Do Changes in the Frequency of Data Affect the Accuracy of Estimation of the Trend Parameter in a Jump Diffusion Process?

¿Los cambios en la frecuencia de los datos afectan la precisión de la estimación del parámetro de tendencia en un proceso de difusión con salto?

Carlos Armando Mejía Vega*

* Master in Finance. Universidad Externado de Colombia (Colombia). Researcher, Observatorio de Economía y Operaciones Numéricas (ODEON). Universidad Externado de Colombia (Colombia). PhD Student in Finance, Universidad Adolfo Ibáñez (Chile). [carlos.mejia@uexternado.edu.co], [ORCID ID: 0000-0002-0863-3342].
Artículo recibido: 2 de mayo de 2023
Aceptado: 15 de junio de 2023

Para citar este artículo:
Mejía Vega, C. A. (2023). Do Changes in the Frequency of Data Affect the Accuracy of Estimation of the Trend Parameter in a Jump Diffusion Process? *Odeon*, 24, 25-54.
DOI: https://doi.org/10.18601/17941113.n24.03

Abstract

This paper explores the effect of the frequency of data on the accuracy (measure by variance) of the maximum likelihood estimator (MLE) of the trend parameter μ in a jump-diffusion process à la Press (1967). First, we consider the case without jumps (i.e., the geometric Brownian motion (GBM)) as a benchmark case to show that the frequency of data is irrelevant in this first setting. Then, we consider the case with jumps and highlight that things are different in this second situation. Specifically, the asymptotic variance of the MLE of the trend parameter turns out to be higher compared to the case without jumps. Nevertheless, we also prove that when sampling occurs infinitely often (i.e., high frequency) it is possible to obtain the same accuracy for the MLE of μ as for the GBM, given that for higher frequencies it is easier to "identify" price discontinuities (i.e., jumps) for this model. Mathematical proofs are performed under the assumption that the MLE of μ is estimated given the other parameters, but numerical (Montecarlo) simulations indicate that this is also the case even when all parameters are estimated together.

Key words: Lévy process; Poisson process; Maximum likelihood; diffusion; jump-diffusion

JEL classification: C13, C63.

Resumen

Este artículo explora el efecto que tiene la frecuencia de los datos en la precisión (medida por la varianza) del estimador de máxima verosimilitud (MLE - maximum likelihood estimator) del parámetro de tendencia μ en un proceso de difusión con salto a la Press (1967). Para ello, consideramos primero el caso sin saltos (es decir, el movimiento Browniano geométrico o GBM - geometric Brownian motion) como el modelo referencia, con el que se evidencia que la frecuencia de los datos es irrelevante. Acto seguido, consideramos el caso con saltos, en donde enfatizamos que las cosas son diferentes. Específicamente, observamos que en este caso la varianza asintótica del MLE del parámetro de tendencia es más alto que cuando no había saltos. Sin embargo, también observamos que cuando la frecuencia ocurre lo suficientemente seguido (alta frecuencia), es posible obtener la misma precisión para el MLE de μ que cuando se tiene el GBM, dado que para frecuencias más altas es más fácil "identificar" discontinuidades (saltos) en el precio para este modelo. Las pruebas matemáticas se llevan a cabo bajo el supuesto de que el MLE de μ se estima dados los

demás parámetros, pero las simulaciones numéricas (Monte Carlo) demuestran que este es el caso también cuando todos los parámetros se estiman en conjunto.

Palabras clave: Procesos de Lévy; Proceso de Poisson; máxima verosimilitud; difusión; difusión con salto.

Clasificación JEL: C13, C63.

Introduction

The estimation of the trend parameter in the classical (geometric) jump-diffusion process proposed by Press (1967) is highly relevant in fields of finance such as forecasting, portfolio choice and portfolio testing. However, one problem arises and that is that discrete observations come at different frequencies (daily, weekly, bi-weekly, monthly, bi-monthly, annual, etc.). Therefore, this paper poses an essential question, which, despite its importance and apparent simplicity, to the best of our knowledge appears to have been missed in the literature: Do changes in the frequency of data (given the same window of observation) affect the accuracy to estimate the drift parameter (measure by its variance) in this stochastic process?

The question stated above has already been formulated in the case of the diffusion parameter (see Ait-Sahalia, 2004). This issue is understandable as the diffusion term is not only essential for portfolio allocation and performance, but also for risk management and option pricing. Furthermore, in the case of jump-diffusion processes, some of the insights suggest that the jump component can affect the estimation of the volatility parameter for different frequencies. However, as indicated before, the drift parameter is also relevant and, consequently, it is important to determine if the frequency of the data also affects the accuracy of its estimation.

Traditionally, parameters of classical stochastic processes like the GBM (i.e., Press, 1967) model without jumps) are estimated by using the maximum (log) likelihood method (Phillips & Yu, 2009). In this case, the method is straightforward since in a discrete time setting the transitional density of each logarithmic return is Gaussian, which enables closed form solutions for both the drift and diffusion estimators. From here, it is possible to estimate the variance of the estimators and check whether the frequency of the data has any influence on them. As we will see in this first case, sampling is irrelevant and does not improve the accuracy of the MLE of the drift parameter.

However, for jump diffusion processes like the one suggested by Press (1967), the direct application of the maximum (log)likelihood method is cumbersome. As noted by Press (1967), Beckers (1981), Ball & Torous (1985), Kiefer (1978), and Honoré (1998), the first order conditions obtained in this second case are highly non-linear, contain an infinite sum that should be truncated somehow, and require the imposition of limits to some parameters to bound the (log)likelihood function. All these requirements make it highly complex to obtain analytical results for the estimators compared with the GBM case[1]. In that sense, if the method is going to be used, it is generally applied using numerical techniques to choose the values that maximize the likelihood function[2]. On the other hand, analytical results can be obtained using other methods such as the cumulant matching method (Press, 1967 or Beckers, 1981 are the classical references), or the efficient method of moments (Chernov et al. 2003 is the classical reference) which, however, possess some problems, and do not necessarily present the efficient properties of maximum likelihood estimators.

Because of the difficulty in obtaining analytical estimators in the jump-diffusion case using the maximum likelihood method, and the need to use that method to compare the results with the GBM benchmark case, we will use a different approach to the one exposed above for the GBM for testing the influence of frequency. First, under the assumption that the other parameters are known, and given that for this type of stochastic process the logarithmic returns are independent and stationary, it is possible to define the Fisher information of the trend parameter for one observation (as the expected value of the square of the derivative of the log-likelihood function with respect to the trend parameter) and then scale it to obtain the Fisher information for the entire sample (Härdle & Simar, 2019). From that result and following some of the ideas of

1 For the GBM, there is one restriction that should be imposed, which is that the diffusion parameter should be strictly positive. However, the analytical result shows that this is not required as the MLE is obtained by using the population standard deviation of logarithmic returns which, by definition, is non-negative (see section 1). Instead, for the jump-diffusion process not only these kinds of requirements should be imposed on the diffusion term, the rate of the Poisson process, and the scale factor of the log-normal IID process, but also upper bounds for them as indicated by Honoré (1998). For Monte Carlo issues in section 3 we impose those requirements.

2 Nowadays, with programming languages such as R or Python, it is possible to implement complex optimization processes as the one that maximizes the joint (log)likelihood function of the jump-diffusion process.

Ait-Sahalia (2004), we can establish one upper and one lower bound that the Fisher information should fulfil in this case and conclude that the presence of jumps increases the variance of the estimator compared with the GBM case, but also that for high frequency data the effect is negligible. These results are also complemented by Monte Carlo simulations which show that the same happens when all the parameters are estimated together.

Despite filling a theoretical gap in the literature, our paper is also relevant for practitioners in finance. In fact, it is well known that logarithmic returns are non-Gaussian (Cont, 2001) so there is an increasing need to use other stochastic processes such as Lévy processes to take financial decisions. However, if the trend parameter is estimated with relatively low frequency data, our paper shows that it will be biased by the presence of jumps, affecting any financial decision taken using that information. In that regard, this article also motivates the need for financial agents to provide information with higher frequency, to guarantee for decisions to be taken with less biases.

To the best of our knowledge, the closest study in spirit to this paper is the doctoral thesis of Mai (2012) later published by Mai (2014) who studied the estimation of the speed of reversion parameter in the drift term of both a zero Ornstein-Uhlenbeck (OU) process and a Lévy-driven zero OU process with continuous-time data. He derived the asymptotic variance in both cases and concluded that the presence of jumps increases the variance of the estimator. Furthermore, he indicated that when we approximate the process from discrete observations it is simpler to detect large jumps and, therefore, for higher frequency data the additional variance of the jump vanishes asymptotically.

Despite the similarities with the approach of this paper, this article differs in some relevant respects. First, Mai (2012, 2014) obtained the variance directly from time-continuous data observations and later adapted it to the case of discrete observations, while in this study the variance of the estimator is obtained directly from discrete observations by using the method adapted from Ait-Sahalia (2004). This difference is important since, in practice, stochastic models should be estimated from discrete observations directly. Second, Mai (2012, 2014) focused on the zero Ornstein-Uhlenbeck (OU) process for general Lévy processes, both with finite and infinite activity, while we specialize on the GBM expanded by a specific Lévy process with finite activity (Poisson process). Third, the emphasis of Mai (2012, 2014) was on the speed of reversion of the process and not on the long-term expected value (given that he worked with a zero OU process) which somehow differs from our approach as we focus precisely on the expected value of logarithmic returns[3].

The rest of this paper is structured as follows; Section 1 introduces the stochastic setting and the benchmark case of the GBM; Section 2 depicts the case of the jump-diffusion process; and Section 3 presents some results using Monte Carlo simulations to show empirically both conclusions obtained in the previous sections, and then the Conclusion follows. Proofs of all results are collected in the "Appendix".

1. Benchmark Case: The Geometric Brownian Motion

Let us fix as a primitive an index set $\mathcal{T} \equiv [0, T]$, with $T > 0$ but $T < \infty$, such that $t \in \mathcal{T}$ can be interpreted as "time". Uncertainty will be modelled by considering as a primitive a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let there be a one-dimensional $\mathbb{F}$-adapted Wiener process denoted by $W \coloneqq \{W_t, \mathcal{F}_t; t \in \mathcal{T}\}$ that takes values on $\mathbb{R}$, with $\mathbb{F} \coloneqq \{\mathcal{F}_t; t \in \mathcal{T}\}$ being the *standard filtration,* which is the one generated by the stochastic process and augmented by all *null sets* of Ω, the subsets of Ω (events) of zero probability. By simplicity we will assume that $\mathcal{F}_0$ is almost trivial and that $\mathcal{F}_T = \mathcal{F}$.

Under the stochastic environment exposed above, let us consider an asset (the *stock)* whose spot price will be modelled as an $\mathbb{F}$-adapted process $S \coloneqq \{S_t, \mathcal{F}_t; t \in \mathcal{T}\}$ that follows a GBM:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{1}$$

with $\mu \in \mathbb{R}$ as the drift and $\sigma > 0$ as the diffusion parameters respectively.

By applying Itô's lemma, we can see that the natural logarithmic of the asset $X = Ln(S)$ follows an arithmetic Brownian motion given by:

$$dX_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t. \tag{2}$$

3 Gloter, Loukianova, and Mai (2016) extend Mai (2012, 2014) results for a Lévy-driven OU with non-zero expected value and a Lévy-driven Cox, Ingersoll, and Ross (CIR) process.

The explicit solution of equation (2) is obtained by integrating between s and t, where $0 \leq s < t \leq T$ to obtain:

$$X_t - X_s = \left(\mu - \frac{\sigma^2}{2}\right)(t-s) + \sigma(W_t - W_s), \ \forall \ 0 \leq s < t \leq T. \quad (3)$$

The logarithmic return $X_t - X_s$ has expected value, variance and Gaussian density given by:

$$E(X_t - X_s) = \left(\mu - \frac{\sigma^2}{2}\right)(t-s), \ \forall \ 0 \leq s < t \leq T. \quad (4)$$

$$Var(X_t - X_s) = \sigma^2(t-s), \ \forall \ 0 \leq s < t \leq T. \quad (5)$$

$$f(X_t - X_s; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2(t-s)}} e^{-\frac{1}{2}\left\{\frac{\left[X_t - X_s - \left(\mu - \frac{\sigma^2}{2}\right)(t-s)\right]^2}{\sigma^2(t-s)}\right\}}, \quad \forall \ 0 \leq s < t \leq T. \quad (6)$$

We proceed to discretize the process as follows. We will assume that the stock's prices are observed at times $0 = t_0, t_1, \dots, t_{\mathcal{N}} = T$, and we will approximate $t - s$ in equation (3) as a fix interval change in time given by $\Delta t = t_i - t_{i-1}$ that expresses the frequency of data, such that $\Delta t = \frac{t_{\mathcal{N}} - t_0}{\mathcal{N}} = \frac{T}{\mathcal{N}}$. Therefore, denoting $\Delta X_i = X_{t_i} - X_{t_{i-1}}$ and $\Delta W_i = W_{t_i} - W_{t_{i-1}}$ we can discretize equation (3) in the following way:

$$\Delta X_i = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma \Delta W_i, \ \forall \ i = 1,2, \dots, \mathcal{N}. \quad (7)$$

The expression ΔW_i which has an expected value $E(\Delta W_i) = 0$ and variance $Var(\Delta W_i) = t_i - t_{i-1} = \Delta t$ can be approximated by $\sqrt{\Delta t}\varepsilon_i$ where ε_i is an IID process that follows a Gaussian distribution with expected value $E(\varepsilon_i) = 0$ and variance $Var(\varepsilon_i) = 1$, such that:

$$E(\sigma \Delta W_i) = E\left(\sigma\sqrt{\Delta t}\varepsilon_i\right) = 0, \forall \ i = 1,2, \dots, \mathcal{N}, \quad (8)$$

$$Var(\sigma \Delta W_i) = Var\left(\sigma\sqrt{\Delta t}\varepsilon_i\right) = \sigma^2 \Delta t, \quad \forall \ i = 1,2, \dots, \mathcal{N}. \quad (9)$$

Therefore, equation (7) can be re-expressed as:

$$\Delta X_i = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}\varepsilon_i, \ \ \forall \ i = 1,2,\dots,\mathcal{N}. \tag{10}$$

The process $\{\Delta X_i\}_{i=1}^{\mathcal{N}}$ is an IID Gaussian process, such that any logarithmic return has an expected value, variance and density given by:

$$E(\Delta X_i) = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t, \ \ \forall \ i = 1,2,\dots,\mathcal{N}, \tag{11}$$

$$Var(\Delta X_i) = \sigma^2 \Delta t, \ \ \forall \ i = 1,2,\dots,\mathcal{N}, \tag{12}$$

$$f(\Delta X_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} e^{-\frac{1}{2}\left\{\frac{\left[\Delta X_i - \left(\mu - \frac{\sigma^2}{2}\right)\Delta t\right]^2}{\sigma^2 \Delta t}\right\}}, \forall \ i = 1,2,\dots,\mathcal{N}. \tag{13}$$

Notice that equations (11), (12), and (13) are consistent with equations (4), (5), and (6) respectively in a discretized framework.

The estimation using the maximum likelihood method has been studied extensively in the literature for this stochastic process using equation (13) (see for example Phillips and Yu, 2009 or Moreno Trujillo, 2011). In fact, departing from equation (13) it is well known that the joint log-likelihood function $\ell(\mu,\sigma)$ is given by:

$$\ell(\mu,\sigma) = -\frac{\mathcal{N}}{2} Ln(2\pi\Delta t) - \mathcal{N} Ln(\sigma) - \frac{1}{2\sigma^2\Delta t}\sum_{i=1}^{\mathcal{N}}\left[\Delta X_i - \left(\mu - \frac{\sigma^2}{2}\right)\Delta t\right]^2, \tag{14}$$

and that from equation (14), the MLE of μ and σ, denoted as $\hat{\mu}$ and $\hat{\sigma}$, are given by:

$$\hat{\mu} = \frac{\delta X}{\delta t} + \frac{\hat{\sigma}^2}{2}, \tag{15}$$

$$\hat{\sigma}^2 = \frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}\frac{(\Delta X_i)^2}{\Delta t} - \frac{1}{\mathcal{N}}\frac{(\delta X)^2}{\delta t}, \tag{16}$$

where $\delta X = \sum_{i=1}^{\mathcal{N}} \Delta X_i = X_{t_{\mathcal{N}}} - X_{t_0}$ and $\delta t = \sum_{i=1}^{\mathcal{N}} \Delta t = t_{\mathcal{N}} - t_0 = \mathcal{N}\Delta t$.

Or alternatively:

$$\left(\hat{\mu} - \frac{\hat{\sigma}^2}{2}\right)\Delta t = \frac{\sum_{i=1}^{\mathcal{N}} \Delta X_i}{\mathcal{N}} = \bar{r}, \tag{17}$$

$$\hat{\sigma}^2 \Delta t = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} (\Delta X_i - \bar{r})^2 = \sigma_r^2, \tag{18}$$

what implies that $\left(\hat{\mu} - \frac{\hat{\sigma}^2}{2}\right)\Delta t$ can be estimated as the mean of the logarithmic returns $\bar{r}$ and $\hat{\sigma}^2 \Delta t$ as the population variance of the logarithmic returns σ_r^2 as indicated by Moreno Trujillo (2011).

Even though it is possible to estimate the asymptotic variance of the estimator $\hat{\mu}$ from equations (15) or (17), under the assumption that $\hat{\sigma}$ is known ($\hat{\sigma} = \sigma$), by directly applying the variance operator, we will obtain it differently. Specifically, under the same assumption that $\hat{\sigma}$ is known, we use Fisher information of $\hat{\mu}$ for one observation and then scale it by $\mathcal{N}$ to obtain the Fisher information for the entire sample. Taking then the inverse function leads to the asymptotic variance. We will do it in this alternative way to use it in the jump-diffusion case later. The result leads to proposition 1.

Proposition 1. Under the assumption that $\hat{\sigma}$ is known ($\hat{\sigma} = \sigma$), Fisher information of $\hat{\mu}$ in the entire sample for the GBM, denoted as I_μ^{GBM}, is given by:

$$I_\mu^{GBM} = \frac{\delta t}{\sigma^2}. \tag{19}$$

Alternatively, the asymptotic variance of $\hat{\mu}$ for the GBM, denoted as $Var^{GBM}(\hat{\mu})$, is given by:

$$Var^{GBM}(\hat{\mu}) = \left(I_\mu^{GBM}\right)^{-1} = \frac{\sigma^2}{\delta t}. \tag{20}$$

Proof See “Appendix A” ■

Proposition 1 shows that the asymptotic variance of the MLE for the trend parameter does not include explicitly the frequency of data term *Δt*, and the distance between the first and final moments of time of the data (window of observation δt) only matter, which is a fact that is also consistent with the maximum likelihood estimation theory (see Appendix A). Therefore, we conclude that in this case the frequency of the data is irrelevant in terms of the accuracy of the estimation of the trend parameter (measured by its variance). This will be the benchmark used in the next section.

2. The Jump-Diffusion Process Case

For the jump-diffusion case, we will continue with the same stochastic setting but, additionally to the Wiener process, let us also consider a one-dimensional -adapted Poisson process with intensity $\lambda > 0$ denoted by $N \coloneqq \{N_t, \mathcal{F}_t; t \in \mathcal{T}\}$ that takes values on $\mathbb{R}_+$, also defined on the probability space, where now $\mathbb{F} \coloneqq \{\mathcal{F}_t; t \in \mathcal{T}\}$ is the one generated by both stochastic processes and augmented by all *null sets* of Ω. Again, for simplicity purposes we will assume that $\mathcal{F}_0$ is almost trivial and that $\mathcal{F}_T = \mathcal{F}$.

Under the stochastic environment presented above, let us consider that the spot price of the stock follows a jump-diffusion process driven by the Wiener process and the Poisson process à la Press (1967) in the following way[4]:

$$\frac{dS_t}{S_{t-}} = \mu dt + \sigma dW_t + (y_t - 1)dN_t, \tag{21}$$

with $\mu \in \mathbb{R}$ as the drift, and $\sigma > 0$ as the diffusion parameters. In contrast, y represents the jump intensity or amplitude of jumps, modelled as an IID process that follows a log-normal distribution, such that its logarithm $Y_t = Ln(y_t)$ follows a normal distribution with expected value $\beta \in \mathbb{R}$ and a variance $\eta > 0$.

By applying Itô's lemma to jump-diffusion processes, we can see that the natural logarithmic of the asset $X = Ln(S)$ follows an arithmetic Brownian motion plus a compound poisson process given by:

$$dX_t = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dW_t + Y_t dN_t. \tag{22}$$

The explicit solution of equation (22) is obtained by integrating between s and t, where $0 \leq s < t \leq T$ to obtain:

4 Given that in this case the stochastic process is right continuous with left limits (càdlàg) then the notation S_{t-} will denote $\lim_{s \uparrow t} S_s$, while S_t denotes $\lim_{s \downarrow t} S_s$.

$$X_t - X_s = \left(\mu - \frac{\sigma^2}{2}\right)(t-s) + \sigma(W_t - W_s) + \sum_{k=1}^{N_t - N_s} Y_k. \ \forall\, 0 \leq s < t \leq T. \quad (23)$$

Notice that by conditioning $N_t - N_s$ to a specific number of jumps n we have:

$$X_t - X_s = \left(\mu - \frac{\sigma^2}{2}\right)(t-s) + \sigma(W_t - W_s) + \sum_{k=1}^{n} Y_k, \qquad \forall\, 0 \leq s < t \leq T. \quad (24)$$

From equation (24) where $N_t - N_s = n$, we know that the increment or logarithmic return $X_t - X_s$ has expected value, variance and Gaussian density given by:

$$E(X_t - X_s | N_t - N_s = n) = \left(\mu - \frac{\sigma^2}{2}\right)(t-s) + n\beta, \ \forall \ 0 \leq s < t \leq T, \qquad (25)$$

$$Var(X_t - X_s | N_t - N_s = n) = \sigma^2(t-s) + n\eta, \ \forall \ 0 \leq s < t \leq T, \qquad (26)$$

$$f(X_t - X_s; \mu, \sigma, \beta, \eta | N_t - N_s = n) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2(t-s) + n\eta}} e^{-\frac{1}{2}\left\{\frac{\left[X_t - X_s - \left(\mu - \frac{\sigma^2}{2}\right)(t-s) - n\beta\right]^2}{\sigma^2(t-s) + n\eta}\right\}},$$

$$\forall \ 0 \leq s < t \leq T. \quad (27)$$

Now, considering that the probability that the Poisson random variable $N_t - N_s$ takes the value n is given by the Poisson distribution with parameter $\lambda(t-s)$:

$$\mathbb{P}[N_t - N_s = n] = \frac{[\lambda(t-s)]^n e^{-[\lambda(t-s)]}}{n!}, \qquad (28)$$

then, the probability density function of the independent and stationary increment or logarithmic return $X_t - X_s$ as a whole is given by an infinite sum (the Poisson random variable can have infinite discrete values) that starts at 0 (the Poisson random variable has a support that starts at zero, where each term is a multiplication of the probability of the random variable $X_t - X_s$ conditioned on a defined number of jumps , multiplied by the probability mass function that the Poisson random variable $N_t - N_s$ takes that number of jumps n (Bayes rule) in the following way:

$$f[X_t - X_s; \mu, \sigma, \lambda\, \beta, \eta]$$

$$= \sum_{n=0}^{\infty} \left\{ \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2(t-s)+n\eta}} e^{-\frac{1}{2}\left\{\frac{\left[X_t - X_s - \left(\mu - \frac{\sigma^2}{2}\right)(t-s) - n\beta\right]^2}{\sigma^2(t-s)+n\eta}\right\}} \frac{[\lambda(t-s)]^n e^{-[\lambda(t-s)]}}{n!} \right\}, \forall\ 0 \leq s < t \leq T. \quad (29)$$

where we can see that the distribution is a Gaussian mixture where mixture weights are given by the Poisson random variables (Honoré, 1998).

We proceed to discretize the process as in the benchmark case. Therefore, we can express equation (23) as:

$$\Delta X_i = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}\varepsilon_i + \sum_{k=1}^{N_{t_i} - N_{t_{i-1}}} Y_k, \quad \forall\ i = 1,2,\ldots,\mathcal{N}. \qquad (30)$$

Again, by conditioning $N_{t_i} - N_{t_{i-1}}$ to any number of jumps n we have:

$$\Delta X_i = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}\varepsilon_i + \sum_{k=1}^{n} Y_k, \quad \forall\ i = 1,2,\ldots,\mathcal{N}. \qquad (31)$$

From equation (31), where $N_{t_i} - N_{t_{i-1}} = n$, we can see that the process $\{\Delta X_i\}_{i=1}^{N}$ is an IID process, such that any logarithmic return has an expected value, variance, and Gaussian density given by:

$$E\left(\Delta X_i | N_{t_i} - N_{t_{i-1}} = n\right) = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \beta n, \ \ \forall\ \ i = 1,2,\ldots,\mathcal{N}, \qquad (32)$$

$$Var\left(\Delta X_i | N_{t_i} - N_{t_{i-1}} = n\right) = \sigma^2 \Delta t + \eta n, \ \ \forall\ \ i = 1,2,\ldots,\mathcal{N}, \qquad (33)$$

$$f\left(\Delta X_i; \mu, \sigma, \beta, \eta | N_{t_i} - N_{t_{i-1}} = n\right) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2\Delta t + n\eta}} e^{-\frac{1}{2}\left\{\frac{\left[\Delta X_i - \left(\mu - \frac{\sigma^2}{2}\right)\Delta t - n\beta\right]^2}{\sigma^2\Delta t + n\eta}\right\}}, \forall\ \ i = 1,2,\ldots,\mathcal{N}. \ (34)$$

Now, considering again that the probability that the Poisson random variable $N_{t_i} - N_{t_{i-1}}$ takes the value *n* is given by the Poisson distribution with parameter $\lambda(t_i - t_{i-1}) = \lambda\Delta t$:

$$\mathbb{P}\left[N_{t_i} - N_{t_{i-1}} = n\right] = \frac{(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{n!}, \tag{35}$$

then, the probability density function of the IID discretized logarithmic return as a whole is given by an infinite sum that starts at 0, where each term is a multiplication of the probability of the random variable ΔX_i, conditioned on a defined number of jumps *n*, multiplied by the probability mass function that the Poisson random variable $N_{t_i} - N_{t_{i-1}}$ takes that number of jumps *n* (Bayes rule):

$$f(\Delta X_i;\ \mu, \sigma, \lambda\ \beta, \eta) = \sum_{n=0}^{\infty} \left\{ \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2\Delta t + n\eta}} e^{-\frac{1}{2}\left\{\frac{\left[\Delta X_i - \left(\mu - \frac{\sigma^2}{2}\right)\Delta t - n\beta\right]^2}{\sigma^2\Delta t + n\eta}\right\}} \frac{(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{n!} \right\},$$

$$\forall\ i = 1,2, \dots, \mathcal{N}. \quad (36)$$

Equations (32), (33), (34), (35), and (36) are consistent with equations (25), (26), (27), (28), and (29) respectively in a discretized framework.

To obtain the asymptotic variance of the estimator $\hat{\mu}$ we might try to firstly employ the maximum likelihood method. Therefore, given the $\mathcal{N}$ observations, we proceed to formulate the joint likelihood function $\mathcal{L}(\ \mu, \sigma, \lambda\ \beta, \eta)$ based on equation (36):

$$\mathcal{L}(\ \mu, \sigma, \lambda\ \beta, \eta) = \prod_{i=1}^{\mathcal{N}} f(\mu, \sigma, \lambda\ \beta, \eta; \Delta X_i) \tag{37}$$

where $f(\mu, \sigma, \lambda\ \beta, \eta; \Delta X_i)$ represents the likelihood function for one observation.

We should proceed by maximizing the joint likelihood function to obtain estimators $\hat{\mu}\ \hat{\sigma},\ \hat{\lambda}, \hat{\beta}$ and $\hat{\eta}$. However, given that the logarithmic function is a monotonic function, we can re-express equation (37) first to obtain the joint log-likelihood function $\ell(\ \mu, \sigma, \lambda\ \beta, \eta)$:

$$\ell(\mu,\sigma,\lambda\,\beta,\eta) = Ln\left\{\prod_{i=1}^{\mathcal{N}} f(\mu,\sigma,\lambda\,\beta,\eta;\Delta X_i)\right\} = \sum_{i=1}^{\mathcal{N}} Ln[f(\mu,\sigma,\lambda\,\beta,\eta;\Delta X_i)] \qquad (38)$$

The expression $Ln[f(\mu,\sigma,\lambda\,\beta,\eta;\Delta X_i)]$ can be written as:

$$Ln[f(\mu,\sigma,\lambda\,\beta,\eta;\Delta X_i)]$$

$$= Ln\left\{\sum_{n=0}^{\infty}\left\{\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2\Delta t + n\eta}}\,e^{-\frac{1}{2}\left\{\frac{\left[\Delta X_i-\left(\mu-\frac{\sigma^2}{2}\right)\Delta t-n\beta\right]^2}{\sigma^2\Delta t+n\eta}\right\}}\frac{(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{n!}\right\}\right\},$$

$$\forall\ i = 1,2,\ldots,\mathcal{N}. \qquad (39)$$

Therefore, the joint log-likelihood function is given by:

$$\ell(\mu,\sigma,\lambda\,\beta,\eta)$$

$$= \sum_{i=1}^{\mathcal{N}} Ln\left\{\sum_{n=0}^{\infty}\left\{\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2\Delta t + n\eta}}\,e^{-\frac{1}{2}\left\{\frac{\left[\Delta X_i-\left(\mu-\frac{\sigma^2}{2}\right)\Delta t-n\beta\right]^2}{\sigma^2\Delta t+n\eta}\right\}}\frac{(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{n!}\right\}\right\}. \qquad (40)$$

Given equation (40) we should proceed, as in the case of the GBM, by differentiating it with respect to each parameter and then equating to zero to obtain the first order conditions. However, as was already indicated by Press (1967) and Beckers (1981), these first order conditions contain an infinite sum and are highly non-linear (given that the sum over the components appears inside the logarithmic function). In this respect, the issue of the infinite sum has been tackled in at least two ways. One way consists of truncating the summatory, an idea originally proposed by Ball and Torous (1985). The other consists of discretizing the process with other finite-support processes instead of the Poisson one. An example is the one which approximates the Poisson random variables with Bernoulli ones as in Ball and Torous (1983) and Honoré (1998).

In any case, the fact that the logarithmic of the process is a Gaussian mixture implies (from the mixture-of-distributions literature) that the joint likelihood function can be unbounded (reaching arbitrary high levels) which causes inconsistencies in the parameters. Therefore, it is necessary to impose certain conditions on them (see Kiefer, 1978 and Honoré, 1998 for further discussions) which combined with the non-linear nature of the first order conditions heavily increases the problem to obtain analytical solutions for the estimators. Therefore, most of the literature on the subject tends to maximize equation (40) not analytically but using numerical methods or techniques that truncate the summatory and impose conditions over the parameters.

Given that we want to obtain an analytical formula for the asymptotic variance of the MLE of the trend parameter, but it is not straightforward to obtain a closed-form formula for $\hat{\mu}$ directly as indicated above, we propose assuming again that the other parameters are known, and work with the Fisher information of $\hat{\mu}$ for one observation and then scale it by $\mathcal{N}$ to obtain the Fisher information for the entire sample. Taking then the inverse function leads to the asymptotic variance. However, given that in this case the log-likelihood function is much more complex, we will adapt the idea of Ait-Sahalia (2004) by imposing bounds that the information should fulfil to obtain two boundaries that asymptotically converge to the one of the GBM as we use higher frequencies of data. The result leads to proposition 2.

Proposition 2. Under the assumption that $\hat{\sigma}, \hat{\lambda}, \hat{\beta}, \hat{\eta}$ are known ($\hat{\sigma} = \sigma,\ \hat{\lambda} = \lambda$, $\hat{\beta} = \beta, \hat{\eta} = \ \eta$) Fisher information of $\hat{\mu}$ in the entire sample for the jump-diffusion à la Press (1967), denoted as I_{μ}^{JD}, when $\Delta t \to 0$ is given by:

$$\lim_{\Delta t \to 0}(I_{\mu}^{JD}) = \frac{\delta t}{\sigma^2} + o(\Delta t) = I_{\mu}^{GBM} \tag{41}$$

Alternatively, the asymptotic variance of $\hat{\mu}$, denoted as $Var^{JD}(\hat{\mu})$, when $\Delta t \to 0$ is given by:

$$\lim_{\Delta t \to 0}[Var^{JD}(\hat{\mu})] = \frac{\sigma^2}{\delta t} + o(\Delta t) = Var^{GBM}(\hat{\mu}) \tag{42}$$

Proof See "Appendix B" ■

Proposition 2 indicates that the presence of jumps increases the asymptotic variance of the drift parameter. However, it is possible to recover the efficiency as if there were no jumps when high frequency data (i.e., $\Delta t \to 0$) is used.

3. Montecarlo simulations

To prove the analytical results obtained in both sections 1 and 2 from a different angle, in this section we develop Monte Carlo simulations to examine the effect of frequency on the accuracy (i.e., variance) of the estimator $\hat{\mu}$ Following Ait-Sahalia et al (2005), we will consider that a daily frequency is high enough to show the effects presented in this paper (in particular of section 2) but without being affected by market microstructures.

3.1. Geometric Brownian Motion Case

For the GBM case, we simulate 5,000 sample paths using equation (10), using a fixed window of observation of 4 years $(\delta t = (t_N = T) - (t_0 = 0) = 4)$ with three different frequencies of data: daily $(\Delta t = 1/252)$, weekly $(\Delta t = 1/52)$, and bi-weekly $(\Delta t = 1/26)$[5]. The parameters used are the following: $\hat{\mu} = 0.1$ (annual expected return of 10%), and $\sigma = 0.2$ (annual volatility of 20%). Once the sample paths are obtained, the maximum likelihood method is applied to each one, such that we look for the values that maximize equation (14) in every case. The means, variances, and standard deviations related with $\hat{\mu}$ are presented in Table 1, where we cannot see a clear trend related with the frequency of the data and the variance of the estimator (variance values oscillate around 0.01 or the standard deviation values around 10%). We can conclude that the variance of the estimator is unaffected by the frequency of the data (given the same window of observation) and are always close to the expression in proposition 1 which in this case is equal to:

$$Var^{GBM}(\hat{\mu}) = \frac{(0.2)^2}{4} = 0{,}01 \tag{43}$$

$$Std^{GBM}(\hat{\mu}) = \sqrt{\frac{(0.2)^2}{4}} = 10\% \tag{44}$$

5 We use 4 years so that, given the $\sigma = 0.2$, even for the lower frequency case (bi-weekly) we would guarantee statistical significance.

Table 1. Results of 5,000 Monte Carlo simulations for estimator $\hat{\mu}$ with $\mu = 0.1$ and $\sigma = 0.2$ for the same window observation of $\delta t = 4$ years with different frequencies: $\Delta t = 1/26$ (bi-weekly), $\Delta t = 1/52$ (weekly), $\Delta t = 1/252$ (daily)

Frequency	Mean	Variance	Std
Bi-Weekly	10,0%	0,0105	10,2%
Weekly	10,0%	0,0105	10,3%
Daily	10,0%	0,0097	9,9%

3.2. Jump-Diffusion Case

For the jump-diffusion case, we also simulate 5,000 sample paths using equation (30), using a fixed window of observation of 4 years ($\delta t = (t_N = T) - (t_0 = 0) = 4$) with three different frequencies of data: daily ($\Delta t = 1/252$), weekly ($\Delta t = 1/52$) and bi-weekly ($\Delta t = 1/26$). The parameters used are the following: $\mu = 0.1$ (annual expected return of 10%), $\sigma = 0.2$ (annual volatility of 20%), $\lambda = 5$ (five jumps per year), $\beta = 0.05$ (expected return of jumps of 5%) and $\eta^{\frac{1}{2}} = 0.4$ (volatility of jumps 40%). Once the sample paths are obtained, the maximum likelihood method is applied to each one such that we look for the values that maximize equation (40). Following Ball and Torous (1985), we truncate the infinite sum to $n = 10$, and following Honoré (1998) we restrict both volatility terms and the frequency of jumps from 0.0001 to a maximum of 10. The estimators, variances, and standard deviations are presented in Table 2, in which we can see a clear trend where the variance of the estimator decreases and approaches to 0.01 (or the standard deviation decreases and approaches to 10%) as the frequency of the data increases, as indicated in proposition 2:

$$\lim_{\Delta t \to 0} [Var^{JD}(\hat{\mu})] = \frac{(0.2)^2}{4} + o(\Delta t) = Var^{GBM}(\hat{\mu}) = 0,01 \tag{45}$$

$$\lim_{\Delta t \to 0} [Std^{JD}(\hat{\mu})] = \sqrt{\frac{(0.2)^2}{4} + o(\Delta t)} = \sqrt{Var^{GBM}(\hat{\mu})} = 10\% \tag{46}$$

Additionally, the numbers obtained in Table 2 are close to the ones obtained by using the same parameters for equations (B32), (B33) and (B34) in the Appendix showing consistency of the theoretical part of section 2.

We can conclude that, given the same window of observation, higher frequencies of data improve the accuracy of the estimation, and when sampling occurs at very short intervals (in the daily example) it is possible to reach the same variance that the GBM case. As indicated by Ait-Sahalia (2004), this is due mainly to the fact that as Δt gets smaller, the ability to identify price discontinuities (i.e., jumps in the context of the model) improves. Specifically, given that β and η are not time-scaled and remain constant when Δt gets smaller (as opposed to μ, σ and λ which are time-scaled by the factor Δt), then for higher frequencies it is easier to determine what logarithmic returns are outliers that might be categorized as jumps. Therefore, given that the maximum likelihood method estimates the jump parameters λ, β and η with those price discontinuities or abnormal return outliers, while the others (namely μ and σ) with the rest of the information, the accuracy of the parameters improves if the data allows for a better discrimination of outliers. In the limit, when $\Delta t \rightarrow 0$, the information used to estimate the drift parameter has no outliers and this is the reason why its variance reaches the one of the GBM.

Frequency	Mean	Variance	Std
Bi-Weekly	10,3%	0,0148	12,2%
Weekly	10,3%	0,0131	11,5%
Daily	10,0%	0,0102	10,1%

Table 2. Results of 5,000 Monte Carlo simulations for estimator $\hat{\mu}$ with $\mu = 0.1$, $\sigma = 0.2$, $\lambda = 5$, $\beta = 0.05$ and $\eta = 0.16$ for the same window observation of $\delta t = 4$ years with different frequencies: $\Delta t = 1/26$ (bi-weekly), $\Delta t = 1/52$ (weekly), $\Delta t = 1/252$ (daily)

Conclusion

This paper shows that the frequency of data when estimating the drift term in the geometric jump-diffusion process à la Press (1967) is relevant. Nevertheless, we also prove that when frequency is sufficiently high, the accuracy improves to the same level of the case without jumps. This work is highly important for practitioners interested in fields such as forecasting or portfolio choice as it reveals the importance of calibrating their models with high frequency information. However, we are still far from a generalization of this property for geometric Lévy processes as the geometric jump-diffusion case is only one member of this family of stochastic processes. Therefore, for a generalization we should expand these ideas to other Lévy processes of finite activity as well as to the cases with infinite activity.

References

Aıt-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics, 74*(3), 487-528. https://doi.org/10.1016/j.jfineco.2004.01.002

Ait-Sahalia, Y., Mykland, P. A., & Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies, 18*(2), 351-416. https://doi.org/10.1093/rfs/hhi016

Ball, C. A., & Torous, W. N. (1983). A simplified jump process for common stock returns. *Journal of Financial and Quantitative Analysis, 18*(1), 53-65. https://doi.org/10.2307/2330710

Ball, C. A., & Torous, W. N. (1985). On jumps in common stock prices and their impact on call option pricing. *The Journal of Finance, 40*(1), 155-173. https://doi.org/10.1111/j.1540-6261.1985.tb04890.x

Beckers, S. (1981). A note on estimating the parameters of the diffusion-jump model of stock returns. *Journal of Financial and Quantitative Analysis, 16*(1), 127-140. https://doi.org/10.2307/2330614

Chernov, M., Gallant, A. R., Ghysels, E., & Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics, 116*(1-2), 225-257. https://doi.org/10.1016/S0304-4076(03)00151-4

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance, 1*(2), 223. https://doi.org/10.1080/713665670

Gloter, A., Loukianova, D. and Mai, H. (2016), Jump filtering and efficient drift estimation for Lévy-Driven SDE'S. Centre de Researche en Économie et Statisique. Série des Documents de Travail.

Härdle, W. K., & Simar, L. (2019). *Applied multivariate statistical analysis*. Springer Nature.

Honore, P. (1998). Pitfalls in estimating jump-diffusion models. Available at SSRN 61998.

Kiefer, N. M. (1978). Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica: Journal of the Econometric Society, 427-434*. https://doi.org/10.2307/1912371

Mai, H. (2012). Drift estimation for jump diffusions: time-continuous and high-frequency observations (Doctoral dissertation, Humboldt Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II).

Mai, H. (2014). Efficient maximum likelihood estimation for Lévy-driven Ornstein–Uhlenbeck processes. *Bernoulli, 20*(2), 919-957. https://doi.org/10.3150/13-BEJ537

Moreno Trujillo, J. F. (2011). Estimación de parámetros en ecuaciones diferenciales estocásticas aplicadas a finanzas. *ODEON,* (6), 131-144. https://doi.org/

Phillips, P. C., & Yu, J. (2009). Maximum likelihood and Gaussian estimation of continuous time models in finance. In *Handbook of Financial Time series* (pp. 497-530). Springer.

Press, S. J. (1967). A compound events model for security prices. *Journal of Business, 40*(3), 317-335. https://doi.org/10.1086/295117

Appendix A. Variance of for the GBM

Let us first obtain the expected value of $\hat{\mu}$, by applying the expected value operator to equation (15):

$$E(\hat{\mu}) = E\left(\frac{\delta X}{\delta t} + \frac{\hat{\sigma}^2}{2}\right), \qquad (A1)$$

from where we obtain:

$$E(\hat{\mu}) = \mu. \qquad (A2)$$

For the asymptotic variance of $\hat{\mu}$, we will use the Fisher information method as indicated by Härdle and Simar (2019). We depart from equation (10) for one single observation but working with the excess with respect to its expected value denoted as L_i:

$$L_i = \Delta X_i - \left(\mu - \frac{\sigma^2}{2}\right)\Delta t, \ \ \forall \quad i = 1,2,\dots,\mathcal{N}. \qquad (A3)$$

Notice that the process $\{L\}_{i=1}^{\mathcal{N}}$ is, by definition, IID such that each observation has an expected value, variance and Gaussian density given by:

$$E(L_i) = 0, \ \ \forall \quad i = 1,2,\dots,\mathcal{N}, \qquad (A4)$$

$$Var(L_i) = E\left(L_i^2\right) = \sigma^2 \Delta t, \ \ \forall \quad i = 1,2,\dots,\mathcal{N}, \qquad (A5)$$

$$f[L_i;\mu,\sigma] = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} e^{-\frac{1}{2}\left(\frac{{L_i}^2}{\sigma^2\Delta t}\right)}, \qquad \forall \quad i = 1,2,\dots,\mathcal{N}. \qquad (A6)$$

Therefore, the log-likelihood function for one observation is given by:

$$Ln[f(\mu,\sigma;L_i)] = -\frac{1}{2} Ln(2\pi\Delta t) - Ln(\sigma) - \frac{{L_i}^2}{2\sigma^2\Delta t}\ . \ \ \forall \quad i = 1,2,\dots,\mathcal{N}. \qquad (A7)$$

The derivate with respect to μ is given by:

$$\frac{\partial Ln[f(\mu,\sigma;L_i)]}{\partial \mu} = \frac{L_i}{\sigma^2}, \ \ \forall \quad i = 1,2,\dots,\mathcal{N}. \qquad (A8)$$

Therefore, we have:

$$\left\{\frac{\partial Ln[f(\mu,\sigma;L_i)]}{\partial\mu}\right\}^2=\frac{{L_i}^2}{\sigma^4},\ \ \forall\quad i=1,2,\dots,\mathcal{N}. \qquad (A9)$$

Then the Fisher information of the trend parameter of one observation for the GBM, denoted as $I_{\mu,\Delta t}^{GBM}$, is equivalent to:

$$I_{\mu,\Delta t}^{GBM}=E\left\{\left\{\frac{\partial Ln[f(\mu,\sigma;L_i)]}{\partial\mu}\right\}^2\right\}=E\left(\frac{{L_i}^2}{\sigma^4}\right)=\frac{1}{\sigma^4}E\left({L_i}^2\right)=\frac{\Delta t}{\sigma^2},\forall\quad i=1,2,\dots,\mathcal{N}. \quad (A10)$$

Alternatively, considering that:

$$\frac{\partial^2 Ln[f(\mu,\sigma;L_i)]}{\partial\mu^2}=-\frac{\Delta t}{\sigma^2},\forall\quad i=1,2,\dots,\mathcal{N}, \qquad (A11)$$

then we obtain the same result that equation (A10) by the properties of Fisher information:

$$I_{\mu,\Delta t}^{GBM}=-E\left\{\frac{\partial^2 Ln[f(\mu,\sigma;L_i)]}{\partial\mu^2}\right\}=-E\left(-\frac{\Delta t}{\sigma^2}\right)=\frac{\Delta t}{\sigma^2}. \qquad (A12)$$

Notice that the Fisher information for the entire sample I_{μ}^{GBM} is obtained by scaling $I_{\mu,\Delta t}^{GBM}$:

$$I_{\mu}^{GBM}=\mathcal{N}I_{\mu,\Delta t}^{GBM}=\frac{\mathcal{N}\Delta t}{\sigma^2}=\frac{\delta t}{\sigma^2}. \qquad (A13)$$

Then, the variance is the one given by equation (20) in proposition 1:

$$Var^{GBM}(\hat{\mu})=\left(I_{\mu}^{GBM}\right)^{-1}=\frac{\sigma^2}{\delta t}. \qquad (A14)$$

As indicated by the maximum likelihood method, $\hat{\mu}$ is a consistent indicator given that it converges in mean square (and hence in probability) to μ when the window of observation goes to infinity:

$$\lim_{\delta t\to\infty}[E(\hat{\mu})]=\mu,\qquad \lim_{\delta t\to\infty}[Var(\hat{\mu})]=0. \qquad (A15)$$

Appendix B. Variance of for the Jump-diffusion Case with GBM

For the asymptotic variance of $\hat{\mu}$, we will use the Fisher information method as indicated by Härdle and Simar (2019). We depart now from equation (30) for one single observation but working again with the excess with respect to its expected value when the number of jumps is equal to n, denoted again as L_i:

$$L_i = \Delta X_i - \left(\mu - \frac{\sigma^2}{2}\right)\Delta t - n\beta, \qquad \forall \quad i = 1,2, \dots, \mathcal{N}. \qquad (B1)$$

Notice that by conditioning the number of jumps to any number n, such that $N_{t_i} - N_{t_{i-1}} = n$, the process $\{L\}_{i=1}^{N}$ is, by definition, IID such that each observation has an expected value, variance and Gaussian density given by:

$$E\left(L_i | N_{t_i} - N_{t_{i-1}} = n\right) = 0, \qquad \forall \quad i = 1,2, \dots, \mathcal{N}, \qquad (B2)$$

$$Var\left(L_i | N_{t_i} - N_{t_{i-1}} = n\right) = \sigma^2 \Delta t + n\eta, \qquad \forall \quad i = 1,2, \dots, \mathcal{N}, \qquad (B3)$$

$$f\left(L_i;\ \mu, \sigma, \beta, \eta | N_{t_i} - N_{t_{i-1}} = n\right) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2(\Delta t) + n\eta}} e^{-\frac{1}{2}\left(\frac{L_i^2}{\sigma^2 \Delta t + n\eta}\right)},$$

$$\forall \quad i = 1,2, \dots, \mathcal{N}. (B4)$$

Now, considering that the probability that the Poisson random variable $N_{t_i} - N_{t_{i-1}}$ takes the value n is given by the Poisson distribution with parameter $\lambda(t_i - t_{i-1}) = \lambda \Delta t$:

$$\mathbb{P}\left[N_{t_i} - N_{t_{i-1}} = n\right] = \frac{(\lambda \Delta t)^n e^{-(\lambda \Delta t)}}{n!}, \qquad (B5)$$

then, the probability density function of the IID excess process as whole is given by:

$$f(L_i;\ \mu, \sigma, \lambda\, \beta, \eta) = \sum_{n=0}^{\infty} \left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2(\Delta t) + n\eta}} e^{-\frac{1}{2}\left(\frac{L_i^2}{\sigma^2 \Delta t + n\eta}\right)} \frac{(\lambda \Delta t)^n e^{-(\lambda \Delta t)}}{n!}\right], \forall \quad i = 1,2, \dots, \mathcal{N}. \quad (B6)$$

Under the assumption that the parameters $\sigma, \lambda\, \beta, \eta$ are known, then the Fisher information on a single observation for the excess process $I_{\mu,\Delta t}^{JD}$ is defined as:

$$I_{\mu,\Delta t}^{JD} = E\left\{\left\{\frac{\partial Ln[f(\mu,\sigma,\lambda\,\beta,\eta;L_i)]}{\partial\mu}\right\}^2\right\}, \qquad \forall \quad i = 1,2,\dots,\mathcal{N}. \qquad (B7)$$

Following Ait-Sahalia (2004) we have:

$$I_{\mu,\Delta t}^{JD} = \int_{-\infty}^{\infty}\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu}\frac{1}{f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}\right]^2 f(\mu,\sigma,\lambda\,\beta,\eta;L_i)dL_i,$$

$$\forall \quad i = 1,2,\dots,\mathcal{N}, \qquad (B8)$$

where $f(\mu,\sigma,\lambda\,\beta,\eta;L_i)$ is given by equation (B6), although with the terms inverted as it is the likelihood function for one observation, which can be also expressed as:

$$f(\mu,\sigma,\lambda\,\beta,\eta;L_i) = \sum_{n=0}^{\infty}\left\{f_n(\mu,\sigma,\lambda\,\beta,\eta;L_i)\frac{(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{n!}\right\}, \qquad \forall \quad i = 1,2,\dots,\mathcal{N}. \qquad (B9)$$

Before we proceed, notice that given the usual properties, the Fisher information for the entire sample I_{μ}^{JD} is obtained by scaling $I_{\mu,\Delta t}^{JD}$:

$$I_{\mu}^{JD} = \mathcal{N} I_{\mu,\Delta t}^{JD} = \mathcal{N}\left\{\int_{-\infty}^{\infty}\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu}\frac{1}{f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}\right]^2 f(\mu,\sigma,\lambda\,\beta,\eta;L_i)dL_i\right\}. \quad (B10)$$

Now, calculating the partial derivative $\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu}$ yields:

$$\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu} = \sum_{n=0}^{\infty}\left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2\Delta t + n\eta}}\frac{(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{n!}e^{-\frac{1}{2}\left(\frac{L_i^2}{\sigma^2\Delta t + n\eta}\right)}\left(\frac{L_i\Delta \mathrm{t}}{\sigma^2\Delta t + n\eta}\right)\right], \forall \quad i$$

$$= 1,2,\dots,\mathcal{N}. \quad (B11)$$

From here, adapting the idea of Ait-Sahalia (2004) we impose two bounds to the Fisher information.

For the upper bound, the presence of jumps cannot increase the available information about μ compared with the case without jumps. Fisher information for the excess in the GBM case is given by (Appendix A):

$$I_{\mu}^{GBM} = \mathcal{N} I_{\mu,\Delta t}^{GBM} = \mathcal{N} \frac{\Delta t}{\sigma^2} = \frac{\delta t}{\sigma^2}. \qquad (B12)$$

Therefore, the first (upper) boundary for the entire sample Fisher information in the case of jumps I_{μ}^{JD} is given by:

$$\frac{\delta t}{\sigma^2} \geq I_{\mu}^{JD}. \qquad (B13)$$

The lower bound is obtained by integrating equation (B8) on a restricted subset of the real line, $(-a_{\Delta t}, +a_{\Delta t})$, yielding from the positivity of the integrand as in Ait-Sahalia (2004), such that:

$$I_{\mu}^{JD} = \mathcal{N} \int_{-\infty}^{\infty} \frac{\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial \mu}\right]^2}{f(\mu,\sigma,\lambda\,\beta,\eta;L_i)} dL_i \geq \mathcal{N} \int_{-a_{\Delta t}}^{a_{\Delta t}} \frac{\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial \mu}\right]^2}{f(\mu,\sigma,\lambda\,\beta,\eta;L_i)} dL_i. \qquad (B14)$$

We proceed to obtain an expression for this second boundary. For that, we set $a_{\Delta t}$ to be the positive solution of:

$$f_0(\mu,\sigma,\lambda\,\beta,\eta;a_{\Delta t}) = f_1(\mu,\sigma,\lambda\,\beta,\eta;a_{\Delta t}). \qquad (B15)$$

Replacing we have:

$$\frac{1}{\sqrt{2\pi\sigma^2\Delta t}} e^{-\frac{1}{2}\left[\frac{(a_{\Delta t})^2}{\sigma^2\Delta t}\right]} = \frac{1}{\sqrt{2\pi(\sigma^2\Delta t+\eta)}} e^{-\frac{1}{2}\left[\frac{(a_{\Delta t})^2}{\sigma^2\Delta t+\eta}\right]}. \qquad (B16)$$

Solving we have:

$$a_{\Delta t} = (\Delta t)^{\frac{1}{2}} (-\eta)^{-\frac{1}{2}} \sigma [Ln(\sigma^2\Delta t) - Ln(\sigma^2\Delta t+\eta)]^{\frac{1}{2}} (\sigma^2\Delta t+\eta)^{\frac{1}{2}}. \qquad (B17)$$

Considering that:

$$[Ln(\sigma^2\Delta t) - Ln(\sigma^2\Delta t+\eta)]^{\frac{1}{2}} = \left[-Ln\left(1+\frac{\eta}{\sigma^2\Delta t}\right)\right]^{\frac{1}{2}}, \qquad (B18)$$

then:

$$a_{\Delta t} = (\Delta t)^{\frac{1}{2}}(-\eta)^{-\frac{1}{2}}\sigma\left[-Ln\left(1+\frac{\eta}{\sigma^2\Delta t}\right)\right]^{\frac{1}{2}}(\sigma^2\Delta t+\eta)^{\frac{1}{2}}. \qquad (B19)$$

Solving:

$$a_{\Delta t} = (\Delta t)^{\frac{1}{2}}(\eta+\sigma^2\Delta t)^{\frac{1}{2}}\eta^{-\frac{1}{2}}\sigma\left[Ln\left(1+\frac{\eta}{\sigma^2\Delta t}\right)\right]^{\frac{1}{2}}. \qquad (B20)$$

For all $y \in (-a_{\Delta t}, +a_{\Delta t})$ we have that:

$$f_0(\mu,\sigma,\lambda\,\beta,\eta;L_i) > f_1(\mu,\sigma,\lambda\,\beta,\eta;L_i) > \cdots > f_n(\mu,\sigma,\lambda\,\beta,\eta;L_i). \qquad (B21)$$

Following Ait-Sahalia (2004), given equation (B21) we have:

$$\frac{1}{f(\mu,\sigma,\lambda\,\beta,\eta;L_i)} \geq \frac{1}{f_0(\mu,\sigma,\lambda\,\beta,\eta;L_i)}. \qquad (B22)$$

Therefore:

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} \frac{\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu}\right]^2}{f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}\,dL_i \geq \mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} \frac{\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu}\right]^2}{f_0(\mu,\sigma,\lambda\,\beta,\eta;L_i)}\,dL_i \qquad (B23)$$

The term to the right (the low boundary) is given by:

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} \frac{\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu}\right]^2}{f_0(\mu,\sigma,\lambda\,\beta,\eta;L_i)}\,dL_i$$

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} \frac{\left\{\sum_{n=0}^{\infty}\left\{\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2\Delta t+n\eta}}\frac{(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{n!}\,e^{-\frac{1}{2}\left[\frac{(L_i)^2}{\sigma^2\Delta t+n\eta}\right]}\left[\frac{L_i\Delta t}{\sigma^2\Delta t+n\eta}\right]\right\}\right\}^2}{f_0(\mu,\sigma,\lambda\,\beta,\eta;L_i)}\,dL_i. \quad (B24)$$

Solving we have:

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} \frac{\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu}\right]^2}{f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}\,dL_i$$

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} \left\{\left\{\sum_{n=0}^{\infty}\left\{\frac{(2\pi)^{-\frac{1}{4}}(\sigma^2\Delta t)^{\frac{1}{4}}(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{[\sigma^2\Delta t+n\eta]^{\frac{3}{2}}n!}(L_i)\Delta t\,e^{-\frac{1}{2}\left[\frac{(L_i)^2}{\sigma^2\Delta t+n\eta}\right]+\frac{1}{4}\left[\frac{(L_i)^2}{\sigma^2\Delta t}\right]}\right\}\right\}^2\right\}dL_i\,. \quad (B25)$$

Denoting:

$$g_n(\mu,\sigma,\lambda\,\beta,\eta;L_i)=\frac{(2\pi)^{-\frac{1}{4}}(\sigma^2\Delta t)^{\frac{1}{4}}(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{[\sigma^2\Delta t+n\eta]^{\frac{3}{2}}n!}\;e^{-\frac{1}{2}\left[\frac{(L_i)^2}{\sigma^2\Delta t+n\eta}\right]+\frac{1}{4}\left[\frac{(L_i)^2}{\sigma^2\Delta t}\right]}L_i\Delta t,\quad (B26)$$

then:

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}}\frac{\left[\frac{\partial f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}{\partial\mu}\right]^2}{f(\mu,\sigma,\lambda\,\beta,\eta;L_i)}dL_i=\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}}\left\{\left[\sum_{n=0}^{\infty}g_n(\mu,\sigma,\lambda\,\beta,\eta;L_i)\right]^2\right\}dL_i.\quad (B27)$$

Following Ait-Sahalia (2004), we have that:

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}}\left\{\left[\sum_{n=0}^{\infty}g_n(\mu,\sigma,\lambda\,\beta,\eta;L_i)\right]^2\right\}dL_i$$

$$=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}}g_n(\mu,\sigma,\lambda\,\beta,\eta;L_i)g_m(\mu,\sigma,\lambda\,\beta,\eta;L_i)\,dL_i.\quad (B28)$$

We proceed to solve the integral of the right-hand side for any n and m. The expressions for n and for m are:

$$g_n\equiv g_n(\,\mu,\sigma,\lambda\,\beta,\eta;L_i)$$

$$=\frac{(2\pi)^{-\frac{1}{4}}(\sigma^2\Delta t)^{\frac{1}{4}}(\lambda\Delta t)^n e^{-(\lambda\Delta t)}}{[\sigma^2\Delta t+n\eta]^{\frac{3}{2}}n!}\;e^{-\frac{1}{2}\left[\frac{(L_i)^2}{\sigma^2\Delta t+n\eta}\right]+\frac{1}{4}\left[\frac{(L_i)^2}{\sigma^2\Delta t}\right]}L_i\Delta t,\qquad (B29)$$

$$g_m\equiv g_m(\mu,\sigma,\lambda\,\beta,\eta;L_i)$$

$$=\frac{(2\pi)^{-\frac{1}{4}}(\sigma^2\Delta t)^{\frac{1}{4}}(\lambda\Delta t)^m e^{-(\lambda\Delta t)}}{[\sigma^2\Delta t+m\eta]^{\frac{3}{2}}m!}\;e^{-\frac{1}{2}\left[\frac{(L_i)^2}{\sigma^2\Delta t+m\eta}\right]+\frac{1}{4}\left[\frac{(L_i)^2}{\sigma^2\Delta t}\right]}L_i\Delta t.\qquad (B30)$$

And so, the integral to be solved is given by:

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} g_n g_m \, dL_i$$

$$= \mathcal{N}\frac{(2\pi)^{-\frac{1}{2}}(\sigma^2\Delta t)^{\frac{1}{2}}(\lambda\Delta t)^{n+m}e^{-2(\lambda\Delta t)}}{[\sigma^2\Delta t + n\eta]^{\frac{3}{2}}[\sigma^2\Delta t + m\eta]^{\frac{3}{2}} n!\, m!}(\Delta t)^2 \int_{-a_\Delta}^{a_\Delta}\left\{L_i^2 e^{-\frac{L_i^2}{2}\frac{\sigma^4(\Delta t)^2 - nm\eta^2}{(\sigma^2\Delta t + n\eta)(\sigma^2\Delta t + m\eta)(\sigma^2\Delta t)}}\right\} dL_i. \quad (B31$$

Solving the integral, we obtain:

$$\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} g_n g_m \, dL_i$$

$$= \delta t \frac{(2\pi)^{-\frac{1}{2}}(\sigma^2\Delta t)^{\frac{1}{2}}(\lambda\Delta t)^{n+m}e^{-2(\lambda\Delta t)}}{[\sigma^2\Delta t + n\eta]^{\frac{3}{2}}[\sigma^2\Delta t + m\eta]^{\frac{3}{2}} n!\, m!}\Delta t\left[\frac{\sqrt{2\pi}\,\mathrm{erf}\left(\frac{a_{\Delta t}\sqrt{b}}{\sqrt{2}}\right)}{b^{\frac{3}{2}}}\right.$$

$$\left. - \frac{2a_\Delta e^{-\frac{(a_\Delta)^2 b}{2}}}{b}\right], \quad (B32)$$

where:

$$b = \frac{\sigma^4(\Delta t)^2 - nm\eta^2}{(\sigma^2\Delta t + n\eta)(\sigma^2\Delta t + m\eta)(\sigma^2\Delta t)}, \quad (B33)$$

$$a_\Delta = (\Delta \mathrm{t})^{\frac{1}{2}}(\eta + \Delta \mathrm{t}\sigma^2)^{\frac{1}{2}}\sigma\eta^{-\frac{1}{2}}\left[Ln\left(1 + \frac{\eta}{\Delta t\sigma^2}\right)\right]^{\frac{1}{2}}. \quad (B34)$$

When $\Delta t \to 0$, equations (B32), (B33) and (B34) produce a value different from zero only when $n = m = 0$[6]. Therefore:

$$\lim_{\Delta t \to 0}\left\{\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}} g_n g_m \, dL_i\right\} = \lim_{\Delta t \to 0}\left\{\mathcal{N}\int_{-a_\Delta}^{a_{\Delta t}}[g_0(\mu, \sigma, \lambda\, \beta, \eta; L_i)]^2 dL_i\right\} + O(\Delta t). \quad (B35)$$

Finally, notice that:

[6] This is done numerically by building the function, assigning a number to the parameters, and then looking that for extremely low values of Δt the only combination that gives a number different to zero is in fact the case $n = m = 0$.

$$\mathcal{N}\int_{-a_\Delta}^{a_{\Delta t}}(g_0)^2\,dL_i = \delta t\frac{(2\pi)^{-\frac{1}{2}}(\Delta t)^{\frac{1}{2}}e^{-2(\lambda\Delta t)}}{\sigma^5(\Delta t)^2}\left[\frac{\sqrt{2\pi}\,\mathrm{erf}\left(\frac{a_{\Delta t}\sqrt{b}}{\sqrt{2}}\right)}{b^{\frac{3}{2}}} - \frac{2a_\Delta e^{-\frac{(a_\Delta)^2 b}{2}}}{b}\right] \quad (B36)$$

where:

$$b = \frac{\sigma^4(\Delta t)^2}{(\sigma^2\Delta t)^3} = \frac{\sigma^4(\Delta t)^2}{\sigma^6(\Delta t)^3} = \frac{1}{\sigma^2\Delta t} \quad (B37)$$

$$a_\Delta = (\Delta \mathrm{t})^{\frac{1}{2}}(\eta + \Delta \mathrm{t}\sigma^2)\sigma\eta^{-\frac{1}{2}}\left[Ln\left(1+\frac{\eta}{\Delta t\sigma^2}\right)\right]^{\frac{1}{2}} \quad (B38)$$

Replacing b and a_Δ in equation (B36) we obtain:

$$\mathcal{N}\int_{-a_\Delta}^{a_{\Delta t}}(g_0)^2\,dL_i$$

$$= \frac{\delta t}{\sigma^2}e^{-2(\lambda\Delta t)}\left\{\mathrm{erf}\left\{\frac{(\eta+\Delta \mathrm{t}\sigma^2)^{\frac{1}{2}}\eta^{-\frac{1}{2}}\left[Ln\left(1+\frac{\eta}{\Delta t\sigma^2}\right)\right]^{\frac{1}{2}}}{\sqrt{2}}\right\}\right.$$

$$- 2(2\pi)^{-\frac{1}{2}}(\eta$$

$$\left. + \Delta \mathrm{t}\sigma^2)^{\frac{1}{2}}\eta^{-\frac{1}{2}}\left[Ln\left(1+\frac{\eta}{\Delta t\sigma^2}\right)\right]^{\frac{1}{2}} e^{-\frac{\left\{(\eta+\Delta \mathrm{t}\sigma^2)^{\frac{1}{2}}\eta^{-\frac{1}{2}}\left[Ln\left(1+\frac{\eta}{\Delta t\sigma^2}\right)\right]^{\frac{1}{2}}\right\}^2}{2}}\right\} \quad (B39)$$

And taking the limit when $\Delta t \to 0$ then:

$$\lim_{\Delta t\to 0}\left[e^{-2(\lambda\Delta t)}\right] = 1,$$

$$\lim_{\Delta t\to 0}\left\{\mathrm{erf}\left\{\frac{(\eta+\Delta \mathrm{t}\sigma^2)^{\frac{1}{2}}\eta^{-\frac{1}{2}}\left[Ln\left(1+\frac{n}{\Delta t\sigma^2}\right)\right]^{\frac{1}{2}}}{\sqrt{2}}\right\}\right\} = 1, \quad (B40)$$

$$\lim_{\Delta t\to 0}\left\{2(2\pi)^{-\frac{1}{2}}(\eta+\Delta t\sigma^2)^{\frac{1}{2}}\eta^{-\frac{1}{2}}\left[Ln\left(1+\frac{\eta}{\Delta t\sigma^2}\right)\right]^{\frac{1}{2}}e^{-\frac{\left\{(\eta+\Delta t\sigma^2)^{\frac{1}{2}}\eta^{-\frac{1}{2}}\left[Ln\left(1+\frac{\eta}{\Delta t\sigma^2}\right)\right]^{\frac{1}{2}}\right\}^2}{2}}\right\}=0. \quad (B41)$$

Therefore,

$$\lim_{\Delta t\to 0}\left\{\mathcal{N}\int_{-a_{\Delta}}^{a_{\Delta t}}[g_0(\mu,\sigma,\lambda\,\beta,\eta;L_i)]^2dL_i\right\}=\frac{\delta t}{\sigma^2}+o(\Delta t). \quad (B42)$$

So we finally have:

$$\lim_{\Delta t\to 0}\left(\mathcal{N}\int_{-a_{\Delta t}}^{a_{\Delta t}}g_n g_m\,dL_i\right)=\frac{\delta t}{\sigma^2}+o(\Delta t). \quad (B43)$$

Following Ait-Sahalia (2004), combining the upper and lower bounds in the limit we obtain:

$$\lim_{\Delta t\to 0}(\,I_{\mu}^{JD})=\frac{\delta t}{\sigma^2}+o(\Delta t)=I_{\mu}^{GBM} \quad (B44)$$

Or alternatively if we denote as $Var^{JD}(\hat{\mu})$ for the variance of $\hat{\mu}$ in the jump-diffusion case, we obtain equation (42) as in proposition 2:

$$\lim_{\Delta t\to 0}[\,Var^{JD}(\hat{\mu})]=\frac{\sigma^2}{\delta t}+o(\Delta t)=Var^{GBM}(\hat{\mu}) \quad (B45)$$