variation. When jumps are present, however, it is only asymptotically, as the number of these absolute moment functions increases, that GMM can reproduce MLE.

Finally, note that at the leading order in Δ, GMM makes little use of the quadratic variation $M(\Delta, \theta, 2)$ when an absolute moment of the type $M_a(\Delta, \theta, r)$ is also part of the h vector. Comparing the asymptotic variance in the case where $(M(\Delta, \theta, 2), \ M_a(\Delta, \theta, 1)'$ are both used together to that where only $M_a(\Delta, \theta, 1)$ is used, we see that the decrease in variance is only

$$2\Delta\sigma^2\left((\pi-2)\sigma^2 + \pi\eta\lambda\right) - 2\Delta\sigma^2\left((\pi-2)\sigma^2 + \frac{(3\pi-8)}{3}\eta\lambda\right) = \frac{16}{3}\Delta\sigma^2\eta\lambda.$$

5.2. *Estimating σ^2 and λ Together*

I now turn to the case where both σ^2 and λ are to be jointly estimated.

Proposition 3. *When (σ^2, λ) are estimated using the moments $(M(\Delta, \theta, 2), M(\Delta, \theta, 4))'$, the resulting* $\mathrm{AVAR}_{\mathrm{GMM}}\left(\sigma^2, \lambda\right)$ *is*

$$\begin{pmatrix} \frac{14\eta^2\lambda}{3} + \frac{2\Delta\left(70\eta^2\lambda^2+22\eta\lambda\sigma^2+3\sigma^4\right)}{3} + O(\Delta^2) & \frac{-20\eta\lambda}{3} - \frac{2\Delta\lambda\left(79\eta\lambda+28\sigma^2\right)}{3} + O(\Delta^2) \\ \bullet & \frac{35\lambda}{3} + \frac{2\Delta\lambda\left(91\eta\lambda+40\sigma^2\right)}{3\eta} + O(\Delta^2) \end{pmatrix} \tag{5.1}$$

When the moments $(M(\Delta,\theta,2), M_a(\Delta,\theta,1/2))'$ *are used, the resulting* $\mathrm{AVAR}_{\mathrm{GMM}}(\sigma^2,\lambda)$ *is*

$$\begin{pmatrix} 2\Delta\sigma^2\left(\sigma^2+\eta\lambda\right)(\pi-2)+O(\Delta^{3/2}) & -2\Delta^{1/2}\eta^{1/2}\lambda\sigma-\frac{2\Delta\sigma^2\left((\pi-2)\eta\lambda+(\pi-3)\sigma^2\right)}{\eta}+O(\Delta^{3/2}) \\ \bullet & 3\lambda+\frac{4\Delta^{1/2}\lambda\sigma}{\eta^{1/2}}+\Delta\left(2\lambda^2+\frac{2\pi\lambda\sigma^2}{\eta}+\frac{2(\pi-3)\sigma^4}{\eta^2}\right)+O(\Delta^2) \end{pmatrix} \tag{5.2}$$

As in the σ^2 alone situation, the introduction of an absolute moment of the type $M_a(\Delta,\theta,r)$ reduces the asymptotic variance of the GMM estimator of σ^2 by an order of magnitude, from $O(1)$ in (5.1) to $O(\Delta)$ in (5.2), which is the same rate as MLE but with a higher constant.

5.3. *Estimating* σ^2, λ *and* η *Together*

The following result gives the asymptotic variance of the GMM estimator of (σ^2,λ,η), estimated jointly.

Proposition 4. *When* (σ^2,λ,η) *are estimated using the moments* $(M(\Delta,\theta,2), M(\Delta,\theta,4), M_a(\Delta,\theta,1/2))'$, *the resulting matrix* $\mathrm{AVAR}_{\mathrm{GMM}}(\sigma^2,\lambda,\eta)$ *has the following elements*

$$\begin{aligned}
\left(\sigma^2,\sigma^2\right) &: \frac{2\Delta\sigma^2\left(3(\pi-2)\sigma^2+(3\pi-7)\eta\lambda\right)}{3}+O(\Delta^{3/2}) \\
\left(\sigma^2,\lambda\right) &: -2\Delta^{1/2}\eta^{1/2}\lambda\sigma-\frac{2\Delta\sigma^2\left((6\pi-13)\eta\lambda+6(\pi-3)\sigma^2\right)}{3\eta}+O(\Delta^{3/2}) \\
\left(\sigma^2,\eta\right) &: 2\Delta^{1/2}\eta^{3/2}\sigma+\frac{2\Delta\sigma^2\left((\pi-2)\eta\lambda+(\pi-3)\sigma^2\right)}{\lambda}+O(\Delta^{3/2}) \\
(\lambda,\lambda) &: \frac{11\lambda}{3}+\frac{8\Delta^{1/2}\lambda\sigma}{\eta^{1/2}}+\Delta\frac{\left(110\lambda^2-\frac{16\lambda\sigma^2}{\eta}+\frac{24\pi\lambda\sigma^2}{\eta}-\frac{24(3-\pi)\sigma^4}{\eta^2}\right)}{3}+O(\Delta^{3/2}) \\
(\lambda,\eta) &: -\frac{8\eta}{3}-6\Delta^{1/2}\eta^{1/2}\sigma+\Delta\left(\frac{4(3-\pi)\sigma^4}{\eta\lambda}-\frac{122\eta\lambda}{3}-\frac{10\sigma^2}{3}-4\pi\sigma^2\right)+O(\Delta^{3/2}) \\
(\eta,\eta) &: \frac{14\eta^2}{3\lambda}+\frac{4\Delta^{1/2}\eta^{3/2}\sigma}{\lambda}+\frac{2\Delta\left(70\eta^2\lambda^2+17\eta\lambda\sigma^2+3\pi\eta\lambda\sigma^2+3(\pi-3)\sigma^4\right)}{3\lambda^2}+O(\Delta^{3/2})
\end{aligned}$$

Comparing the asymptotic variance of σ^2 in the case where only λ is estimated along with σ^2 (the upper left element of (5.2)) to that obtained in Proposition 4 measures the cost associated with not knowing η, given the moment functions used. That cost is given here by:

$$\frac{2\Delta\sigma^2\left(3(\pi-2)\sigma^2+(3\pi-7)\eta\lambda\right)}{3}-2\Delta\sigma^2\left(\sigma^2+\eta\lambda\right)(\pi-2)=\frac{4}{3}\Delta\sigma^2\eta\lambda.$$

6. Disentangling the Diffusion from Other Jump Processes: The Cauchy Case

Theorem 1 demonstrated the ability of maximum-likelihood to fully distinguish the diffusive component from the jump component on the basis of the full sample path. I then showed under what circumstances (i.e., choices of moment functions) GMM was able to approach this result, although not fully reproduce the efficiency of MLE. I now examine whether the perfect distinction afforded by MLE 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. Among the class of continuous-time Markov processes, it is natural to look at Lévy processes. As I will discuss below, Poisson jumps are a unique case in the Lévy universe. Yet, it is possible to find examples of other pure jump processes for which the same result continues to hold, which cannot be explained away as easily as in the Poisson case.

6.1. Lévy Processes

I start by briefly reviewing the main properties of Lévy processes that I will use in the rest of the paper (see e.g., Bertoin (1998), for further details). A process is a Lévy process if it has stationary and independent increments and is continuous in probability. A Lévy process can be decomposed as the sum of three independent components: a linear drift, a Brownian motion and a pure jump process. Correspondingly, the log-characteristic function of a sum of independent random variables being the sum of their individual characteristic functions, the characteristic function of a Lévy process is given by the Lévy-Khintchine formula, which states that there exist constants $\gamma_c \in \mathbb{R}$, $\sigma \geq 0$ and a positive sigma-finite measure $\nu(\cdot)$ on $\mathbb{R}\backslash\{0\}$ (extended to $\mathbb{R}$ by setting $v(\{0\}) = 0$) satisfying

$$\int_{-\infty}^{+\infty} \min\left(1, z^2\right) \nu(dz) < \infty \tag{6.1}$$

such that the log-characteristic function $\psi(u)$, defined by

$$E[e^{iuX_\Delta}|X_0 = 0] = e^{\psi(u)\Delta}$$

has the form for $u \in \mathbb{R}$:

$$\psi(u) = i\gamma_c u - \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{+\infty} \left(e^{iuz} - 1 - iuzc(z)\right) \nu(dz). \tag{6.2}$$

The three quantities $(\gamma_c, \sigma, \nu(\cdot))$, called the characteristics of the Lévy process, completely describe the probabilistic behavior of the process. γ_c is the drift rate of the process, σ its volatility from the Brownian component and the measure $\nu(\cdot)$ describes the pure jump component. It is known as the Lévy measure and has the interpretation that $\nu(E)$ 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. Sample paths of the process are continuous if and only if $\nu \equiv 0$. Note that $\nu(\cdot)$ is not necessarily a probability measure, in that $\nu(\mathbb{R})$ may be finite or infinite.

The function $c(z)$ is a weighting function whose role is to make the integrand in (6.2) integrable. When $|z|\nu(dz)$ is integrable near 0, it is enough to have

$$e^{iuz} - 1 - iuzc(z) = O(|z|) \text{ as } z \to 0$$

and this can be achieved simply by setting $c = 0$ near $z = 0$. But if $|z|v(dz)$ is not integrable near 0 (and only

$z^2 v(dz)$ is, as required by (6.1)), then one needs a non-zero $c(z)$ function, which must insure in light of (6.1) that

$$e^{iuz} - 1 - iuzc(z) = O(z^2) \text{ as } z \to 0, \tag{6.3}$$

that is, $c(z) \sim 1$ near $z = 0$. When $|z|\nu(dz)$ is integrable near ∞, it is enough to have

$$e^{iuz} - 1 - iuzc(z) = O(|z|) \text{ as } z \to \pm\infty$$

and this can be achieved simply by setting $c = O(1)$ near $z = \infty$. But if $|z|v(dz)$ is not integrable near ∞ (and only $v(dz)$ is, as required by (6.1)), then one needs

$$e^{iuz} - 1 - iuzc(z) = O(1) \text{ as } z \to \pm\infty, \tag{6.4}$$

that is, $c(z) = O\left(1/|z|\right)$ near $z = \pm\infty$. Typical examples include $c(z) = 1/(1+z^2)$, $c(z) = 1(|z| < \varepsilon)$ for some $\varepsilon > 0$, etc.

The function $c(z)$ can be replaced by another one. Any change in the weighting function from $c(z)$ to $c'(z)$ is absorbed by a matching change in γ_c, which is replaced by γ_c' in such a way that

$$\gamma_c - \int_{-\infty}^{+\infty} zc(z)\nu(dz) = \gamma_c' - \int_{-\infty}^{+\infty} zc'(z)\nu(dz).$$

The infinitesimal generator of the process is given by

$$\begin{aligned} A \cdot f(\Delta, x, x_0) \quad = \quad & \frac{\partial f(\Delta, x, x_0)}{\partial \Delta} + \left(\gamma_c - \int_{-\infty}^{+\infty} zc(z)\nu(dz)\right) \frac{\partial f(\Delta, x, x_0)}{\partial x} \\ & + \frac{1}{2}\sigma^2 \frac{\partial^2 f(\Delta, x, x_0)}{\partial x^2} + \int_{-\infty}^{+\infty} \{f(\Delta, x+z, x_0) - f(\Delta, x, x_0)\} \, \nu(dx). \end{aligned}$$

Examples of Lévy processes include the Brownian motion ($c = 0$, $\gamma_c = 0$, $\sigma = 1$, $\nu = 0$), the Poisson process ($c = 0$, $\gamma_c = 0$, $\sigma = 0$, $\nu(dx) = \lambda\delta_1(dx)$ where δ_1 is a Dirac point mass at $x = 1$) and the Poisson jump diffusion I considered above in (2.1), corresponding to $c = 0$, $\gamma_c = \mu$, $\sigma > 0$, $\nu(dx) = \lambda n(x; \beta, \eta)dx$ where $n(x; \beta, \eta)$ is the Normal density with mean β and variance η.

The question I now address is whether it is 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 $\nu(\mathbb{R})$, i.e., a finite number of jumps in a finite time interval (and the sample paths are piecewise constant). In that case, define $\lambda = \nu(\mathbb{R})$ and the distribution of the jumps has measure $n(dx) = v(dx)/\lambda$. 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.[2] Intuitively, these tiny jumps ought to be harder to distinguish from Brownian noise, which is itself made up of many small moves. Can the

[2] The number of "big" jumps remains finite: $\nu((-\infty, -\varepsilon) \cup (\varepsilon, +\infty)) < \infty$.

likelihood still tell them perfectly apart from Brownian noise?

I will consider as an example the Cauchy process, which is the pure jump process ($\sigma = 0$) with Lévy measure

$$\nu(dx) = \frac{\alpha}{x^2}dx \tag{6.5}$$

and, with weight function $c(z) = 1/(1+z^2)$, $\gamma_c = 0$. This is an example of a symmetric stable distribution of index $0 < \xi < 2$ and rate $\alpha > 0$, with log characteristic function proportional to $\psi(u) = -(\alpha\,|u|)^{\xi}$, and Lévy measure

$$\nu(dx) = \frac{\alpha^{\xi}\xi}{|x|^{1+\xi}}dx. \tag{6.6}$$

The Cauchy process corresponds to $\xi = 1$, while the limit $\xi \to 2$ (from below) produces a Gaussian distribution.

While, as a result of (6.1), all Lévy processes have finite quadratic variation almost surely, the absolute variation of the process will be finite only if $\sigma = 0$ and if $|z|\,\nu(dz)$ is integrable near 0, a condition that fails for the Cauchy process but is satisfied by the Poisson process (and gamma, beta, and simple homogeneous examples). More generally, for $r > 0$,

$$\Pr\left(\sum_{0\leq s\leq t} |X_s - X_{s-}|^r < \infty\right) = 1 \Longleftrightarrow \int_{-\infty}^{+\infty} \min\left(1, |z|^r\right)\nu(dz) < \infty \tag{6.7}$$

which in the case (6.6) is equivalent to $r > \xi$.

6.2. *Mixing Cauchy Jumps with Brownian Noise*

So I now look at the situation where

$$dX_t = \mu dt + \sigma dW_t + dC_t \tag{6.8}$$

where C_t is a Cauchy process independent of the Brownian motion W_t. Focusing on the ability to disentangle σ^2 from the jumps, let's consider again the case where $\mu = 0$. The solution of the stochastic differential equation (6.8) is

$$\begin{aligned} X_\Delta - X_0 &= \sigma W_\Delta + C_\Delta \\ &= \sigma\sqrt{\Delta}Z_\Delta + C_\Delta \end{aligned} \tag{6.9}$$

where $Z_\Delta \sim N(0,1)$. Equation (6.9) implies again that the log-returns $Y_{n\Delta} = X_{n\Delta} - X_{(n-1)\Delta}$ are i.i.d.

By independence of C and W, the transition density of the process X is given by the convolution of their respective densities

$$f_{X_\Delta}(y) = f_{\sigma\sqrt{\Delta}Z_\Delta + C_\Delta}(y) = \int_{-\infty}^{+\infty} f_{\sigma\sqrt{\Delta}Z_\Delta}(y-z) f_{C_\Delta}(z)dz.$$

To obtain the density $f_{C_\Delta}(z)$, recall that the log-characteristic function of C is given by (6.2),

$$\psi_C(u) = \int_{-\infty}^{+\infty} \left(e^{iuz} - 1 - \frac{izu}{1+z^2} \right) \frac{\alpha}{z^2} dz = -\pi\alpha |u|, \tag{6.10}$$

with the density following by Fourier inversion since the characteristic function $\exp(\psi_C(u))$ is integrable:

$$\begin{aligned} f_{C_\Delta}(z) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left(-iuz + \psi_C(u)\Delta\right) du \\ &= \frac{\Delta\alpha}{\Delta^2\alpha^2\pi^2 + z^2}. \end{aligned} \tag{6.11}$$

Finally, for this process the density $q(y|\Delta;\theta) = p(x+y|x_0,\Delta;\theta)$ is given by the convolution

$$f_{X_\Delta}(y) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sqrt{\Delta\sigma^2}} \exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2} \right) \frac{\Delta\alpha}{\Delta^2\alpha^2\pi^2 + z^2} dz. \tag{6.12}$$

which is known as the Voigt function.

The question now becomes whether it is still possible, using maximum likelihood, to identify σ^2 with the same degree of precision as if there were no jumps, despite the fact that the Cauchy process contaminates the Brownian motion with infinitely many infinitesimal jumps. The answer is, surprisingly, yes:

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

$$\mathrm{AVAR}_{MLE}\left(\sigma^2\right) = 2\sigma^4\Delta + o(\Delta). \tag{6.13}$$

6.3. *Intuition for the Result: How Big is That Infinite Number of Small Jumps?*

Theorem 2 has shown 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 surprising result is the following: while there is an infinite number of small jumps in a Cauchy process, this "infinity" remains relatively small (just like the cardinal of the set of integers is smaller than the cardinal 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. In other words, they are harder to pick up from inspection of the sample path than Poisson jumps are, but with a fine enough microscope, still possible. And the likelihood is the best microscope there is, per Cramer-Rao.

I now show formally how this works:

Lemma 1. *Fix $\varepsilon > 0$. If Y_Δ is the log-return from a pure Brownian motion, then*

$$\Pr\left(|Y_\Delta| > \varepsilon\right) = \frac{\Delta^{1/2}\sigma}{\varepsilon}\sqrt{\frac{2}{\pi}} \exp\left(-\frac{\varepsilon^2}{2\Delta\sigma^2} \right)(1+o(1)) \tag{6.14}$$

is exponentially small as $\Delta \to 0$. However, if Y_Δ results from a Lévy pure jump process with jump measure $v(dz)$, then under regularity conditions

$$\Pr\left(|Y_\Delta| > \varepsilon\right) = \Delta \times \int_{|y|>\varepsilon} v(dy) + o(\Delta) \tag{6.15}$$

which decreases only linearly in Δ.

For example, for a Cauchy process a direct calculation based on the density (6.11) yields

$$\Pr\left(|Y_\Delta| > \varepsilon\right) = \int_{|y|>\varepsilon} f_{C_\Delta}(y)dy = \Delta\frac{2\alpha}{\varepsilon} - \Delta^3\frac{2\alpha^3\pi^2}{3\varepsilon^3} + O(\Delta^5) \tag{6.16}$$

whose leading term coincides with (6.15) when v is replaced by its Cauchy expression (6.5). More generally, with a symmetric stable process with order ξ, whose Lévy measure is given in (6.6), we have

$$\Pr\left(|Y_\Delta| > \varepsilon\right) = \Delta \times \frac{2\alpha^\xi}{\varepsilon^\xi} + o(\Delta). \tag{6.17}$$

The key aspect here is that the order in Δ of $\Pr\left(|Y_\Delta| > \varepsilon\right)$ for a pure jump Lévy process is always $O(\Delta)$. In other words, Lévy jump processes will always produce moves of size greater than ε at a rate far greater than the Brownian motion. Brownian motion will have all but an exponentially small fraction of its increments of size less than any given ε. Lévy pure jump processes with infinite $\nu(\mathbb{R})$ (i.e., all except the compound Poisson process), despite producing an infinite amount of small jumps will not produce quite as many small moves as Brownian motion does: "only" a fraction $1 - O(\Delta)$ of their increments are smaller than ε. It's a question of two "infinities" one growing linearly, the other exponentially.

In that sense, all of these Lévy pure jump processes produce tiny jumps (those of size less than ε) at the same rate $1 - O(\Delta)$ as a compound Poisson process does:

$$\Pr\left(|Y_\Delta| > \varepsilon\right) = \Delta \times \lambda \int_{|y|>\varepsilon} n(y;\beta,\eta)dy + o(\Delta) \tag{6.18}$$

since in the example considered above the jumps J have density $n(x;\beta,\eta)$. The probability of seeing a move greater than ε is at the first order in Δ the probability that one jump occurs, i.e.,

$$\Pr\left(N_\Delta = 1\right) = \Delta\lambda + O(\Delta^2)$$

times the probability that J will be of size at least ε.

Do jumps always have to behave that way? The answer is yes, in light of the following. Ray (1956) showed that the sample paths of a Markov process are almost surely continuous if and only if, for every $\varepsilon > 0$,

$$\Pr\left(|Y_\Delta| > \varepsilon\right) = o\left(\Delta\right) \tag{6.19}$$

Ray's condition maps out the continuity of the sample path into a bound on the size of the probability of leaving a given neighborhood in the amount of time Δ; intuitively, this probability must be small as Δ goes to zero if the sample paths are to remain continuous. Condition (6.19) says how small this probability must be as Δ gets smaller. But, since condition (6.19) is necessary and sufficient, it also establishes a lower bound for how big the probability of making a move greater than ε must be if the process is *not* continuous, i.e., can jump. Based on what we have seen, it is therefore natural to have for a Lévy pure jump process $\Pr(|Y_\Delta| > \varepsilon) = O(\Delta)$ as stated in (6.15), and not $o(\Delta)$. Further, while I wrote (6.19) assuming that the process has independent increments (i.e., be Lévy), this condition is valid also for processes with dependent increments: replace $\Pr(|Y_\Delta| > \varepsilon)$ with $\Pr(|X_\Delta - X_0| > \varepsilon | X_0 = x_0)$ and add the requirement that it be satisfied uniformly for x_0 in a compact.

Since

$$\Pr(|Y_\Delta| \leq \varepsilon) = 1 - \Pr(|Y_\Delta| > \varepsilon)$$

is the probability of making small moves (the ones that look like Brownian motion), this effectively puts an upper bound on the ability of a jump process to imitate the behavior of Brownian volatility. So the result is likely not driven by the fact that the divergence of $\nu(dx)$ near 0 is only $O(|x|^{-2})$ for the Cauchy process, instead of for instance $O(|x|^{-(1+\xi)})$ with ξ greater than 1 but smaller than 2 ($\xi \to 2$ provides the maximum admissible amount of small jumps per unit of time, while still satisfying the requirement (6.1)).

7. Monte Carlo Simulations

A legitimate question at this point is whether Theorems 1 and 2, which are statements about the behavior of the estimators at high frequency, have relevance at the observation frequencies that we typically encounter in asset pricing. So, in this section, I report the results of Monte Carlo simulations designed to examine the empirical adequacy of the theoretical results when we observe asset prices once a day. The daily frequency is generally considered to be low enough to be largely unaffected by the market microstructure issues that can substantially derail the performance of high frequency quantities such as the realized quadratic variation, etc. (see Aït-Sahalia and Mykland (2003b) for an analysis of the effect of market microstructure noise on high frequency estimates).

Starting with the jump-diffusion (2.1), I simulate $5,000$ sample paths, each of length $n = 1,000$ at the daily frequency ($\Delta = 1/252$). To demonstrate the ability of the likelihood to disentangle the volatility parameter from the jumps, I purposefully set the arrival rate of the jumps at a high level, $\lambda = 5$ in the Poisson case. Five jumps per year on average is much higher than would be realistic based on actual estimates for stock index returns (I use $\lambda = 0.2$ in all figures above). I set the value of σ at a realistic level, $\sigma = 0.3$. So there is relatively little volatility given the amount of jumps, which should make it more difficult to distinguish volatility from jumps among the overall amount of noise. The standard deviation of the jump size is $\eta^{1/2} = 0.6$. The process is symmetric ($\mu = \beta = 0$). In the Cauchy case, I set $\alpha = 0.2$ and $\sigma = 0.3$. Again, jumps are plentiful.

I then estimate the parameter σ^2 using MLE. I also repeated the experiment estimating both (σ^2, λ) in the Poisson case, (σ^2, α) in the Cauchy case to investigate the effects of joint estimation on our ability to distinguish the volatility parameter from the jump component. Figures 10 and 11 report the small sample and asymptotic distributions for (σ^2, λ) and (σ^2, α) respectively. The histograms show that despite the large number of jumps the estimates of σ^2 remain in a tight interval around the true value of 0.09. And the sample variance is quite close to the values predicted by Theorems 1 and 2 (corresponding to the asymptotic distribution) despite the fact that the data are only sampled once a day. These results show that the theoretical asymptotic distribution of MLE as derived above provides a good approximation to the small sample behavior of the estimators at the daily frequency. Finally, Figures 12 and 13 show the resulting confidence regions for the two joint estimation problems in the Poisson and Cauchy cases. Note that in the Poisson case, the estimators are largely uncorrelated hence the roughly circular shape of the joint confidence interval. This is not the case in the Cauchy situation however.

8. Conclusions

I studied the effect of the presence of jumps on our ability to identify the volatility component of the log returns process and found that, somewhat surprisingly, jumps had no detrimental effect as far as maximum likelihood estimation was concerned. I also discussed which moment conditions are better than others in terms of approaching the efficiency of maximum likelihood for this problem.

Even more surprisingly, the result did not depend on the jumps being large and infrequent, i.e., Poisson jumps. It remains valid in the case of Cauchy jumps which can be infinitely small in magnitude, and infinitely frequent. Finally, I provided an explanation of this phenomenon based on the fact that all jump processes, despite having an infinite number of small jumps, have distinguishing characteristics relative to Brownian motion that ultimately can be picked up by the likelihood. As discussed in the introduction, this result means that the recent literature in finance that introduces Lévy processes in option pricing, portfolio choice or risk management will not face major obstacles from the econometric side, despite what the intuition might have initially suggested.

Appendix A: Proof of Proposition 1

Writing the expected value in terms of the density of the discrete increments $Y_\Delta = X_\Delta - X_0$, we have, with $b \equiv \Delta(\mu + \beta\lambda)$

$$\begin{aligned}\Delta_a(\Delta,\theta,r) &= E\left[\left|Y_\Delta - \Delta(\mu+\beta\lambda)\right|^r\right] \\ &= \int_{-\infty}^{+\infty} \left|y - \Delta(\mu+\beta\lambda)\right|^r q(y,\Delta;\theta)\,dy \\ &= \int_{-\infty}^{b} (b-y)^r q(y,\Delta;\theta)\,dy + \int_{b}^{+\infty} (y-b)^r q(y,\Delta;\theta)\,dy\end{aligned}$$

Then recall from (2.4) that

$$q(y,\Delta;\theta) = \sum_{n=0}^{+\infty} \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{\sqrt{2\pi}\sqrt{n\eta+\Delta\sigma^2}n!} \exp\left(-\frac{(y-\mu\Delta-n\beta)^2}{2(n\eta+\Delta\sigma^2)}\right)$$

and compute each of the two integrals term by term.

Each term can be computed from the form

$$\begin{aligned}&\int_{-\infty}^{b} (b-y)^r \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(y-a)^2}{2v}\right) dy + \int_{b}^{+\infty} (y-b)^r \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(y-a)^2}{2v}\right) dy \\ = \; & (2v)^{r/2}\pi^{-1/2} \exp\left(-\frac{(b-a)^2}{2v}\right) \Gamma\left(\frac{1+r}{2}\right) F\left(\frac{1+r}{2}, \frac{1}{2}, \frac{(b-a)^2}{2v}\right)\end{aligned}$$

based on Section 13.2 in Abramowitz and Stegun (1972). Summing the terms over n after replacing the values a and v by their expressions $a = \mu\Delta + n\beta$ and $v = \Delta\sigma^2 + n\eta$ yields the result.

Appendix B: Proof of Theorem 1

Fisher's Information for σ^2 is

$$\begin{aligned}I_{\sigma^2} &= E\left[\left(\frac{\partial \ln q(y|\Delta;\theta)}{\partial\sigma^2}\right)^2\right] \\ &= E\left[\left(\frac{\partial q(y|\Delta;\theta)}{\partial\sigma^2}\frac{1}{q(y|\Delta;\theta)}\right)^2\right] \\ &= \int_{-\infty}^{+\infty} \left(\frac{\partial q(y|\Delta;\theta)}{\partial\sigma^2}\right)^2 \frac{dy}{q(y|\Delta;\theta)} \qquad \text{(B.1)}\end{aligned}$$

where q is given by (2.4):

$$\begin{aligned}q(y|,\Delta;\theta) &= \sum_{n=0}^{+\infty} \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{\sqrt{2\pi}\sqrt{n\eta+\Delta\sigma^2}n!} \exp\left(-\frac{y^2}{2(n\eta+\Delta\sigma^2)}\right) \\ &\equiv \sum_{n=0}^{+\infty} q_n(y|,\Delta;\theta)\frac{\exp(-\lambda\Delta)(\lambda\Delta)^n}{n!} \qquad \text{(B.2)}\end{aligned}$$

so that

$$\frac{\partial q\left(y|\Delta;\theta\right)}{\partial\sigma^{2}}=\sum_{n=0}^{+\infty}\frac{\exp(-\lambda\Delta)(\lambda\Delta)^{n}\Delta}{2\sqrt{2\pi}\left(n\eta+\Delta\sigma^{2}\right)^{5/2}n!}\exp\left(-\frac{y^{2}}{2\left(n\eta+\Delta\sigma^{2}\right)}\right)\left(y^{2}-\left(n\eta+\Delta\sigma^{2}\right)\right).$$

Since the presence of the jumps cannot increase the information we have about σ^2 relative to the no-jumps case, it must be that

$$I_{\sigma^{2}}\leq\frac{1}{2\sigma^{4}}. \tag{B.3}$$

The idea is now to integrate (B.1) on a restricted subset of the real line, $(-a_\Delta,+a_\Delta)$, yielding from the positivity of the integrand

$$I_{\sigma^{2}}=\int_{-\infty}^{+\infty}\frac{\left(\frac{\partial q(y|\Delta;\theta)}{\partial\sigma^{2}}\right)^{2}}{q\left(y|\Delta;\theta\right)}dy\geq\int_{-a_{\Delta}}^{+a_{\Delta}}\frac{\left(\frac{\partial q(y|\Delta;\theta)}{\partial\sigma^{2}}\right)^{2}}{q\left(y|\Delta;\theta\right)}dy \tag{B.4}$$

and then to select a_Δ small enough that $q\left(y|\Delta;\theta\right)$ has a simpler expression on $(-a_\Delta,+a_\Delta)$ yet with enough of the support of the density included in $(-a_\Delta,+a_\Delta)$ that

$$\int_{-a_{\Delta}}^{+a_{\Delta}}\frac{\left(\frac{\partial q(y|\Delta;\theta)}{\partial\sigma^{2}}\right)^{2}}{q\left(y|\Delta;\theta\right)}dy=\frac{1}{2\sigma^{4}}+o(\Delta).$$

Combining the upper and lower bounds (B.3)-(B.4) will give the desired result

$$I_{\sigma^{2}}=\frac{1}{2\sigma^{4}}+o(\Delta) \tag{B.5}$$

which, in light of (4.9), will prove the Theorem.

Set a_Δ to be the positive solution of

$$q_{0}(a_{\Delta}|,\Delta;\theta)=q_{1}(a_{\Delta}|,\Delta;\theta), \tag{B.6}$$

that is

$$\begin{aligned} a_{\Delta} &= \Delta^{1/2}\left(\eta+\Delta\sigma^{2}\right)\sigma\eta^{-1/2}\left[\ln\left(1+\frac{\eta}{\Delta\sigma^{2}}\right)\right]^{1/2} \\ &= \sigma\left[-\Delta\ln\left(\Delta\right)\right]^{1/2}\left(1+o(1)\right). \end{aligned}$$

For all $y\in(-a_\Delta,+a_\Delta)$, we have

$$q_{0}(y|,\Delta;\theta)>q_{1}(y|,\Delta;\theta)>...>q_{n}(y|,\Delta;\theta)>...$$

and so from (B.2)

$$\frac{1}{q\left(y|\Delta;\theta\right)}\geq\frac{1}{q_{0}\left(y|\Delta;\theta\right)}.$$

Therefore

$$
\begin{aligned}
\int_{-a_\Delta}^{+a_\Delta} \frac{\left(\frac{\partial q(y|\Delta;\theta)}{\partial\sigma^2}\right)^2}{q\left(y|\Delta;\theta\right)} dy &\geq \int_{-a_\Delta}^{+a_\Delta} \frac{\left(\frac{\partial q(y|\Delta;\theta)}{\partial\sigma^2}\right)^2}{q_0\left(y|\Delta;\theta\right)} dy \\
&= \int_{-a_\Delta}^{+a_\Delta} \frac{\left(\sum_{n=0}^{+\infty} \frac{\exp(-\lambda\Delta)(\lambda\Delta)^n\Delta}{2\sqrt{2\pi}(n\eta+\Delta\sigma^2)^{5/2}n!} \exp\left(-\frac{y^2}{2(n\eta+\Delta\sigma^2)}\right)\left(y^2-(n\eta+\Delta\sigma^2)\right)\right)^2}{\frac{1}{\sqrt{2\pi}\sqrt{\Delta\sigma^2}}\exp\left(-\frac{y^2}{2\Delta\sigma^2}\right)} dy \\
&= \int_{-a_\Delta}^{+a_\Delta} \left(\sum_{n=0}^{+\infty} \frac{\sqrt{2\pi}e^{-\lambda\Delta}(\lambda\Delta)^n\Delta\left(\Delta\sigma^2\right)^{1/4}}{2\left(n\eta+\Delta\sigma^2\right)^{5/2}n!} e^{-\frac{y^2}{2(n\eta+\Delta\sigma^2)}+\frac{y^2}{4\Delta\sigma^2}}\left(y^2-(n\eta+\Delta\sigma^2)\right)\right)^2 dy \\
&\equiv \int_{-a_\Delta}^{+a_\Delta} \left(\sum_{n=0}^{+\infty} f_n(y|,\Delta;\theta)\right)^2 dy \\
&= \sum_{n=0}^{+\infty}\sum_{m=0}^{+\infty}\int_{-a_\Delta}^{+a_\Delta} f_n(y|,\Delta;\theta)f_m(y|,\Delta;\theta)dy
\end{aligned}
$$

The leading term in that double sum comes from $n = m = 0$ and is the only one with a final limit as $\Delta \to 0$:

$$
\sum_{n=0}^{+\infty}\sum_{m=0}^{+\infty}\int_{-a_\Delta}^{+a_\Delta} f_n(y|,\Delta;\theta)f_m(y|,\Delta;\theta)dy = \int_{-a_\Delta}^{+a_\Delta} f_0^2(y|,\Delta;\theta)dy + o(1)
$$

with

$$
\begin{aligned}
\int_{-a_\Delta}^{+a_\Delta} f_0^2(y|,\Delta;\theta)dy &= \frac{e^{-2\Delta\lambda}}{2\sqrt{2\pi}\eta^{3/2}\left(1+\frac{\eta}{\Delta\sigma^2}\right)^{\frac{\eta+\Delta\sigma^2}{2\eta}}\left(\sigma^2\right)^{3/2}} \\
&\quad\times\left(\sqrt{2\pi}\eta^{3/2}\left(1+\frac{\eta}{\Delta\sigma^2}\right)^{\frac{\eta+\Delta\sigma^2}{2\eta}}\left(2\Phi\left(\eta^{-1/2}\sqrt{\eta+\Delta\sigma^2}\left[\ln\left(1+\frac{\eta}{\Delta\sigma^2}\right)\right]^{1/2}\right)-1\right)\right. \\
&\quad\left.-\sqrt{\eta+\Delta\sigma^2}\left[\ln\left(1+\frac{\eta}{\Delta\sigma^2}\right)\right]^{1/2}\left(\eta+\left(\eta+\Delta\sigma^2\right)\ln\left(1+\frac{\eta}{\Delta\sigma^2}\right)\right)\right) \\
&= \frac{1}{2\sigma^4}+o(\Delta)
\end{aligned}
$$

Therefore

$$
\int_{-a_\Delta}^{+a_\Delta} \frac{\left(\frac{\partial q(y|\Delta;\theta)}{\partial\sigma^2}\right)^2}{q_0\left(y|\Delta;\theta\right)} dy = \frac{1}{2\sigma^4}+o(\Delta)
$$

which achieves the proof.

Appendix C: Proof of Propositions 2, 3 and 4

In all cases, what needs to be computed are the matrices D and S. With polynomial moment functions in h (including possibly absolute values and non integer powers) of the type

$$
h(y,\delta,\theta) = y^r - \Delta\left(\delta,\theta,r\right) \tag{C.1}
$$

or

$$h(y,\delta,\theta)=|y|^{r}-\Delta_{a}\left(\delta,\theta,r\right), \tag{C.2}$$

the functionals $\dot{h}$ and hh' retain the polynomial form in y. Thus D and S can be computed explicitly using the moments $\Delta\left(\Delta,\theta,r\right)$ and $\Delta_{a}\left(\Delta,\theta,r\right)$ calculated in Proposition 1. Indeed, say we used the single (C.2) as our moment condition. Then

$$D=E\left[\dot{h}(Y_{\Delta},\Delta,\theta)\right]=-\Delta_{a}\left(\delta,\theta,r\right) \tag{C.3}$$

which can be calculated by differentiating with respect to the parameter of the expression for $\Delta_{a}\left(\Delta,\theta,r\right)$ given in Proposition 1 and

$$\begin{aligned} S &= E\left[h(Y_{\Delta},\Delta,\theta)h(Y_{\Delta},\Delta,\theta)'\right] \\ &= E\left[\left(|Y_{\Delta}|^{r}-\Delta_{a}\left(\Delta,\theta,r\right)\right)^{2}\right] \\ &= E\left[|Y_{\Delta}|^{2r}\right]-2E\left[|Y_{\Delta}|^{r}\right]\Delta_{a}\left(\Delta,\theta,r\right)+\Delta_{a}\left(\Delta,\theta,r\right)^{2} \\ &= \Delta_{a}\left(\Delta,\theta,2r\right)-\Delta_{a}\left(\Delta,\theta,r\right)^{2} \end{aligned} \tag{C.4}$$

which is calculated using again the expressions for the moments in Proposition 1.

Given D and S in (C.3)-(C.4), I then calculate a Taylor expansion in Δ for Ω in (4.7),

$$\Omega=\Delta\left(D'S^{-1}D\right)^{-1}=\Delta\frac{S^{2}}{D^{2}}.$$

The leading terms of the Taylor expansions of the moments are given in Proposition 1. In the vector case, repeat the calculations (C.3)-(C.4) for each element of the matrices D and S. The leading term of the Taylor expansion of the AVAR matrix Ω is reported for each combination of moments and parameters in the three propositions.

Appendix D: Proof of Theorem 2

The essence of the argument is to compute the leading term of Fisher's Information by using the convergence of the Cauchy density as $\Delta\rightarrow 0$ to a Dirac delta function, using the Brownian density as the test function, after a change of variable, to get a fixed function. Fisher's Information for σ^{2} is

$$I_{\sigma^{2}}=\int_{-\infty}^{+\infty}\left(\frac{\partial q\left(y|\Delta;\theta\right)}{\partial\sigma^{2}}\right)^{2}\frac{dy}{q\left(y|\Delta;\theta\right)}$$

Replace now q, the density of the Brownian plus Cauchy process, by its expression (6.12), yielding

$$\frac{\partial q\left(y|\Delta;\theta\right)}{\partial\sigma^{2}}=\int_{-\infty}^{+\infty}\frac{\alpha}{2\sqrt{2\pi}\Delta^{1/2}(\sigma^{2})^{5/2}}e^{-\frac{(y-z)^{2}}{2\Delta\sigma^{2}}}\frac{\left((y-z)^{2}-\Delta\sigma^{2}\right)}{\Delta^{2}\alpha^{2}\pi^{2}+z^{2}}dz$$

and therefore

$$
\begin{aligned}
I_{\sigma^2} &= \int_{-\infty}^{+\infty}\left(\frac{\partial q\left(y|\Delta;\theta\right)}{\partial\sigma^2}\right)^2\frac{dy}{q\left(y|\Delta;\theta\right)} \\
&= \left(\frac{\alpha}{2\sqrt{2\pi}\Delta^{1/2}(\sigma^2)^{5/2}}\right)^2\frac{\sqrt{2\pi}\sqrt{\sigma^2}}{\Delta^{1/2}\alpha} \\
&\quad\times\int_{-\infty}^{+\infty}\frac{\left(\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right)\frac{\left((y-z)^2-\Delta\sigma^2\right)}{\Delta^2\alpha^2\pi^2+z^2}dz\right)^2}{\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right)\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dz}dy \\
&= \frac{\alpha}{4\sqrt{2\pi}\Delta^{3/2}(\sigma^2)^{9/2}}\times\int_{-\infty}^{+\infty}\frac{\left(\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right)\frac{\left((y-z)^2-\Delta\sigma^2\right)}{\Delta^2\alpha^2\pi^2+z^2}dz\right)^2}{\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right)\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dz}dy
\end{aligned}
$$

Concentrating on the numerator inside the integral, we have

$$
\begin{aligned}
\left(\int_{-\infty}^{+\infty}e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{\left((y-z)^2-\Delta\sigma^2\right)}{\Delta^2\alpha^2\pi^2+z^2}dz\right)^2 &= \left(\int_{-\infty}^{+\infty}e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{(y-z)^2}{\Delta^2\alpha^2\pi^2+z^2}dz\right)^2 \\
&\quad -2\Delta\sigma^2\left(\int_{-\infty}^{+\infty}e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{(y-z)^2}{\Delta^2\alpha^2\pi^2+z^2}dz\right) \\
&\quad\times\left(\int_{-\infty}^{+\infty}e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dz\right) \\
&\quad +\Delta^2\sigma^4\left(\int_{-\infty}^{+\infty}e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dz\right)^2
\end{aligned}
$$

so that

$$
\begin{aligned}
\int_{-\infty}^{+\infty}\frac{\left(\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right)\frac{\left((y-z)^2-\Delta\sigma^2\right)}{\Delta^2\alpha^2\pi^2+z^2}dz\right)^2}{\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right)\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dz}dy &= \int_{-\infty}^{+\infty}\frac{\left(\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right)\frac{(y-z)^2}{\Delta^2\alpha^2\pi^2+z^2}dz\right)^2}{\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right)\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dz}dy \\
&\quad -2\Delta\sigma^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{(y-z)^2}{\Delta^2\alpha^2\pi^2+z^2}dzdy \\
&\quad +\Delta^2\sigma^4\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dzdy \\
&\equiv A+B+C
\end{aligned}
$$

From

$$
\int_{-\infty}^{+\infty}q\left(y|\Delta;\theta\right)dy=1,
$$

it follows that

$$
\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dzdy=\frac{\sqrt{2\pi}(\sigma^2)^{1/2}}{\Delta^{1/2}\alpha} \tag{D.1}
$$

and hence

$$C = \Delta^2\sigma^4 \left(\frac{\sqrt{2\pi}(\sigma^2)^{1/2}}{\Delta^{1/2}\alpha} \right) = \frac{\Delta^{3/2}\sqrt{2\pi}(\sigma^2)^{5/2}}{\alpha}.$$

Similarly, from the expected value of the score being zero, i.e.,

$$\int_{-\infty}^{+\infty} \frac{\partial q\,(y|\Delta;\theta)}{\partial\sigma^2} dy = 0,$$

it follows that

$$\begin{aligned}
\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(y-z)^2}{2\Delta\sigma^2}} \frac{(y-z)^2}{\Delta^2\alpha^2\pi^2+z^2} dzdy &= \Delta\sigma^2 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(y-z)^2}{2\Delta\sigma^2}} \frac{1}{\Delta^2\alpha^2\pi^2+z^2} dzdy \\
&= \frac{\Delta^{1/2}\sqrt{2\pi}(\sigma^2)^{3/2}}{\alpha} \qquad \text{(D.2)}
\end{aligned}$$

and hence

$$B = -2\Delta\sigma^2 \left(\frac{\Delta^{1/2}\sqrt{2\pi}(\sigma^2)^{3/2}}{\alpha} \right) = -2 \left(\frac{\Delta^{3/2}\sqrt{2\pi}(\sigma^2)^{5/2}}{\alpha} \right).$$

Thus

$$\begin{aligned}
I_{\sigma^2} &= \frac{\alpha}{4\sqrt{2\pi}\Delta^{3/2}(\sigma^2)^{9/2}} \times (A+B+C) \\
&= \frac{\alpha}{4\sqrt{2\pi}\Delta^{3/2}(\sigma^2)^{9/2}} \times A + \frac{\alpha}{4\sqrt{2\pi}\Delta^{3/2}(\sigma^2)^{9/2}} \left(-2\left(\frac{\Delta^{3/2}\sqrt{2\pi}(\sigma^2)^{5/2}}{\alpha} \right) + \frac{\Delta^{3/2}\sqrt{2\pi}(\sigma^2)^{5/2}}{\alpha} \right) \\
&= \frac{\alpha}{4\sqrt{2\pi}\Delta^{3/2}(\sigma^2)^{9/2}} \times A - \frac{1}{4\sigma^4} \\
&\equiv \tilde{A} - \frac{1}{4\sigma^4} \qquad \text{(D.3)}
\end{aligned}$$

That leaves us with the computation of

$$\tilde{A} \equiv \frac{\alpha}{4\sqrt{2\pi}\Delta^{3/2}(\sigma^2)^{9/2}} \times A = \frac{\alpha}{4\sqrt{2\pi}\Delta^{3/2}(\sigma^2)^{9/2}} \int_{-\infty}^{+\infty} \frac{\left(\int_{-\infty}^{+\infty} \exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right) \frac{(y-z)^2}{\Delta^2\alpha^2\pi^2+z^2} dz \right)^2}{\int_{-\infty}^{+\infty} \exp\left(-\frac{(y-z)^2}{2\Delta\sigma^2}\right) \frac{1}{\Delta^2\alpha^2\pi^2+z^2} dz} dy.$$

To handle the integral A, I first do two changes of variable: from z to $w = z/(\Delta\sigma^2)^{1/2}$ in the two inner integrals and from y to $x = y/(\Delta\sigma^2)^{1/2}$ in the outer integral, yielding

$$\begin{aligned}
A &= \int_{-\infty}^{+\infty} \frac{\left(\int_{-\infty}^{+\infty} \exp\left(-\frac{\left(y-w(\Delta\sigma^2)^{1/2}\right)^2}{2\Delta\sigma^2}\right) \frac{\left(y-w(\Delta\sigma^2)^{1/2}\right)^2}{\Delta^2\alpha^2\pi^2+\Delta\sigma^2 w^2} dw (\Delta\sigma^2)^{1/2}\right)^2}{\int_{-\infty}^{+\infty} \exp\left(-\frac{\left(y-w(\Delta\sigma^2)^{1/2}\right)^2}{2\Delta\sigma^2}\right) \frac{1}{\Delta^2\alpha^2\pi^2+\Delta\sigma^2 w^2} dw (\Delta\sigma^2)^{1/2}} dy \\
&= \int_{-\infty}^{+\infty} \frac{\left(\int_{-\infty}^{+\infty} \exp\left(-\frac{\left(x(\Delta\sigma^2)^{1/2}-w(\Delta\sigma^2)^{1/2}\right)^2}{2\Delta\sigma^2}\right) \frac{\left(x(\Delta\sigma^2)^{1/2}-w(\Delta\sigma^2)^{1/2}\right)^2}{\Delta^2\alpha^2\pi^2+\Delta\sigma^2 w^2} dw\right)^2}{\int_{-\infty}^{+\infty} \exp\left(-\frac{\left(x(\Delta\sigma^2)^{1/2}-w(\Delta\sigma^2)^{1/2}\right)^2}{2\Delta\sigma^2}\right) \frac{1}{\Delta^2\alpha^2\pi^2+\Delta\sigma^2 w^2} dw} (\Delta\sigma^2)^{1/2} dx (\Delta\sigma^2)^{1/2} \\
&= \left(\Delta\sigma^2\right)^3 \int_{-\infty}^{+\infty} \frac{\left(\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \frac{(x-w)^2}{\Delta^2\alpha^2\pi^2+\Delta\sigma^2 w^2} dw\right)^2}{\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \frac{1}{\Delta^2\alpha^2\pi^2+\Delta\sigma^2 w^2} dw} dx \\
&= \Delta^2\sigma^6 \int_{-\infty}^{+\infty} \frac{\left(\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \frac{(x-w)^2}{\Delta\alpha^2\pi^2+\sigma^2 w^2} dw\right)^2}{\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \frac{1}{\Delta\alpha^2\pi^2+\sigma^2 w^2} dw} dx
\end{aligned}$$

so that

$$\begin{aligned}
\tilde{A} &= \frac{\Delta^{1/2}\alpha}{4\sqrt{2\pi}(\sigma^2)^{3/2}} \int_{-\infty}^{+\infty} \frac{\left(\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \frac{(x-w)^2}{\Delta\alpha^2\pi^2+\sigma^2 w^2} dw\right)^2}{\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \frac{1}{\Delta\alpha^2\pi^2+\sigma^2 w^2} dw} dx \\
&= \frac{1}{4\sqrt{2\pi}\sigma^4} \int_{-\infty}^{+\infty} \frac{\left(\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) (x-w)^2 \omega_\Delta(w)\, dw\right)^2}{\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \omega_\Delta(w)\, dw} dx
\end{aligned}$$

where

$$\omega_\Delta(w) \equiv \frac{\alpha\Delta^{1/2}\sigma}{\Delta\alpha^2\pi^2+\sigma^2 w^2}.$$

The function ω_Δ integrates to 1, and as Δ tends to zero, it converges to a Dirac point mass at $w = 0$. Therefore

$$\begin{aligned}
\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \omega_\Delta(w)\, dw &= \exp\left(-\frac{x^2}{2}\right) + o(1) \\
\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) (x-w)^2 \omega_\Delta(w)\, dw &= \exp\left(-\frac{x^2}{2}\right) x^2 + o(1)
\end{aligned}$$

Let

$$h_\Delta(x) \equiv \frac{\left(\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) (x-w)^2 \omega_\Delta(w)\, dw\right)^2}{\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-w)^2}{2}\right) \omega_\Delta(w)\, dw}.$$

From the limits above we have that

$$h_\Delta(x) = \frac{\left(\exp\left(-\frac{x^2}{2}\right) x^2\right)^2}{\exp\left(-\frac{x^2}{2}\right)} + o(1) = \exp\left(-x^2\right) x^4 + o(1) \equiv h_0(x) + o(1).$$

To establish that

$$\int_{-\infty}^{+\infty} h_\Delta(x)dx = \int_{-\infty}^{+\infty} h_0(x)dx + o(1) \tag{D.4}$$
$$= 3\sqrt{2\pi} + o(1)$$

I apply Fatou's Lemma (see e.g., 6.8.8 in Haaser and Sullivan (1991)) to yield

$$\int_{-\infty}^{+\infty} h_0(x)dx \leq \underline{\lim} \int_{-\infty}^{+\infty} h_\Delta(x)dx \tag{D.5}$$

To obtain an upper bound for the r.h.s. of (D.5), I use the Cauchy-Schwarz Inequality as follows

$$\frac{\left(\int_{-\infty}^{+\infty} u(w,x)dw\right)^2}{\int_{-\infty}^{+\infty} v(w,x)dw} \leq \int_{-\infty}^{+\infty} \frac{u(w,x)^2}{v(w,x)}dw$$

to obtain

$$\begin{aligned}
\int_{-\infty}^{+\infty} h_\Delta(x)dx &\leq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\left(\exp\left(-\frac{(x-w)^2}{2}\right)(x-w)^2\omega_\Delta(w)\right)^2}{\exp\left(-\frac{(x-w)^2}{2}\right)\omega_\Delta(w)} dwdx \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(x-w)^2}{2}}(x-w)^4\omega_\Delta(w)\,dwdx \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(x-w)^2}{2}}(x-w)^4\frac{\alpha\Delta^{1/2}\sigma}{\Delta\alpha^2\pi^2+\sigma^2w^2}dwdx
\end{aligned}$$

with the change of variable from w to $z = w - x$, followed by a change of variable from x to $y = -x$.

From

$$\begin{aligned}
0 &= \int_{-\infty}^{+\infty} \frac{\partial^2 q(y|\Delta;\theta)}{\partial(\sigma^2)^2}dy \\
&= \frac{1}{4\sqrt{2\pi}\sigma^4\Delta^{5/2}\sigma^9}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\left(\frac{(y-z)^4 - 6(y-z)^2\Delta\sigma^2 + 3\Delta^2\sigma^4}{\Delta^2\alpha^2\pi^2+z^2}\right)dzdy
\end{aligned}$$

it follows that

$$\begin{aligned}
\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{(y-z)^4}{\Delta^2\alpha^2\pi^2+z^2}dzdy &= 6\Delta\sigma^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{(y-z)^2}{\Delta^2\alpha^2\pi^2+z^2}dzdy \\
&\quad -3\Delta^2\sigma^4\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(y-z)^2}{2\Delta\sigma^2}}\frac{1}{\Delta^2\alpha^2\pi^2+z^2}dzdy \\
&= 6\Delta\sigma^2\left(\frac{\Delta^{1/2}\sqrt{2\pi}(\sigma^2)^{3/2}}{\alpha}\right) - 3\Delta^2\sigma^4\left(\frac{\sqrt{2\pi}(\sigma^2)^{1/2}}{\Delta^{1/2}\alpha}\right) \\
&= 3\frac{\Delta^{3/2}\sqrt{2\pi}(\sigma^2)^{5/2}}{\alpha}
\end{aligned} \tag{D.6}$$

with the last step from (D.1) and (D.2). Using now the changes of variables from w to $z = w(\Delta\sigma^2)^{1/2}$ in the inner integral and from x to $y = x(\Delta\sigma^2)^{1/2}$ in the outer integral yields

$$\begin{aligned}
& \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(x-w)^2}{2}}(x-w)^4 \frac{\alpha\Delta^{1/2}\sigma}{\Delta\alpha^2\pi^2+\sigma^2 w^2} dwdx \\
= \quad & \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(y-z)^2}{2\Delta\sigma^2}} \frac{(y-z)^4}{(\Delta\sigma^2)^2} \frac{\alpha\Delta^{1/2}\sigma}{\Delta\alpha^2\pi^2+\sigma^2 z^2/(\Delta\sigma^2)} \frac{dz}{(\Delta\sigma^2)^{1/2}} \frac{dy}{(\Delta\sigma^2)^{1/2}} \\
= \quad & \frac{\alpha}{\Delta^{3/2}(\sigma^2)^{5/2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(y-z)^2}{2\Delta\sigma^2}} \frac{(y-z)^4}{\Delta^2\alpha^2\pi^2+z^2} dzdy \\
= \quad & \frac{\alpha}{\Delta^{3/2}(\sigma^2)^{5/2}} \left(3\frac{\Delta^{3/2}\sqrt{2\pi}(\sigma^2)^{5/2}}{\alpha}\right) \\
= \quad & 3\sqrt{2\pi}
\end{aligned}$$

proving that

$$\int_{-\infty}^{+\infty} h_\Delta(x)dx \leq 3\sqrt{2\pi}. \tag{D.7}$$

Combined with (D.5), we have therefore

$$3\sqrt{2\pi} = \int_{-\infty}^{+\infty} h_0(x)dx \leq \lim \int_{-\infty}^{+\infty} h_\Delta(x)dx \leq 3\sqrt{2\pi}$$

which establishes (D.4):

$$\lim \int_{-\infty}^{+\infty} h_\Delta(x)dx = 3\sqrt{2\pi}.$$

Hence I have obtained that

$$\tilde{A} = \frac{1}{4\sqrt{2\pi}\sigma^4}\int_{-\infty}^{+\infty} h_\Delta(x)dx = \frac{3}{4\sigma^4} + o(1)$$

and from (D.3) Fisher's Information for σ^2 is

$$I_{\sigma^2} = \tilde{A} - \frac{1}{4\sigma^4} = \frac{1}{2\sigma^4} + o(1)$$

which, in light of (4.9), proves the Theorem.

Appendix E: Proof of Lemma 1

For a Brownian motion with density

$$f_{W_\Delta}(y) = \frac{1}{\sqrt{2\pi}\sqrt{\Delta\sigma^2}} \exp\left(-\frac{y^2}{2\Delta\sigma^2}\right)$$

we have

$$
\begin{aligned}
\Pr\left(|Y_{\Delta}|>\varepsilon\right) &= \int_{|y|>\varepsilon} f_{W_{\Delta}}(y)dy \\
&= 2\Phi\left(\frac{\varepsilon}{\Delta^{1/2}\sigma}\right)-1 \\
&= \frac{\Delta^{1/2}\sigma}{\varepsilon}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\varepsilon^{2}}{2\,\Delta\,\sigma^{2}}\right)(1+o(1))
\end{aligned}
$$

with the last equation following from the known asymptotic behavior of the Normal cdf Φ near infinity (see e.g., 26.2.12 in Abramowitz and Stegun (1972)):

$$
1-\Phi\left(x\right)=\frac{\phi\left(x\right)}{x}\left(1+o(1)\right)
$$

as $x \to +\infty$, where ϕ is the Normal pdf.

For a Lévy pure jump process with jump measure $v(dz)$ and probability measure $f_{L_{\Delta}}(dy)$ it is known that for points $y \neq 0$, under regularity conditions,

$$
f_{L_{\Delta}}(dy)=\Delta\times v(dy)+o(\Delta)
$$

(see e.g., Corollary 1 in Rüschendorf and Woerner (2002) as a special case for Lévy processes of Léandre (1987) and Picard (1997) for points that can be reached in one jump from 0). The regularity conditions referred to in the statement of the lemma are those of Theorem 1 in Rüschendorf and Woerner (2002). They are satisfied for instance by the symmetric stable class emphasized here.

Then, by Fatou's Lemma,

$$
\begin{aligned}
\Pr\left(|Y_{\Delta}|>\varepsilon\right) &= \int_{|y|>\varepsilon} f_{L_{\Delta}}(dy) \\
&= \Delta\times\int_{|y|>\varepsilon} v(dy)+o(\Delta).
\end{aligned}
$$

References

Abramowitz, M., Stegun, I. A., 1972. Handbook of Mathematical Functions. Dover, New York, N.Y.

Aït-Sahalia, Y., 2002a. Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., 2002b. Telling from discrete data whether the underlying continuous-time model is a diffusion. Journal of Finance 57, 2075–2112.

Aït-Sahalia, Y., Mykland, P. A., 2003a. The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica 71, 483–549.

Aït-Sahalia, Y., Mykland, P. A., 2003b. How often to sample a continuous-time process in the presence of market microstructure noise. Tech. rep., Princeton University.

Andersen, T. G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115–158.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Ball, C. A., Torous, W. N., 1983. A simplified jump process for common stock returns. The Journal of Financial and Quantitative Analysis 18, 53–65.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Power variation with stochastic volatility and jumps. Tech. rep., University of Aarhus.

Beckers, S., 1981. A note on estimating the parameters of the diffusion-jump model of stock returns. The Journal of Financial and Quantitative Analysis 16, 127–140.

Bertoin, J., 1998. Lévy Processes. Cambridge University Press, Cambridge, UK.

Carr, P., Geman, H., Madan, D. B., Yor, M., 2002. The fine structure of asset returns: An empirical investigation. Journal of Business 75, 305–332.

Carr, P., Wu, L., 2003a. Time-changed Lévy processes and option pricing. Journal of Financial Economics, forthcoming , xxx–xxx.

Carr, P., Wu, L., 2003b. What type of process underlies options? A simple robust test. Journal of Finance, forthcoming , xxx–xxx.

Chan, K., 1999. Pricing contingent claims on stocks driven by Lévy processes. Annals of Applied Probability 9, 504–528.

Chernov, M., Gallant, A. R., Ghysels, E., Tauchen, G. T., 2002. Alternative models of stock price dynamics. Journal of Econometrics, forthcoming , xxx–xxx.

Ding, Z., Granger, C. W., Engle, R. F., 1993. A long memory property of stock market returns and a new model markets. Journal of Empirical Finance 1, 83–106.

Eberlein, E., Keller, U., Prause, K., 1998. New insights into smile, mispricing and Value at Risk: The hyperbolic model. Journal of Business 71, 371–405.

Eberlein, E., Raible, S., 1999. Term structure models driven by general Lévy processes. Mathematical Finance 9, 31–53.

Eraker, B., Johannes, M. S., Polson, N., 2003. The impact of jumps in equity index volatility and returns. Journal of Finance, forthcoming , xxx–xxx.

Haaser, N. B., Sullivan, J. A., 1991. Real Analysis. Dover, New York.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Honoré, P., 1998. Pitfalls in estimating jump-diffusion models. Tech. rep., Aarhus School of Business.

Kiefer, N. M., 1978. Discrete parameter variation: Efficient estimation of a switching regression model. Econometrica 46, 427–434.

Léandre, R., 1987. Densite en temps petit d'un processus de sauts. In: Séminaire de Probabilités, XXI, Lecture Notes in Mathematics, 1247. Springer, Berlin, Germany, pp. 81–99.

Lepingle, D., 1976. La variation d'ordre p des semi-martingales. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 36, 295–316.

Madan, D. B., Carr, P. P., Chang, E. E., 1998. The variance gamma process and option pricing. European Finance Review 2, 79–105.

Merton, R. C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3, 125–144.

Picard, J., 1997. Density in small time for lévy processes. ESAIM Probability and Statistics 1, 357–389.

Press, S. J., 1967. A compound events model for security prices. The Journal of Business 40, 317–335.

Ray, D., 1956. Stationary markov processes with continuous paths. Transactions of the American Mathematical Society 82, 452–493.

Rüschendorf, L., Woerner, J. H. C., 2002. Expansions of transition distributions of lévy processes in small time. Bernoulli 8, 81–96.

Schaumburg, E., 2001. Maximum likelihood estimation of jump processes with applications to finance. Ph.D. thesis, Princeton University.

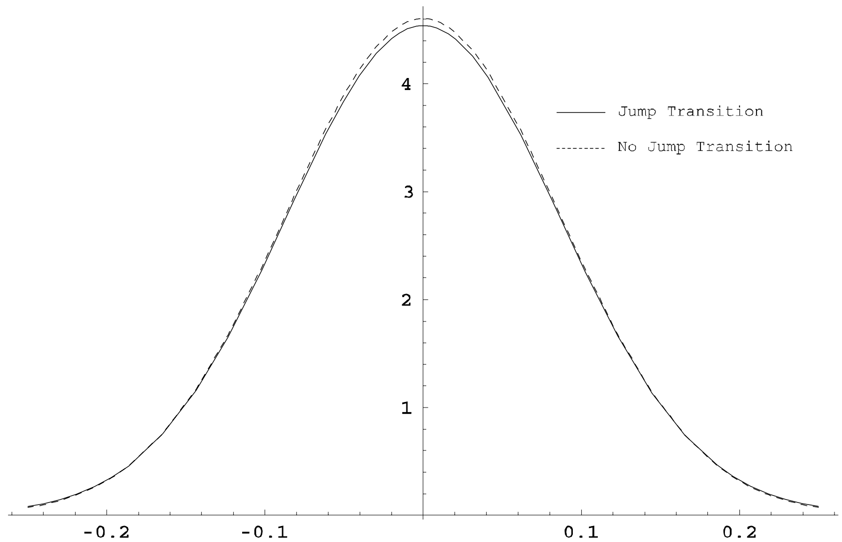


Fig. 1
The Transition Density

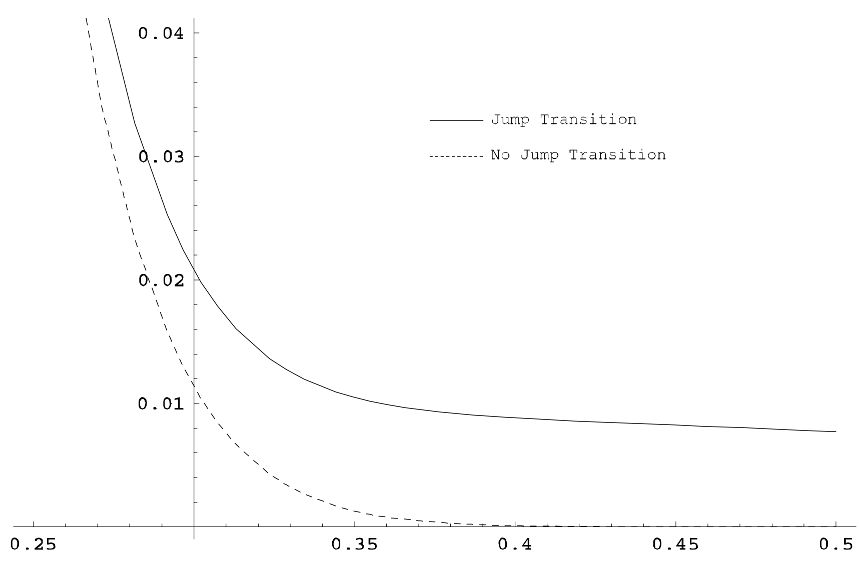


Fig. 2
Tail of the Transition Density

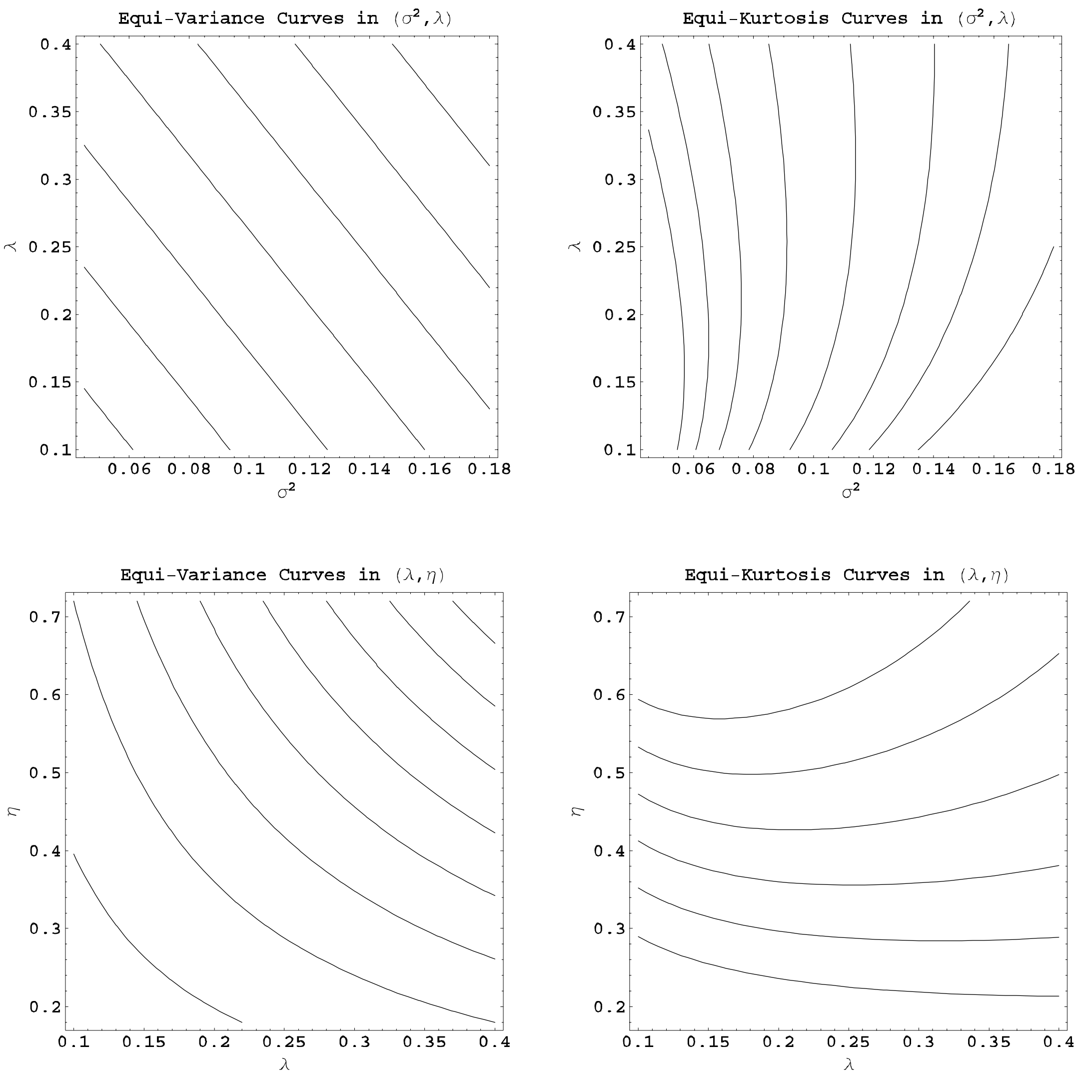


Fig. 3
Isonoise Curves

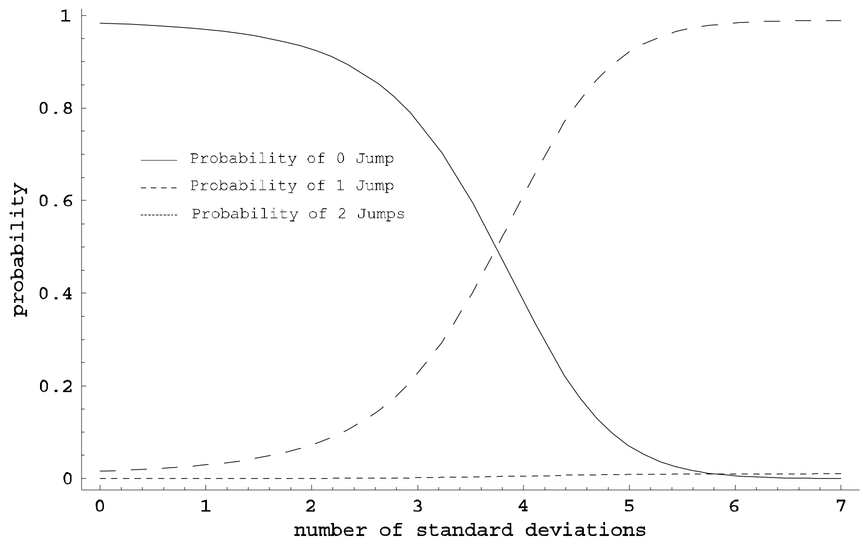


Fig. 4
Jump Probabilities Inferred from Observing a Jump Greater than a Given Threshold

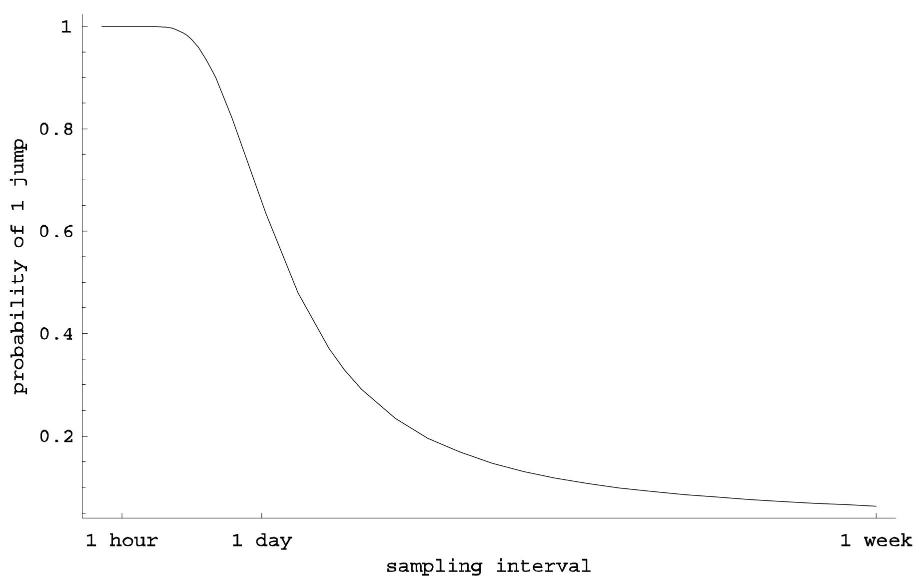


Fig. 5
Probability that a 10% Log-Return Involved 1 Jump as a Function of the Sampling Interval

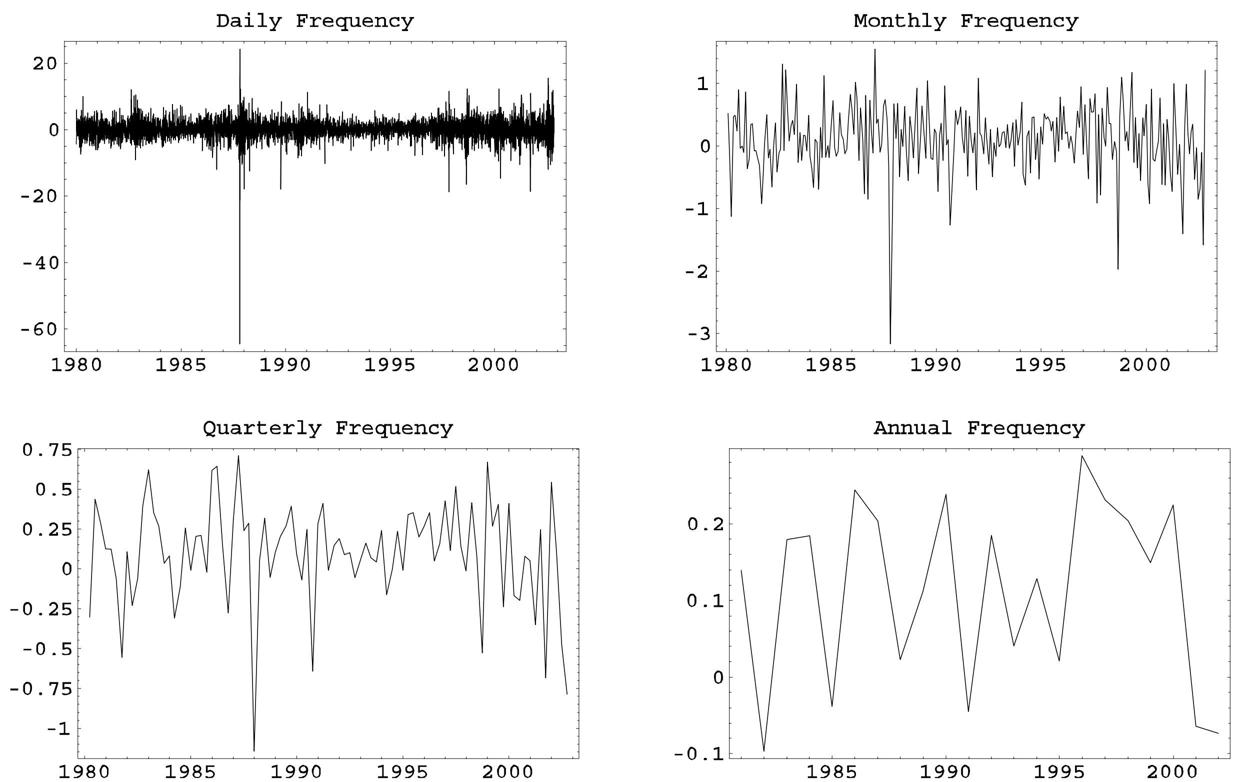


Fig. 6
Log-Returns on the Dow Jones Industrial Average at Different Observation Frequencies

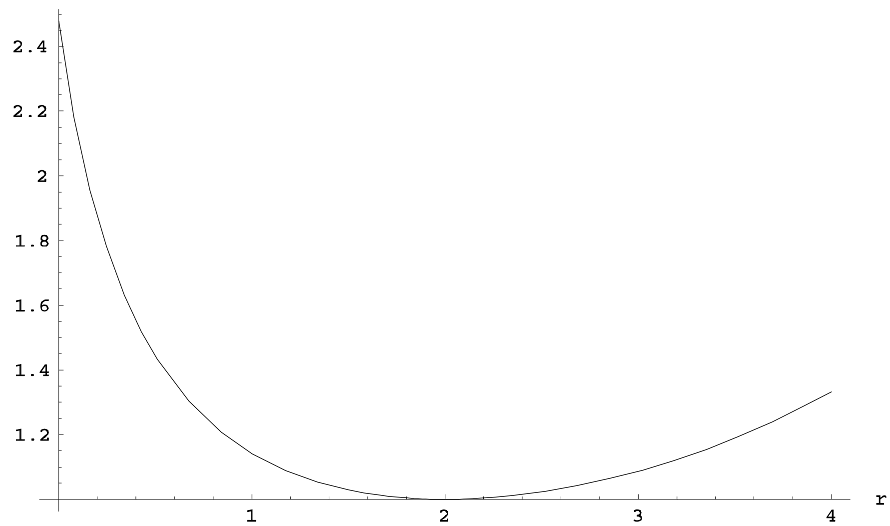


Fig. 7
Efficiency of the $M_a(\Delta, \theta, r)$ Moment Condition Relative to MLE in the Absence of Jumps

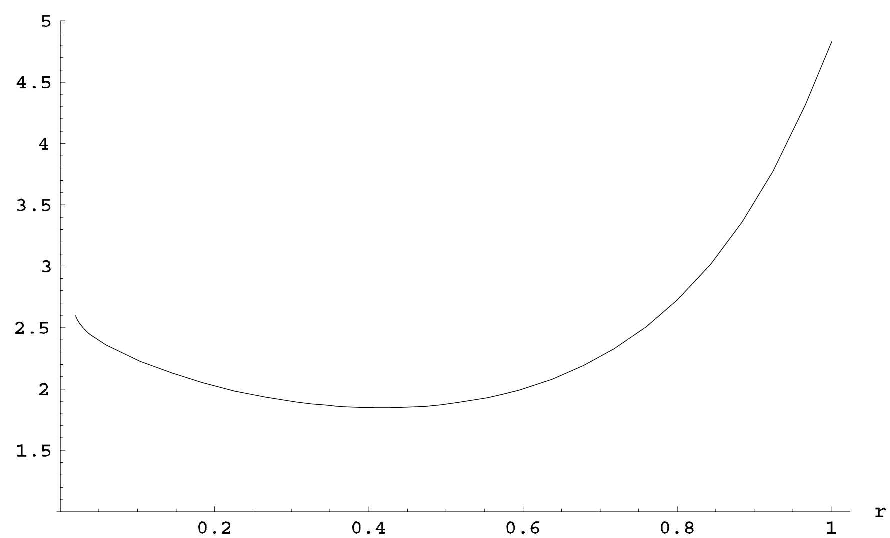


Fig. 8
Efficiency of the $M_a(\Delta, \theta, r)$ Moment Condition Relative to MLE When Jumps Are Present

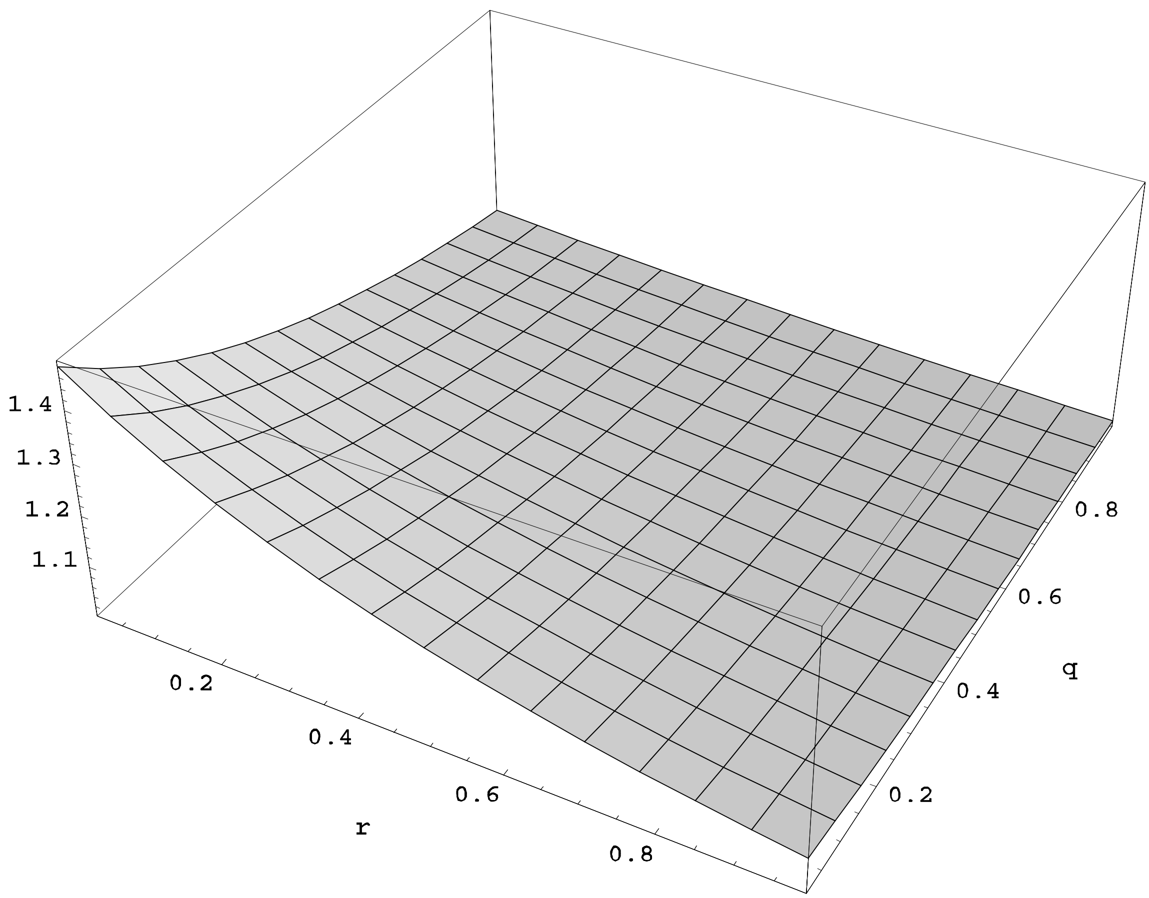


Fig. 9
Efficiency of the $(M_a(\Delta, \theta, r), M_a(\Delta, \theta, q))'$ Moment Conditions Used Jointly Relative to MLE in the Absence of Jumps

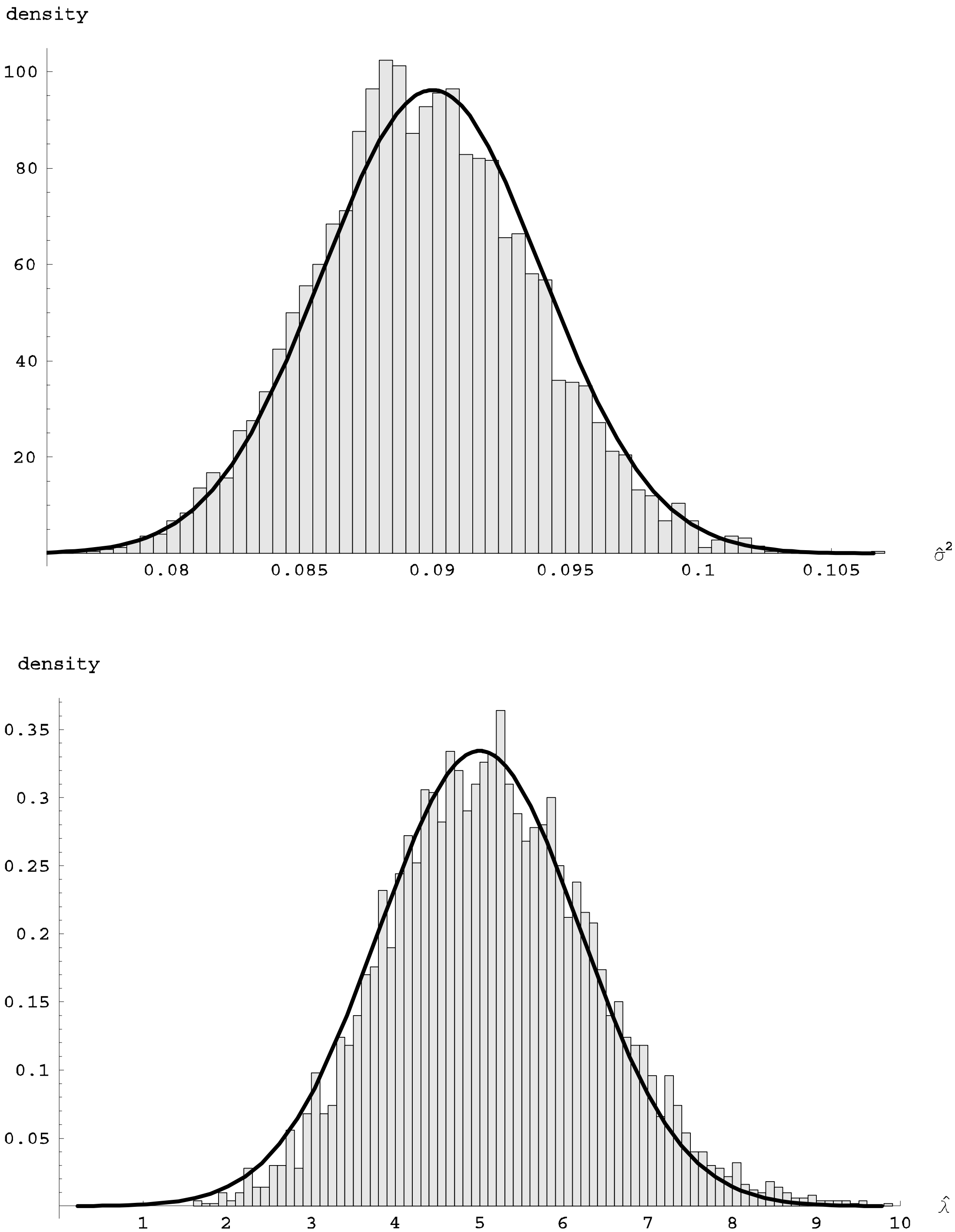
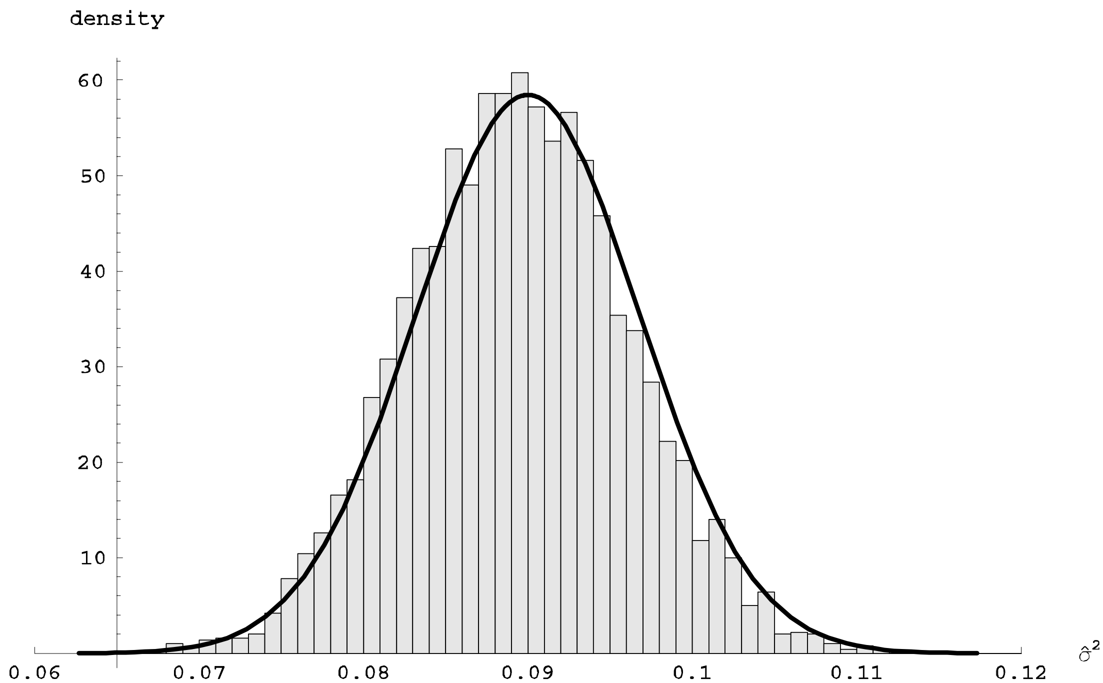


Fig. 10
Distinguishing Volatilty from Poisson Jumps: Small Sample and Asymptotic Distributions of the MLE Estimators of (σ^2, λ)



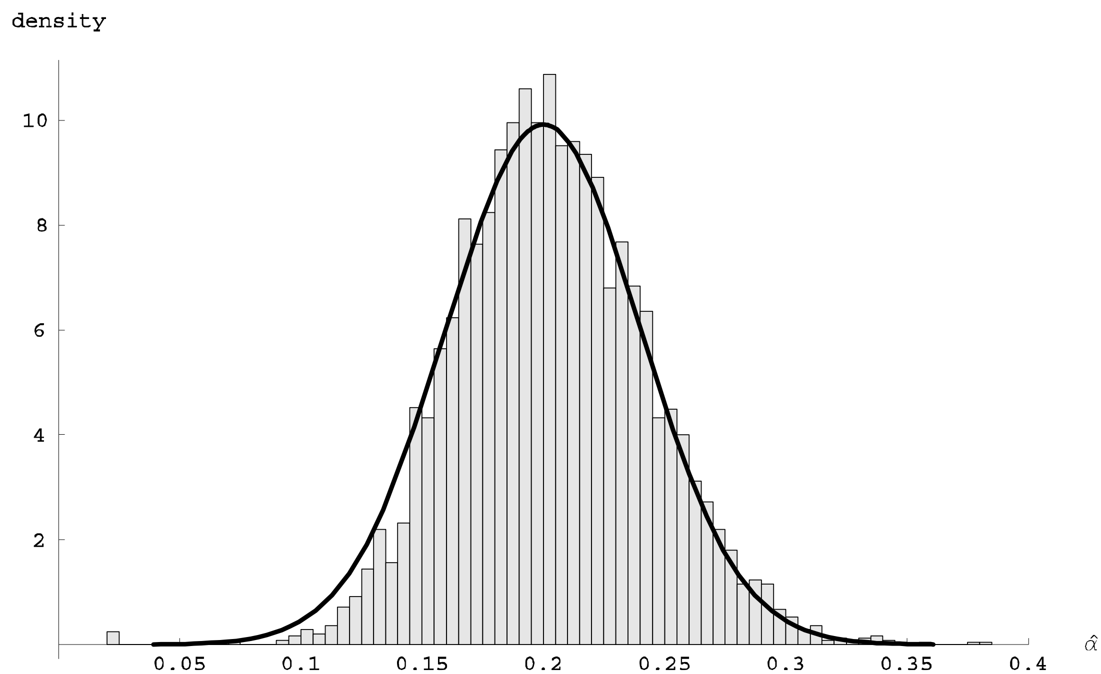


Fig. 11
Distinguishing Volatilty from Cauchy Jumps: Small Sample and Asymptotic Distributions of the MLE Estimators of (σ^2, α)

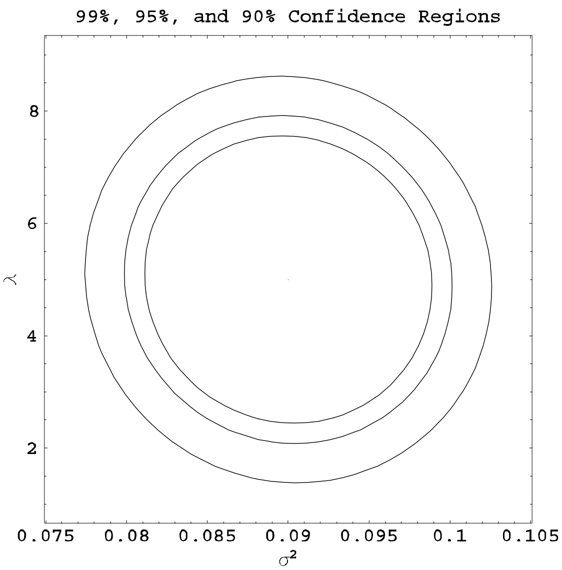


Fig. 12
Distinguishing Volatilty from Poisson Jumps: Confidence Regions for the MLE Estimates of (σ^2, λ)

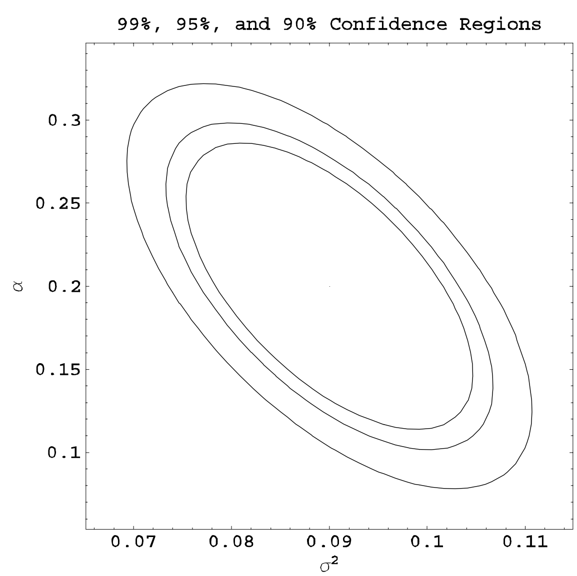


Fig. 13
Distinguishing Volatilty from Cauchy Jumps: Confidence Regions for the MLE Estimates of (σ^2, α)