Maximum Entropy Evaluation of Asymptotic Hedging Error under a Generalised Jump-Diffusion Model

Farzad Alavi Fard
Firmin Doko Tchatoka
Sivagowry Sriananthakumar

Working Paper No. 2015-17

September 2015

Copyright the authors
Maximum Entropy Evaluation of Asymptotic Hedging Error under a Generalised Jump-Diffusion Model

Farzad Alavi Fard*   Firmin Doko Tchatoka†   Sivagowry Sriananthakumar‡

September 18, 2015

Abstract

In this paper we propose a maximum entropy estimator for the asymptotic distribution of the hedging error for options. Perfect replication of financial derivatives is not possible, due to market incompleteness and discrete-time hedging. We derive the asymptotic hedging error for options under a generalised jump-diffusion model with kernel biased, which nests a number of very important processes in finance. We then obtain an estimation for the distribution of hedging error by maximising Shannon’s entropy subject to a set of moment constraints, which in turn yield the value-at-risk and expected shortfall of the hedging error. The significance of this approach lies in the fact that the maximum entropy estimator allows us to obtain a consistent estimate of the asymptotic distribution of hedging error, despite the non-normality of the underlying distribution of returns.

Key Words: Generalised Jump; kernel biased; Asymptotic Hedging Error; Esscher Transform; Maximum Entropy Density; Value-at-Risk; Expected Shortfall

JEL: C13; C51; G13

1 Introduction

The theory of pricing and hedging options has been the centre of attention in modern mathematical finance since the seminal Black-Scholes model. It provides a theoretical value and hedging strategy for European options, under the key assumption that there exists a trading strategy which constructs a portfolio that perfectly replicates the pay-off of an option. Further, Black and Scholes assume the underlying stock price follows a geometric Brownian motion, and trading may take place in continuous time. With these assumptions, they show that the initial value of the replicating portfolio provides the initial price of the option. Moreover, the Black-Scholes analysis demonstrates that an option can be created synthetically by dynamically trading in

---

*RMIT University, Victoria, Australia. Email: farzad.alavifard@rmit.edu.au
†The University of Adelaide, South Australia, Australia. Email: firmin.dokotchatoka@adelaide.edu.au
‡RMIT University, Victoria, Australia. Email: sivagowry.sriananthakumar@rmit.edu.au
the underlying asset. Nevertheless, it is well accepted that the perfect replication of options by any self-financing strategy is impossible, due to market incompleteness as well as discrete-time hedging. These two sources of error are termed the jump error and the gamma error.

Many researchers, (see for example Jacod et al. (2000), Hubalek et al. (2006), and Cont et al. (2007)) have studied the problem of hedging an option in an incomplete market, particularly where stock prices may jump. It is well understood that, except in very special cases, martingales with respect to the filtration of a discontinuous processes cannot be represented in the form of a unique self-financing strategy, which leads to market incompleteness. At the jump time, both the model price of the option and the value of the hedging portfolio jump. The former is a non-linear function of the stock price but the latter is a linear function of the stock price. Therefore, the jump induces a discrepancy between the value of the option and its replicating portfolio, and thus leads to jump error.

Furthermore, these researchers generally assume that the hedging portfolio can be continuously rebalanced, which is only possible in the absence of transcription costs. In practice, this level of liquidity is not possible and market practitioners rebalance their hedging portfolio using discrete-time observations, just a few times per trading day. The discrete hedging of derivatives securities leads to the gamma error. This error is not easy to measure because the stochastic analysis techniques are not available in discrete time. Basing it on the seminal work of Bertsimas et al. (2000), Hayashi and Mykland (2005) developed a methodology to analyse the discrete hedging error in a continuous-time framework using an asymptotic approach. This methodology was further developed by Tankov and Voltchkova (2009) who investigate the gamma risk via establishing a limit theorem for the renormalised error when the discretisation step tends to zero. Additionally, Rosenbaum and Tankov (2014) discuss the optimality conditions of discretised hedging strategies in the presence of jump.

In this paper we contribute to the literature by approaching the problem from a different angle. We characterise the risk in dynamic hedge of options through the asymptotic distribution of hedging error. In particular, we investigate the case of conventional delta hedging for a European call option, although other types of options may be treated in a similar manner. Further, we obtain an estimation for the distribution of hedging error by maximising Shannon’s entropy (Shannon (1948)) subject to a set of moment constraints, which in turn yield the value-at-risk (VaR) and expected shortfall (ES) of the hedging error, two widely used risk metrics in finance. In the literature there exist two dominant approaches for constructing the distribution of hedging error, namely, the parametric and non-parametric approaches. The new approach that we propose in this paper chooses the probability distribution with the most uncertainty, or maximum entropy, subject to what is known. This allows us to obtain a consistent estimate of the asymptotic distribution of hedging error, despite the non-normality of the underlying distribution of returns. As a result, we can drive a very generalised modelling framework, which can be applied in different areas of derivatives pricing.

We first extend the methodology introduced in Hayashi and Mykland (2005) to model the asymptotic hedging error for vanilla call options when the underlying asset is governed by a generalised jump-diffusion model with kernel bias. The class of kernel biased completely random measures is a wide class of jump-type processes. It has a very nice representation, which is a generalised kernel biased mixture of Poisson random measures. The main idea of the kernel
biased completely random measure is to provide various forms of distortion of jump sizes of a completely random measure using the kernel function. This provides a great deal of flexibility in modelling different types of finite and infinite jump activities when compared with some existing models in the literature (for further discussion, see Fard and Siu (2013)).

Next, we estimate the probability density function of the hedging error using maximum entropy (ME) methodology. The principle of entropy maximisation is widely used in a variety of applications ranging from statistical thermodynamics, communications and engineering and quantum mechanics to decision and finance theories. The concept of entropy was first introduced in Shannon (1948) and various generalisations of Shannon entropy and other entropy measures exist in the literature, including Vasicek (1976) and Theil and Fiebig (1981). Some of these may be more convenient computationally in special cases, however, Shannon entropy is the best known and has a well-developed maximisation framework (Geman et al. (2014)).

The ME density is used in various areas of finance, for instance, modelling the distribution of financial returns (Chan (2009)), equities risk analysis (Chan (2009); Geman et al. (2014)), robust portfolio construction (Xu et al. (2014)), assessing financial contagion (Mistrulli (2011) and Zhou et al. (2013)). Geman et al. (2014), who use ME density in the Value-at-Risk (VaR) context, make an interesting observation that the real world is mostly ignorant about the importance of true probability distributions. They further point out that historically, finance theory has had a preference for parametric, less robust, methods. An approach that is based on distributional and parametric certainties may be useful for research purposes but does not accommodate responsible risk taking. Their study shows the importance of the use of true probability distributions in VaR calculations.

The remainder of this paper is structured as follows. In Section 2 we present the calculations for pricing European options under a generalised jump-diffusion model with kernel-bias. Further, we generalise the Hayashi and Mykland (2005) framework to derive the asymptotic hedging error, stemmed from the market incompleteness and discrete hedging. In Section 3, we obtain an estimation for the distribution of hedging error by maximising Shannon’s entropy subject to a set of moment constraints, which in turn yield the value-at-risk and expected shortfall of the hedging error. Section 4 provides a numerical analysis to highlight the applicability of the method. Section 5 concludes the paper.

2 Modelling Framework

2.1 Preliminaries

We fix a complete probability space \((\Gamma, \mathcal{F}, \mathbb{P})\), where \(\mathbb{P}\) is the real-world probability measure. Let \(\mathcal{F}\) denote the time index set \([0, T]\) of the economy. Let \(\{r(t)\}_{t \in \mathcal{F}}\) be the instantaneous market interest rate of a money market account. Then, the dynamics of the value of the risk-free asset, \(\{B_t\}_{t \in \mathcal{F}}\) would be

\[
\frac{dB_t}{B_t} = r(t)dt, \quad B_0 = 1.
\]
James (2002, 2005) propose a kernel biased representation of completely random measures, which provide a great deal of flexibility in modelling different types of finite and infinite jump activities by choosing different kernel functions. The approach is an amplification of Bayesian techniques developed by Lo and Weng (1989) for the gamma-Dirichlet processes. Perman et al. (1992) consider applications to the models, which all fall within an inhomogeneous spatial extension of the size biased framework. In this sequel, we adopt the kernel biased representation of completely random measures proposed by James (2002, 2005).

Let \((T, \mathcal{B}(T))\) denote a measurable space, where \(\mathcal{B}(T)\) is the Borel \(\sigma\)-field generated by the open subsets of \(T\). Write \(\mathcal{B}_0\) for the family of Borel sets \(U \in \mathcal{B}^+\), whose closure \(\bar{U}\) does not contain the point 0. Let \(X\) denote \(T \times \mathcal{B}^+\). The measurable space \((X, \mathcal{B}(X))\) is then given by \((T \times \mathcal{B}^+, \mathcal{B}(T) \otimes \mathcal{B}_0)\).

For each \(U \in \mathcal{B}_0\), let \(N(.,U)\) denote a Poisson random measure. Write \(N(dt,dz)\) for the differential form of measure \(N(t,U)\). Let \(\rho(dz|t)\) denote a Lévy measure, depending on \(t\); \(\eta\) is a \(\sigma\)-finite (nonatomic) measure on \(T\). As in James (2005), the existence of the kernel biased completely random measure is ensured by supposing an arbitrary positive function on \(\mathcal{B}^+\), \(h(z)\), \(\rho\) and \(\eta\) are selected in such a way that for each bounded set \(B\) in \(T\)

\[
\sum_{i=1}^{N} \int_{\mathcal{B}_0} \int_{\mathcal{B}^+} \min(h(z),1)\rho(dz|t)\eta(dt) < \infty.
\]

and \(h^2(z) \leq F_t\rho(z)\), where \(F\) is a càdlàg \(\mathcal{F}_t\)-adapted process. Define the intensity measure \(\nu(dt,dz) := \rho(dz|t)\eta(dt)\), as well as the kernel biased completely random measure

\[
\mu(dt) := \int_{\mathcal{B}^+} h(z)N(dt,dz).
\]

The latter is a kernel biased Poisson random measure \(N(dt,dz)\) over the state space of the jump size \(\mathcal{B}^+\) with the mixing kernel function \(h(z)\). We can replace the Poisson random measure with any random measure and choose some quite exotic functions for \(h(z)\) to generate different types of finite and infinite jump activities. Let \(\{W_t\}_{t \in \mathcal{T}}\) denote a standard Brownian motion on \((\Omega, \mathcal{F}, \mathbb{P})\) with respect to the \(\mathbb{P}\)-augmentation of its natural filtration \(\mathcal{F}^W := \{\mathcal{F}^W_t\}_{t \in \mathcal{T}}\). Let \(\tilde{N}_{X_t}(dt,dz)\) denote the compensated Poisson random measure defined by

\[
\tilde{N}(dt,dz) = N(dt,dz) - \rho(dz|t)\eta(dt).
\]

Let \(\mu_t\) and \(\sigma_t\) denote the drift and volatility of the market value of the underlying asset, respectively. Consider a random jump process \(A := \{A(t)|t \in \mathcal{T}\}\), such that

\[
dA_t = A_t\left[\mu_t dt + \sigma_t dW_t + \int_{\mathcal{B}^+} h(z)\tilde{N}(dt,dz)\right],
\]

where \(A_0 = 0\). We assume under \(\mathbb{P}\) the price process \(\{S_t\}_{t \in \mathcal{T}}\) is defined as \(S_t := \exp(A_t)\) so that
\[ dS_t = \left( \mu_t + \frac{1}{2} \sigma^2_t \right) dt + \sigma_t dW_t - \int_{\mathbb{R}^+} \left\{ h(z) - e^{h(z) + 1} \right\} \rho d(z|t) \eta(dt) + \int_{\mathbb{R}^+} \left( e^{h(z)} - 1 \right) \tilde{N}(dt, dz), \]

with \( S_0 = 1. \)

### 2.2 Pricing by the Esscher Transform

For the fair valuation of the option we need to ensure there are no arbitrage opportunities in the market through the determination of the equivalent risk neutral martingale measure (Pilska (1997)). In incomplete markets, as in this paper, there are more than one equivalent martingale measure, and hence, more than one no-arbitrage price. Different approaches have been proposed for pricing and hedging derivative securities in incomplete financial markets. For instance, Follmer and Sondermann (1986), Schweizer (1995), and Follmer and Schweizer (1991) select an equivalent martingale measure by minimising the quadratic utility of the terminal hedging errors. Davis (1997) adopts an economic approach based on the marginal rate of substitution to pick a pricing measure via a utility maximisation problem. Avellaneda (1998), Frittelli (2000), and Fard and Siu (2012) employ the minimum entropy martingale measure method to choose the equivalent martingale measure.

In this paper we employ the Esscher transform to determine an equivalent martingale measure for the valuation of the option. Gerber and Shiu (1994) pioneered the use of the Esscher transform, a popular tool in actuarial science. The Esscher transform provides market practitioners with a convenient and flexible way to value options. It has been shown in Elliott et al. (2005) that for exponential Lévy models the Esscher martingale transform for the linear process is also the minimal entropy martingale measure, i.e., the equivalent martingale measure which minimises the relative entropy, and that this measure also has the property of preserving the Lévy structure of the model. In the framework of exponential Lévy models the study of equivalent martingale measures, their relationships, and their optimality properties, has been developed in several directions, see Esche and Schweizer (2005), Hubalek and Sgarra (2006), and Tankov (2003) and the references therein.

Let \( \mathcal{F}^A := \{ \mathcal{F}^A_t \}_{t \in \mathcal{T}} \) and \( \mathcal{F}^S := \{ \mathcal{F}^S_t \}_{t \in \mathcal{T}} \) denote the \( \mathbb{P} \)-augmentation of the natural filtration generated by \( A \) and \( S \), respectively. Since, \( \mathcal{F}^A \) and \( \mathcal{F}^S \) are equivalent, we can use either one as an observed information structure. Write \( B(\mathcal{T}) \) for the Borel \( \sigma \)-field of \( \mathcal{T} \) and let \( BM(\mathcal{T}) \) denote the collection of \( B(\mathcal{T}) \)-measurable and nonnegative functions with compact support on \( \mathcal{T} \). For each process \( \theta \in BM(\mathcal{T}) \), write

\[ (\theta.A)_t := \int_0^t \theta(u)dA(u), \quad t \in \mathcal{T}, \]

such that \( \theta \) is integrable with respect to the return process.

Let \( \{ \Lambda_t \}_{t \in \mathcal{T}} \) denote a \( \mathcal{G} \)-adapted stochastic process.
\[ \Lambda_t := \frac{e^{(\theta A)_t}}{\mathcal{M}(\theta)_t}, \quad t \in \mathcal{T}, \]

where \[ \mathcal{M}(\theta)_t := E[e^{(\theta A)_t}|\mathcal{F}_t^A] \] is a Laplace cumulant process and takes the following form

\[ \mathcal{M}(\theta)_t = \exp \left[ \int_0^t \theta_s \left( \mu_s - \frac{1}{2} \sigma_s^2 \right) ds + \frac{1}{2} \int_0^t \theta_s^2 \sigma_s^2 ds + \int_0^t \int_{\mathcal{A}^+} \left( e^{\theta_s h(z)} - 1 - \theta h(z) \right) \rho(dz|s) \eta(ds) \right]. \]

Therefore

\[ \Lambda_t = \exp \left[ \int_0^t \theta_s \sigma_s dW_s - \frac{1}{2} \int_0^t \theta_s^2 \sigma_s^2 ds + \int_0^t \int_{\mathcal{A}^+} \theta_s h(z) \tilde{N}(dz, ds) \quad (2.3) \]

\[ - \int_0^t \int_{\mathcal{A}^+} \left( e^{\theta_s h(z)} - 1 + \theta_s h(z) \right) \rho(dz|s) \eta(ds) \right]. \]

Equation (2.3) is an essential part of our pricing formulation, since we aim to use \( \Lambda_t \) as the Radon-Nikodym derivative to change the measure from the historical measure to the risk-neutral measure. One key characteristic of risk-neutral measure is that under this measure, every discounted price process is a martingale. So it is also essential to demonstrate that (2.3) is \( \mathcal{G}_t \)-martingale.

**Lemma 1.** \( \Lambda_t \) is \( \mathbb{P} \) martingale w.r.t \( \mathcal{G}_t \).

**Proof.** James (2002, 2005) shows that

\[ E \left[ \exp \left( \int_0^t \int_{\mathcal{A}^+} \theta_s h(z) \tilde{N}(ds, ds) \right) \right] |\mathcal{G}_t] = \exp \left( \int_0^t \int_{\mathcal{A}^+} \left( e^{\theta_s h(z)} - 1 + \theta_s h(z) \right) \rho(dz|s) \eta(ds) \right). \]

Then, by taking the conditional expectations of (2.3), the results follow. \( \square \)

For each \( \theta \in L(A) \) define a new probability measure \( \mathbb{P}^\theta \sim \mathbb{P} \) on \( \mathcal{G}(T) \) by the Radon-Nikodym derivative

\[ \left. \frac{d\mathbb{P}^\theta}{d\mathbb{P}} \right|_{\mathcal{G}(T)} := \Lambda_T. \quad (2.4) \]

This new measure \( d\mathbb{P}^\theta \) is defined by the Esscher transform \( \Lambda_T \) associated with \( \theta \in L(A) \).

According to the fundamental theorem of asset pricing, the absence of arbitrage means there exists an equivalent martingale measure under which discounted asset prices are local-martingales, which is widely known as the *local-martingale condition*. Now we stipulate a necessary and sufficient condition for the local martingale condition.
Proposition 1. For each \( t \in \mathcal{T} \), let the discounted price of the risky asset at time \( t \) be

\[
\tilde{S}(t) := e^{-rt}S(t).
\]

Then the discounted price process \( \tilde{S} := \{\tilde{S}(t) | t \in \mathcal{T}\} \) is an \( \mathbb{P}^\theta \)-local-martingale if and only if \( \theta_t := \langle \theta, X_t \rangle \), \( t \in \mathcal{T} \), is such that \( \theta := (\theta_1, \theta_2, ..., \theta_N) \in \mathbb{R}^N \) satisfies the following equation

\[
\theta_t \sigma_t^2 + \int_{\mathbb{R}^+} \left\{ e^{\theta h(z)}(e^{h(z)} - 1) - h(z) \right\} \rho(dz|t) \eta'(t) = r_t - \mu_t.
\] (2.5)

Proof. See Appendix \( \square \)

The results from the Lemma 1, Equation (2.4), and Proposition 1, allow us to use (2.3) to drive the risk-neutral dynamics of the return process.

Proposition 2. Suppose \( \tilde{W}_t = W_t - \int_0^t \sigma_s \theta ds \) is a \( \mathbb{P}^\theta \)-Brownian motion, \( \rho^\theta(dz|t) := e^{\theta h(z)} \rho(dz|t) \) is the \( \mathbb{P}^\theta \) compensator of \( N^\theta(dt,dz) \) then

\[
dA_t = (\mu_t + 2\theta \sigma^2_s - \frac{1}{2} \sigma^2_s)dt + \sigma_t \tilde{W}_t + \int_{\mathbb{R}^+} h(z)(1 - e^{-\theta h(z)}) \rho^\theta(dz|t) \eta(dt) + \int_{\mathbb{R}^+} h(z) \tilde{N}^\theta(dz,dt).
\] (2.6)

Proof. See Appendix \( \square \)

Similarly, we can derive the risk-neutral price process of the reference portfolio.

Proposition 3. The price process of the reference portfolio \( S \) under \( \mathbb{P}^\theta \) is

\[
dS_t = (r_t - \frac{1}{2} \sigma^2_t)dt + \sigma_t \tilde{W}_t + \int_{\mathbb{R}^+} \left( e^{\theta h(z)}(e^{h(z)} - 1) - h(z) \right) \tilde{N}^\theta(dt,dz)
\] (2.7)

Proof. Recall \( S_t := \exp(A_t) \). Then the proof can easily follow by applying Ito’s Lemma and the martingale condition (2.5) to (2.6). \( \square \)

We study the hedging of a European option with pay-off function \( G \) using the popular delta hedging strategy. The option price is given by

\[
C(t, S) = E^\theta[G(S_T)|S_t = S],
\] (2.8)

where we assume \( C \in C^\infty([0, T] \times \mathcal{X}) \). Further, the delta hedging strategy is \( H_t := \frac{\partial C(t,S)}{\partial S} \), which is the most widely used hedging strategy with a mathematically tractable structure. Detailed discussion about the hedging strategy is provided in the next subsection.
2.3 Continuous Hedging Strategy

We suppose that there exists a continuous-time trading strategy $H$ which is the strategy that the agent would follow if continuous-time hedging was possible. In incomplete markets, this strategy need not lead to perfect replication, and can be chosen in many different ways. Here we do not discuss the relative advantages of different choices of $H$ but simply suppose that it is given by another generalised jump-diffusion process satisfying the same assumptions as $S$. Therefore, under $\mathbb{P}$ we suppose

$$dH_t = a_t dt + b_t dW_t + \int_{\mathbb{R}^+} \gamma_t \tilde{N}(dt, dz), \quad (2.9)$$

where $a_t$, $b_t$, and $\gamma_t$ are the parameters of the process that will be determined below.

By applying the Itô Lemma to the definition of $H_t$, we can show the following decomposition

$$dH_t := \frac{d\partial^2 C(t,S)}{dS} = \left\{ \frac{\partial^2 C}{\partial t \partial S}(t,S) + \left( \mu_t + \frac{1}{2} \sigma_t^2 + \int_{\mathbb{R}^+} h(z) \rho(dz|t)\eta \right) \frac{\partial^3 C}{\partial S^3}(t,S) + \frac{1}{2} \sigma_t^2 \frac{\partial^3 C}{\partial S^3}(t,S) \right\} dt$$

$$+ \sigma \frac{\partial^2 C}{\partial S^2}(t,S) dW_t + \int_{\mathbb{R}^+} \left\{ \frac{\partial C}{\partial S}(t, S + e^{h(z)} - 1) - \frac{\partial C}{\partial S}(t, S) \right\} \rho(dz|t)\eta' dt$$

$$+ \int_{\mathbb{R}^+} \left\{ \frac{\partial C}{\partial S}(t, S + e^{h(z)} - 1) - \frac{\partial C}{\partial S}(t, S) \right\} \tilde{N}(dt, dz).$$

Then

$$a_t = \frac{\partial^2 C}{\partial t \partial S}(t,S) + \left( \mu_t + \frac{1}{2} \sigma_t^2 + \frac{1}{2} \sigma_t^2 \frac{\partial^3 C}{\partial S^3}(t,S) \right) \frac{\partial^3 C}{\partial S^3}(t,S)$$

$$+ \int_{\mathbb{R}^+} \left\{ \frac{\partial C}{\partial S}(t, S + e^{h(z)} - 1) - \frac{\partial C}{\partial S}(t, S) + h(z) \frac{\partial^2 C}{\partial S^2}(t, S) \right\} \rho(dz|t)\eta' dt$$

$$b_t = \sigma \frac{\partial^2 C}{\partial S^2}(t,S),$$

$$\gamma_t = \frac{\partial C}{\partial S}(t, S + e^{h(z)} - 1) - \frac{\partial C}{\partial S}(t, S).$$

2.4 Asymptotic Hedging Error

In what follows we drive the asymptotic distribution of the hedging error, generalising the methodology proposed in Hayashi and Mykland (2005) and Tankov and Voltchkova (2009).

Let

$$\mu^n_t := (-n) \vee \mu_t \wedge n; \quad \sigma^n_t := (-n) \vee \sigma_t \wedge n; \quad h^n(z) := -\sqrt{n\rho(n)} \vee h(z) \wedge \sqrt{n\rho(n)};$$

$$a^n_t := (-n) \vee a_t \wedge n; \quad b^n_t := (-n) \vee b_t \wedge n; \quad \gamma^n_t(z) := -\sqrt{n\rho(n)} \vee \gamma_t(z) \wedge \sqrt{n\rho(n)).$$

Then the processes $S^n$ and $H^n$ are Lévy-Itô processes with bounded coefficients and bounded jumps that coincide with $S$ and $H$ on the set.
Further, define the renormalised hedging error process by

\[ \Omega_n := \{ \sup_{0 \leq t \leq T} \max(|\mu_t|, |\sigma_t|, |F_t|) \leq n; \sup_{0 \leq t \leq T} \max(|a_t|, |b_t|, |F_t|) \leq n; \mathcal{N}([0, T], ((-\infty, -n) \cup (n, \infty)) = 0) \}. \]

Since all processes are supported càdlàg, \( P[\Omega_n] \to 1 \).

The continuous re-balancing of a portfolio is practically unfeasible. Typically, holders of a position in an option, \( \Delta \)-hedge in discrete time intervals of \( t_i = iT/n \). Therefore, the trading strategy is piecewise constant and given by \( F_{\phi_n(t)} \), where \( \phi_n(t) = \sup\{t_i, t_i < t\} \). The value of the hedging portfolio at time \( t \) is \( V_0 + \int_0^t H_{s^-}dS_s \) with continuous hedging and \( V_0 + \int_0^t H_{\phi_n(t)}dS_s \) with discrete hedging. Then the asymptotic distribution of the difference between discrete and continuous hedging is

\[ U^n_t = \int_0^t (H_{s^-} - H_{\phi_n(t)})dS_s = \int_0^t H_{s^-}dS_s \]  \hfill (2.10)

where \( n \to \infty \). For any process \( A \) we set \( A^n_t := A_t - A_{\phi_n(t)} \). Under the above conditions, Hayashi and Mykland (2005) provides a thorough discussion on the stable convergence of the bounded processes to their respective original processes.

Further, define the renormalised hedging error process by

\[ Z^n_t = \sqrt{n}U^n_t = \sqrt{n} \int_0^t H_{s^-}dS_s. \]  \hfill (2.11)

Let \( \{\hat{W}_t\}_{t \in \mathcal{F}} \) be a standard Brownian motion independent of \( W \) and \( N \), and let \( (\xi_k)_{k \geq 1} \) and \( (\xi'_{k})_{k \geq 1} \) be two sequences of standard normal random variable and \( (\zeta_k)_{k \geq 1} \) sequence of independent uniform random variables on \([0,1]\), such that the three sequences are independent from each other and other random elements. Then, applying the Theorem 1 in Tankov and Voltchkova (2009), the renormalised discrete delta hedging error asymptotically converges stably in finite-dimensional laws to

\[ Z_t = \frac{\sqrt{T}}{2} \int_0^t \sigma_s^2 \frac{\partial^2 C}{\partial S^2}d\hat{W}_s + \sqrt{T} \sum_{s \leq t; \Delta S_t \neq 0} \left( \frac{\partial C}{\partial S}(s, S_s) - \frac{\partial C}{\partial S}(s, S_{s^-}) \right) \sqrt{\zeta_t} \xi_s \sigma_s \]  \hfill (2.12)

where \( T_i \) is an estimation of the jump time of \( N \).

### 3 Estimation of the Density of the Hedging Error

Suppose that we have a random sample of \( n \) i.i.d. observations on the hedging error \( Z \), \( \{Z_t: t = 1, \ldots, n\} \), each with pdf \( f_Z \) and cdf \( F_Z \) on a support \( \Omega \subseteq \mathbb{R} \). Value-at-risk is a popular risk metric, which for level \( \alpha \) associated to \( Z \) is defined as

\[ \text{VaR}_\alpha(Z) = \inf \{z \in \mathbb{R} : F_Z(z) \geq \alpha\} = F_Z^{-1}(\alpha). \]  \hfill (3.13)
The expected shortfall is given as
\[
\text{ES}_\alpha(Z) = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\gamma(Z) d\gamma
\]
\[
= -\frac{1}{\alpha} \left( E[Z 1_{\{Z \leq z_\alpha\}}] + z_\alpha [\alpha - F_Z(z_\alpha)] \right),
\]
(3.14)
where \(z_\alpha = \inf \{ z \in \mathbb{R} : F_Z(z) \geq \alpha \}\) is the lowest \(\alpha\)-quantile and \(1_A(z) = 1\) if \(x \in A\) and \(1_A(z) = 0\) else. The dual representation of (3.14) is given by
\[
\text{ES}_\alpha(Z) = \operatorname{inf}_{Q \in \mathcal{Q}} E_Q[Z],
\]
(3.15)
where \(\mathcal{Q}_\alpha\) is a set of probability measure \(Q\) which is absolutely continuous to the physical measure \(P\) such that \(dQ/dP \leq \alpha^{-1}\) almost surely.\(^1\) If \(F_Z\) (or \(f_Z\)) is known, then the computation of \(\text{VaR}_\alpha(Z)\) and \(\text{ES}_\alpha(Z)\) is straightforward from (3.13)-(3.15). The problem, however, is that the distribution of the hedging error is unknown and needs to be estimated. Suppose that we have a consistent estimate of \(f_Z\), say \(\hat{f}_Z\). So, we also estimate its cdf and compute the estimate of the value-at-risk and the expected shortfall as
\[
\hat{\text{VaR}}_\alpha(Z) = \hat{F}_Z^{-1}(\alpha), \quad \hat{\text{ES}}_\alpha(Z) = \frac{1}{\alpha} \int_0^\alpha \hat{\text{VaR}}_\gamma(Z) d\gamma.
\]
(3.16)
In this section, we seek an estimator \(\hat{f}_Z\) of \(f_Z\) that captures the maximum uncertainty in \(Z\). Before proceeding, it will be useful to first define the concept of information entropy.

3.1 Information Entropy and Density Estimation

The information entropy associated with \(Z\) is defined as:
\[
I_E(Z) = -\int f_Z(z) \ln f_Z(z) dz,
\]
(3.17)
where, by convention, we assume that \(0 \times \ln(0) = 0\). \(I_E(Z)\) is a measure of the information carried by \(Z\). As data are communicated more, they are corrupted with more noise so that the entropy increases, therefore they carry less information.

Let \(g_z^{(k)} : \Omega \to \mathbb{R} (k = 0, 1, 2, \ldots)\) be a moment function of \(Z\). Then, the moment of \(Z\) with respect to \(g_z^{(k)}\) is defined as:
\[
\mu_k = \mathbb{E}[g_z^{(k)}(z)] = \int \Omega g_z^{(k)}(z) f_Z(z) dx, \quad k = 0, 1, 2, \ldots
\]
(3.18)
In practice, polynomial functions are often used for the moment function\(^2\) \(g_z^{(k)} (k = 0, 1, 2, \ldots)\), i.e., \(g_z^{(k)}(z) = z^k\), \(k = 0, 1, 2, \ldots\), with the normalisation \(g_z^{(0)}(z) = 1\). In this case, we have:
\[
\mu_0 = \int \Omega f_Z(z) dz = 1, \quad \mu_k = \int \Omega z^k f_Z(z) dz, \quad k = 1, 2, \ldots
\]
(3.19)
\(^1\)If the distribution of \(Z\) is continuous, then the expected shortfall is equivalent to the tail conditional expectation defined by \(\text{TCE}_\alpha(Z) = E[-Z | Z \leq -\text{VaR}_\alpha(Z)]\).
\(^2\)For example, see Zellner and Highfield (1988).
and \( \mu_k \) is called the noncentral \( k \)th moment of \( Z \). We seek \( f_Z(z) \) such that:

\[
\max_{f_Z} \left\{ I_E(Z) = -\int_{\Omega} f_Z(z) \ln f_Z(z) dz \right\}
\]

subject to

\[
\bar{\mu}_k = \int_{\Omega} g^{(k)}_Z(z) f_Z(z) dz, \quad k = 0, 1, 2, \ldots, m \tag{3.20}
\]

where \( \bar{\mu}_k = n^{-1} \sum_{t=1}^{n} g^{(k)}_Z(Z_t) \) is the empirical moment of \( Z \) with respect to \( g^{(k)}_Z \), and \( m \) is the number of moments of the pdf that have been chosen to match the empirical moments. We assume that \( g^{(0)}_Z = 1 \). So, the zeroth moment is such that \( \mu_0 = \int_{\Omega} f_Z(z) dz = \bar{\mu}_0 = 1 \). For example if \( g^{(k)}_Z(z) = z^k \), \( k = 0, 1, 2, 3 \) and \( m = 4 \) with \( g^{(0)}_Z = 1 \), then problem (3.20) becomes

\[
\max_{f_Z} I_E(Z) \quad \text{subject to} \quad \bar{\mu}_0 = \int_{\Omega} f_Z(z) dz = 1, \quad \bar{\mu}_1 = \int_{\Omega} z f_Z(z) dz, \quad \bar{\mu}_2 = \int_{\Omega} z^2 f_Z(z) dz, \quad \bar{\mu}_3 = \int_{\Omega} z^3 f_Z(z) dz, \quad \bar{\mu}_4 = \int_{\Omega} z^4 f_Z(z) dz. \tag{3.21}
\]

In what follows we call the maximum-entropy estimator of \( f_Z(z) \) the solution \( \hat{f}_Z(z) \) of the problem (3.20). We will now prove that this solution \( \hat{f}_Z(z) \) is unique. Let

\[
L(f_Z, \lambda_0, \ldots, \lambda_m) = -I_E(Z) + \sum_{k=0}^{m} \lambda_k \left( \int_{\Omega} g^{(k)}_Z(z) f_Z(z) dz - \bar{\mu}_k \right) \tag{3.22}
\]

be the Lagrange function associated with (3.20), where \( \lambda_0, \ldots, \lambda_m \) are the Lagrange multipliers. By noting that solving (3.20) is equivalent to static problem

\[
\min_{f_Z, \lambda_0, \ldots, \lambda_m} L(f_Z, \lambda_0, \ldots, \lambda_m), \tag{3.23}
\]

it is sufficient to show that (3.23) has a unique solution with respect to \( f_Z \). The necessary conditions for a minimum in (3.23) are:

\[
\frac{\partial L(f_Z, \lambda_0, \ldots, \lambda_m)}{\partial f_Z} = 0, \quad \int_{\Omega} g^{(k)}_Z(z) f_Z(z) dz = \bar{\mu}_k, \quad k = 0, 1, \ldots, m. \tag{3.24}
\]

First, observe that

\[
\frac{\partial L(f_Z, \lambda_0, \ldots, \lambda_m)}{\partial f_Z} = \frac{\partial}{\partial f_Z} \left\{ \int_{\Omega} \left[ f_Z(z) \ln f_Z(z) + \sum_{k=0}^{m} \lambda_k g^{(k)}_Z(z) f_Z(z) \right] dz - \sum_{k=0}^{m} \lambda_k \bar{\mu}_k \right\}
\]

\[
= \int_{\Omega} \left[ \ln f_Z(z) + 1 + \sum_{k=0}^{m} \lambda_k g^{(k)}_Z(z) \right] dz
\]

\[
= 0
\]
where the last equality holds by the Leibniz integral rule. Therefore,

$$
\ln f_z(z) + 1 + \sum_{k=0}^{m} \lambda_k g_z^{(k)}(z) = 0 \iff f_z(z; \lambda) = \exp \left( -1 - \lambda_0 - \sum_{k=1}^{m} \lambda_k g_z^{(k)}(z) \right)
$$

$$
= \frac{1}{\theta} \exp \left( - \sum_{k=1}^{m} \lambda_k g_z^{(k)}(z) \right)
$$

(3.25)

where \( \lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)' \), \( \theta = \exp(1 + \lambda_0) \), and the zeroth-moment equality implies that

$$
\theta = \int_{\Omega} \exp \left( - \sum_{k=1}^{m} \lambda_k g_z^{(k)}(z) \right) dz.
$$

To complete the closed form of \( f_z(z; \lambda) \), we must substitute \( \lambda \) in (3.25) by an optimal value \( \hat{\lambda} \).

Zellner and Highfield (1988) consider the case in which \( g_z^{(k)}(z) = z^k \) where \( k = 1, \ldots, m = 4 \), and use an algorithm based on a Newton method to compute \( \lambda_1, \ldots, \lambda_4 \) from the restrictions in (3.24). This numerical approximation is cumbersome even for a moderate choice of \( m = 4 \).

In this paper, we propose the maximum likelihood (ML) method to estimate \( \lambda_1, \ldots, \lambda_m \), and \( \theta \).

Since \( \{Z_t : t = 1, \ldots, n\} \) are i.i.d. with common pdf \( \frac{1}{\theta} \exp \left( - \sum_{k=1}^{m} \lambda_k g_z^{(k)}(z_t) \right) \), the likelihood function of the sample can be written as

$$
\mathcal{L}(Z_1, \ldots, Z_n; \lambda) = \prod_{t=1}^{n} \frac{1}{\theta} \exp \left( - \sum_{k=1}^{m} \lambda_k g_z^{(k)}(z_t) \right) = \frac{1}{\theta^n} \exp \left[ - \sum_{t=1}^{n} \sum_{k=1}^{m} \lambda_k g_z^{(k)}(z_t) \right].
$$

(3.26)

The log-likelihood function from (3.26) is then given by

$$
\ell(Z_1, \ldots, Z_n; \lambda) = -n \ln(\theta) - \sum_{t=1}^{n} \sum_{k=1}^{m} \lambda_k g_z^{(k)}(z_t).
$$

The ML estimator of \( \lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)' \) satisfies:

$$
\max_{\lambda} \ell(Z_1, \ldots, Z_n; \lambda)
$$

(3.27)

and we can prove the following on the existence of unique solutions for both problems (3.27) and (3.20).

**Proposition 4.** Suppose that \( \bar{\mu}_k = n^{-1} \sum_{t=1}^{n} g_z^{(k)}(Z_t) < \infty \) \( \forall k = 1, \ldots, m \). Then:

(a) problem (3.27) has a unique solution with respect to \( \lambda \);

(b) problem (3.20) has a unique solution with respect to \( f_z(\cdot) \).

**Proof.** The first order condition of (3.27) is given by:

$$
\frac{\partial \ell(Z_1, \ldots, Z_n; \lambda)}{\partial \lambda_k} = -n \frac{\partial \ln(\theta)}{\partial \lambda_k} - \sum_{t=1}^{n} g_z^{(k)}(z_t) = 0
$$

$$
\iff \frac{\partial \ln(\theta)}{\partial \lambda_k} = -n^{-1} \sum_{t=1}^{n} g_z^{(k)}(z_t) = -\bar{\mu}_k, \ k = 1, \ldots, m.
$$

(3.28)
Moreover, $\theta = \exp(1 + \lambda_0) = \int_{\Omega} \exp\left(-\sum_{k=1}^{m} \lambda_k g_z^{(k)}(z_t)\right) dz_t$ from the zeroth-moment equality. Thus

$$
\frac{\partial \ln(\theta)}{\partial \lambda_k} = \frac{1}{\theta} \frac{\partial}{\partial \lambda_k} \int_{\Omega} \exp\left(-\sum_{k=1}^{m} \lambda_k g_z^{(k)}(z_t)\right) dz_t = -\int_{\Omega} g_z^{(k)}(z_t) \frac{1}{\theta} \exp\left(-\sum_{k=1}^{m} \lambda_k g_z^{(k)}(z_t)\right) dz_t
$$

where $\mu_k(\lambda)$ is a continuous function of $\lambda$. Therefore, (3.28) and (3.29) entail that

$$
\mu_k(\lambda) = \bar{\mu}_k, \quad k = 1, \ldots, m. \tag{3.30}
$$

So, as long as $\bar{\mu}_k < \infty$, (3.30) has a solution with respect to $\lambda$, i.e., the likelihood function $\ell(Z_1, \ldots, Z_n; \lambda)$ has a critical point $\hat{\lambda}$. We show that this critical point, $\hat{\lambda}$, is the unique point where $\ell(Z_1, \ldots, Z_n; \lambda)$ is maximised.

Let $\mathcal{H}(Z_1, \ldots, Z_n; \lambda)$ denote the hessian matrix of the log-likelihood function $\ell(Z_1, \ldots, Z_n; \lambda)$. The $(k,p)$ entree of $\mathcal{H}(Z_1, \ldots, Z_n; \lambda)$ is:

$$
\frac{\partial}{\partial \lambda_p} \left( \frac{\partial \ell(Z_1, \ldots, Z_n; \lambda)}{\partial \lambda_k} \right) = \frac{\partial}{\partial \lambda_p} \left( -n \frac{\partial \ln(\theta)}{\partial \lambda_k} - \sum_{t=1}^{n} \frac{g_z^{(k)}(z_t)}{\theta} \right) = -n \frac{\partial \ln(\theta)}{\partial \lambda_p} \frac{\partial \ln(\theta)}{\partial \lambda_k}
$$

$$
= n \frac{\partial}{\partial \lambda_p} \left( \frac{1}{\theta} \int_{\Omega} g_z^{(k)}(z) \exp\left(-\sum_{k=1}^{m} \lambda_k g_z^{(k)}(z)\right) dz \right)
$$

$$
= -n \frac{1}{\theta} \int_{\Omega} g_z^{(p)}(z) g_z^{(k)}(z) \exp\left(-\sum_{k=1}^{m} \lambda_k g_z^{(k)}(z)\right) dz + n \int_{\Omega} g_z^{(p)}(z) \exp\left(-\sum_{k=1}^{m} \lambda_k g_z^{(k)}(z)\right) dz \times \int_{\Omega} g_z^{(k)}(z) \exp\left(-\sum_{k=1}^{m} \lambda_k g_z^{(k)}(z)\right) dz
$$

$$
= -n \left[ \int_{\Omega} g_z^{(p)}(z) g_z^{(k)}(z) f_Z(z) dz - \int_{\Omega} g_z^{(p)}(z) f_Z(z) dz \int_{\Omega} g_z^{(k)}(z) f_Z(z) dz \right]
$$

$$
= -n \left[ \mathbb{E}(g_z^{(p)}(z) g_z^{(k)}(z)) - \mathbb{E}(g_z^{(p)}(z)) \mathbb{E}(g_z^{(k)}(z)) \right]
$$

$$
= -n \text{CoV}(g_z^{(p)}(z), g_z^{(k)}(z)), \quad \forall \ p, k = 1, \ldots, m, \tag{3.31}
$$

where $\text{CoV}(a, b)$ is the covariance of the two random variables $a$ and $b$. From (3.31), it is clear that $\mathcal{H}(Z_1, \ldots, Z_n; \lambda)$ is symmetric and strictly negative definite for all values of the vector of Lagrange multipliers. Thus, the critical point $\hat{\lambda}$ is a unique maximum, which completes the proof of Proposition 4(a). It follows immediately from (3.25) that $\hat{f}_z(z; \hat{\lambda})$ is also a unique maximum, thus establishing Proposition 4(b).

### 3.2 Selection of Moment Function

Note that from (3.25), we have

$$
f_z(z; \lambda) = \frac{1}{\theta} \exp\left(-\sum_{k=1}^{m} \lambda_k g_z^{(k)}(z)\right). \tag{3.32}
$$
If each \( g_z^{(k)}(z) \) is replaced with a truncated Taylor series expansion of \( g_z^{(k)}(z) \) around the expectation of \( z \), then the problem of looking for the proper \( g_z^{(k)}(z) \), \( k = 0, 1, \ldots, m \), moment functions is the same as finding an optimal order of truncation for each of the expansion:

\[
\hat{\lambda} = \arg \max_{\lambda} \prod_{t=1}^{n} \frac{1}{\theta} \exp \left( -\sum_{k=1}^{m} \lambda_k \left[ a_{1k} + a_{2k} z_t^2 + a_{3k} z_t^3 + \ldots \right] \right)
\]

\[
= \arg \max_{\lambda} \prod_{t=1}^{n} \frac{1}{\theta} \exp \left( \beta_0 + \beta_1 z_t^2 + \beta_3 z_t^3 + \ldots \right), \tag{3.33}
\]

where \( \beta_0, \beta_1, \beta_3, \ldots \) are constant to be estimated. We seek the truncation order \( l_0 \) of the power series in (3.33). So, the search for the moment functions is converted to a search for an optimal truncation order, \( l_0 \), that yields the best fit of \( \hat{f}_z(z; \hat{\lambda}) \) to the data. We suggest using Bayesian information criterion (BIC) or Schwartz information criterion (SIC) based on (3.33) to select the optimal \( l_0 \). Let \( \hat{\lambda}_{MLE} \) denote the ML estimator of \( \lambda \) using the optimal order of truncation \( l_0 \). Then, the estimated pdf of \( f_z(z) \) is \( \hat{f}_z(z; \hat{\lambda}_{MLE}) \), which is used to compute \( \hat{\text{VaR}}_{\alpha}(Z) \) and \( \hat{\text{ES}}_{\alpha}(Z) \) in Equation (3.16).

4 Numerical Analysis

In this section we conduct a numerical experiment to analyse the sensitivity of the hedging error with respect to model parameters. In the previous sections we have defined a general jump-diffusion process with the jump component specified by a kernel biased completely random measure. This generalised framework nests a number of very important models in mathematical finance, including, but not limited to, the jump diffusion model of Merton (1976), the Generalised Gamma process discussed in Lo and Weng (1989), the Variance Gamma process by Madan et al. (1998) and the CGMY model of Carr et al. (2002). Here, for simplicity, we only use the Generalised Gamma (GG) process. The analysis can be easily extended to other classes of models, or even their Markovian regime switching versions discussed in Fard and Siu (2013).

The GG process is the generalised form of some famous models in finance, namely, the weighted gamma (WG) process and inverse gamma (IG) process. Let \( \varsigma \leq 1 \) denote a constant shape parameter and \( \delta(t) \) denote the time-dependent scale parameter of the GG process. Then the intensity process of the GG process is

\[
\varrho(dz|t)\eta(dt) = \frac{1}{\Gamma(\varsigma \delta(t)) z^{\varsigma}} dz\eta(dt),
\]

where \( \Gamma \) is the gamma density function. When \( \varsigma = 1 \) or \( \varsigma = \frac{1}{2} \), the GG process reduces to a WG process or IG process, respectively.

The GG process is obtained by setting the kernel function \( h(z) \) to \( cz^q \), where \( c \) and \( q \) are constants, and choosing a particular parametric form of the compensator measure. The scale-distorted version is achieved when the kernel function \( c \) is a positive constant and \( q = 1 \). When \( c > 1 \), jump sizes are overstated. When \( 0 < c < 1 \), jump sizes are understated. For the power-distorted version of the GG process, the kernel function \( q \) is a positive constant and \( c = 1 \).
When $q > 1$, small jump sizes (i.e., $0 < z < 1$) are understated and large jump sizes (i.e., $z > 1$) are overstated. When $0 < q < 1$, small jump sizes are overstated and large jump sizes are understated.

For simulating the GG process, we adopt the Poisson weighted algorithm by Lee and Kim (2004) to simulate completely random measures. The Poisson weighted algorithm is applicable for a wide class of completely random measures, which are very difficult to simulate directly. The main idea of the Poisson weighted algorithm is that instead of generating jump sizes of a completely random measure directly from a non-standard density function, one can first generate jump sizes from a proposed density function (e.g. a gamma density) and then adjust the simulated jump sizes by the corresponding Poisson weights. The Poisson weights are simulated from a Poisson distribution with an intensity parameter given by the odd ratio of the compensator of the completely random measure and the compensator corresponding to the proposed density.

To implement the algorithm, divide the time interval to maturity $[0, T]$ into $nT$ equally spaced subintervals. Then for each $j = 0, 1, ..., n - 1$, let $[t_j, t_{j+1}]$ be the $(j + 1)$st subinterval. Let $M$ denote the number of jumps of the completely random measure over the term to maturity, such that, $M$ controls the accuracy of the approximation of the algorithm. To implement the Poisson weighted algorithm, we take the following steps:

Step 1. Let $\mathcal{N}$ denote a normalised density function defined by $\mathcal{N}^{-1} := \frac{1}{\eta'(t)} \int_0^T \eta'(s) ds$.

Step 2. Generate a set of i.i.d random variables $T_i \geq 0$ for each $i = 1, 2, ..., MT$, from $\mathcal{N}$.

Step 3. Generate the jump size $\Omega_i$ from the conditional density function $g_{T_i}$ (i.e. gamma).

Step 4. Evaluate whether $T_i \in [t_j, t_{j+1})$. If yes, calculate

$$
\lambda_i = \frac{\phi^\theta(\Omega_i|T_i)}{MT g_{T_i}(\Omega_i)} \int_0^T \eta'(s) ds.
$$

$\lambda_i$ is the intensity of the Poisson distribution, used to generate the Poisson weights $\mathcal{W}$.

We provide a comparative analysis of the density function of the hedging error obtained by our ME estimation of the hedging error with that obtained by Monte Carlo simulation. In Monte Carlo simulation, we calculate the hedging error at $t_n$, $\Delta \Pi_n$, $n = 1, \ldots, N$, with hedging portfolio $\Pi_n := \Pi(t_n, S)$ defined by

$$
\Pi(t_n, s) = S(t_n) H(t_n, S) - C(t_n, S)
$$

where the option price and delta are computed by evaluating the risk neutral expectation in equation (2.8).

We consider hedging a one-month at the money call option with $S(0) = K = 1$. We analyse different hedging frequencies, namely, $N = 4$ (weekly hedging), $N = 20$ (daily hedging), $N = 80$ (hedging every hour and a half), and $N = 320$ (hedging every 20 minutes). For brevity, we suppose that the WG process with $c = q = 1$, equal model parameters under $\mathbb{P}$ and $\mathbb{P}^\theta$, zero interest rate, and zero drift for the underlying asset. Further, we assume constant volatility,
Figure 1: Distribution of hedging error of a one-month European call option under the Weighted Gamma, obtained using Monte Carlo simulation and computing Max Entropy estimation

\( \sigma = 18\% \). Then, the set of rebalancing times is \( \{t_n\} \), \( t_n = n\delta t \), \( n = 1, \ldots, N \). All computations are fully vectorised in JULIA\textsuperscript{TM}.

The kernel biased completely random measure on \( \mathcal{T} \) is \( \kappa(t) = \sum_{i=1}^{MT} h(\Omega_i \mathcal{W}_i) \mathbf{1}(T_i \leq t) \). The process generated by the Poisson weighting algorithm converges in distribution to the completely random measure \( \kappa(t) \) of real-valued functions defined on the compact domain \([0, T] \) with the Skorohod topology as \( M \to \infty \). For the proof of this convergence, interested readers may refer to Lee and Kim (2004).

Figure 1 compares the distribution of hedging error obtained from the Monte Carlo Simulation and estimated from the ME method. In Table 1 we report the summary statistics for the distribution of hedging error, as well as the 99% and 95% VaR, obtained by the above two methods. From the numerical analysis, it follows that the distribution of hedging error for a small number of trades has a high volatility and is negatively skewed, which indicates that it is more likely to have large losses than large wins. It is noteworthy that the ME estimation of the distribution fails to sufficiently capture the right tail of the empirical distribution, nevertheless it adequately describes the left tail.
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M1 (%)</th>
<th>M2 (%)</th>
<th>M3 (%)</th>
<th>M4 (%)</th>
<th>99% VaR</th>
<th>95% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Entropy</td>
<td>4</td>
<td>0.173</td>
<td>43.21</td>
<td>-1.97</td>
<td>7.01</td>
<td>99.87</td>
<td>74.38</td>
</tr>
<tr>
<td>Max Entropy</td>
<td>20</td>
<td>0.154</td>
<td>28.44</td>
<td>-2.41</td>
<td>16.01</td>
<td>88.31</td>
<td>72.11</td>
</tr>
<tr>
<td>Max Entropy</td>
<td>80</td>
<td>-0.058</td>
<td>26.04</td>
<td>-3.36</td>
<td>19.74</td>
<td>77.59</td>
<td>51.6</td>
</tr>
<tr>
<td>Max Entropy</td>
<td>320</td>
<td>-0.041</td>
<td>22.14</td>
<td>-3.99</td>
<td>14.32</td>
<td>61.33</td>
<td>46.31</td>
</tr>
<tr>
<td>Monte Carlo</td>
<td>4</td>
<td>0.131</td>
<td>42.52</td>
<td>-1.03</td>
<td>6.51</td>
<td>101.73</td>
<td>75.12</td>
</tr>
<tr>
<td>Monte Carlo</td>
<td>20</td>
<td>0.143</td>
<td>29.37</td>
<td>-2.33</td>
<td>15.36</td>
<td>76.12</td>
<td>62.49</td>
</tr>
<tr>
<td>Monte Carlo</td>
<td>80</td>
<td>-0.072</td>
<td>25.41</td>
<td>-3.21</td>
<td>19.85</td>
<td>74.06</td>
<td>49.33</td>
</tr>
<tr>
<td>Monte Carlo</td>
<td>320</td>
<td>-0.074</td>
<td>23.33</td>
<td>-4.01</td>
<td>18.94</td>
<td>64.83</td>
<td>48.99</td>
</tr>
</tbody>
</table>

Table 1: Summary statistics of the distribution of hedging error under the GG process

As we increase the frequency of the delta hedging trades, the distribution becomes more symmetric, and our ME estimation performs better. For $N = 20$, our distribution is strongly leptokurtic, indicating option writers’ large exposure to jumps. When trading frequency increases to 80 and 320, the distribution peaks around zero and the volatility significantly decreases, however, the left tail is still noticeable. From our analysis, it appears that the volatility is negatively related to the frequency of delta hedging trades. However, further research is required to conclusively establish the relationship between the number of trades and the moments of the distributions.

5 Conclusion

Perfect replication of financial derivatives is not possible, given market incompleteness and discrete-time hedging. We characterise the risk in dynamic hedge of European options through the asymptotic distribution of hedging error. Further, we obtain an estimation for the distribution of hedging error by maximising Shannon’s entropy (Shannon (1948)) subject to a set of moment constraints, which in turn yield the value-at-risk (VaR) and expected shortfall (ES) of the hedging error, two widely used risk metrics in finance. This new approach chooses the probability distribution with the most uncertainty subject to what is known. Thus we obtain a consistent estimate of the asymptotic distribution of hedging error, despite the non-normality of the underlying distribution of returns. As a result, we derive a very generalised modelling framework, which can be applied in different areas of derivatives pricing. Some parametric specifications of this framework include, but are not limited to, the jump diffusion model of Merton (1976), the Generalised Gamma process discussed in Lo and Weng (1989), the Variance Gamma process by Madan et al. (1998), and the CGMY model of Carr et al. (2002).

Finally, we conduct a robust numerical simulation of the result, to highlight the practical applications of our model.
Acknowledgement

The authors would like to thank Professor Mervyn Silvapulle of the Department of Econometrics and Business Statistics at Monash University, for his insightful comments and helpful suggestions.

References


**Appendix**

**Proof for Proposition 1**

*Proof.* Since \( \bar{S} \) is \( \mathcal{F}^A \)-adapted, \( \bar{S} \) is an \((\mathcal{F}^Y, \mathbb{P}^\theta)\)-local-martingale if and only if it is a \((\mathcal{G}, \mathbb{P}^\theta)\)-local-martingale. By Lemma 7.2.2 in Elliott and Kopp (2005), \( \bar{S} \) is an \((\mathcal{G}^Y, \mathbb{P}^\theta)\)-local-martingale if and only if \( \Lambda \bar{S} := \{ \Lambda(t) \bar{S}(t) | t \in \mathcal{F} \} \) is a \((\mathcal{G}, \mathbb{P}^\theta)\)-local-martingale. First, by Bayes’ rule

\[
E^\theta \left[ \exp \left( - \int_0^t r_s ds \right) S_t \mid \mathcal{G}_0 \right] = \exp \left( - \int_0^t r_s ds \right) E \left[ \Lambda_t \exp \left( \int_0^t dA_u \right) \mid \mathcal{G}_0 \right]
\]

\[
= \exp \left( - \int_0^t r_s ds \right) \frac{E \left[ \exp \left( \int_0^t (\theta + 1) dA_u \right) \mid \mathcal{G}_0 \right]}{\mathcal{M}(\theta)_t}
\]

\[
= \exp \left( - \int_0^t r_s ds \right) \frac{\mathcal{M}(\theta + 1)_t}{\mathcal{M}(\theta)_t}
\]

\[
= \exp \left( \int_0^t (\mu - r_s - \frac{1}{2} \sigma^2_s) ds + \frac{1}{2} \int_0^t (2\theta + 1) \sigma^2_s ds \right)
\]

\[
+ \int_0^t \int_{\mathbb{R}^+} \left\{ e^{\theta h(z)} (e^{h(z)} - 1) - h(z) \right\} \rho d(z | t) \eta(dt).
\]
Then by setting time $s = 0$, and applying the martingale condition we achieve
\[
\int_0^t (\mu_s - r_s - \frac{1}{2}\sigma_s^2) ds + \frac{1}{2} \int_0^t (1 + 2\theta)\sigma_s^2 + \int_0^t \int_{\mathcal{R}^+} \{ e^{\theta h(z)}(e^{h(z)} - 1) - h(z) \} \rho(dz|s)\eta'(s) ds = 0.
\]

Hence, for each $t \in \mathcal{T}$, (2.5) must hold. \qed

**Proof for Proposition 2**

**Proof.** Assume that $\mathbb{P} \sim \mathbb{P}^\theta$ with density process $\Lambda_t$. Suppose $\mathcal{Z}_u \in BM(\mathcal{T})$. Then by Bayes' rule
\[
\mathcal{M}^\theta_A(\mathcal{Z})_t := E^{\theta}[e^{(\mathcal{Z}.A)t}|\mathcal{G}_0] = E[\Lambda_t e^{(\mathcal{Z}.A)t}|\mathcal{G}_0]
\]
\[
= \exp \left( \int_0^t \mathcal{Z}(\mu_s - \frac{1}{2}\sigma_s^2) ds + \int_0^t \frac{1}{2}(\mathcal{Z} + \theta)^2 \sigma_s^2 ds 
+ \int_0^t \int_{\mathcal{R}^+} \left\{ e^{(\mathcal{Z} + \theta)h(z)} - 1 - (\mathcal{Z} + \theta)h(z) \right\} \rho(dz|s)\eta(ds) 
- \int_0^t \frac{1}{2}(\theta, \sigma_s)^2 ds 
- \int_0^t \int_{\mathcal{R}^+} \left\{ e^{\theta h(z)} - 1 - \theta h(z) \right\} \rho(dz|s)\eta(ds) \right)
\]
\[
= \exp \left( \int_0^t \mathcal{Z}(\mu_s + 2\theta\sigma_s^2 - \frac{1}{2}\sigma_s^2) ds 
+ \int_0^t \int_{\mathcal{R}^+} (e^{\theta h(z)} - 1) \rho(dz|s)\eta(ds) + \frac{1}{2} \int_0^t \mathcal{Z}^2 \sigma_s^2 ds 
+ \int_0^t \int_{\mathcal{R}^+} \left\{ e^{\mathcal{Z} h(z)} - 1 - \mathcal{Z}^2 h(z) \right\} e^{\theta h(z)} \rho(dz|s)\eta(ds) \right).
\]

Then under $\mathbb{P}^\theta$, (2.6) holds. \qed