HYPOTHESIS TESTING ON TIME SERIES DRIVEN BY UNDERLYING LÉVY PROCESSES, WITH MACHINE LEARNING APPLICATIONS

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Michael Roberts

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Mathematics

May 2021

Fargo, North Dakota

NORTH DAKOTA STATE UNIVERSITY

Graduate School

Title

HYPOTHESIS TESTING ON TIME SERIES DRIVEN BY UNDERLYING LÉVY PROCESSES, WITH MACHINE LEARNING APPLICATIONS

By

Michael Roberts

The supervisory committee certifies that this dissertation complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Indranil SenGupta
Chair

Dr. Azer Akhmedov

Dr. Artem Novozhilov

Dr. William Nganje

Approved:

5/14/2021
Date

Dr. Friedrich Littmann
Department Chair

ABSTRACT

In this dissertation, we study the testing of hypotheses on streams of observations that are driven by Lévy processes. This is applicable for sequential decision making on the state of two-sensor systems. In one case, each sensor receives or does not receive a signal obstructed by noise. In another, each sensor receives data driven by Lévy processes with large or small jumps. In either case, these give rise to four possible outcomes for the hypotheses. Infinitesimal generators are presented and analyzed. Bounds for likelihood functions in terms of *super-solutions* and *sub-solutions* are computed. As an application, we study a *change point detection* hypothesis test for the detection of the distribution of jump size in one-dimensional Lévy processes. This is shown to be implementable in relation to various classification problems for a crude oil price data set. Machine and deep learning algorithms are implemented to extract a specific deterministic component from the data set, and the deterministic component is implemented to improve the Barndorff-Nielsen & Shephard (BN-S) model, a commonly used stochastic model for derivative and commodity market analysis.

ACKNOWLEDGEMENTS

I'd like to thank all of the professors I've had here at NDSU, especially Dr. Novozhilov for providing many exciting courses, and Dr. SenGupta for leading me to an intriguing thesis topic and guiding me through the entire research process with great patience.

DEDICATION

This thesis is dedicated to my parents. They instilled in me a deep desire for education: my mother, by lovingly insisting I can do better when getting a single B in middle school math, and my father, by noticing the seed for a love of math and nourishing it with lectures on material I could not yet fully understand but still found fascinating.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

One of the most classical problems arising in statistical sequential analysis is *sequential hypothesis testing* (see [32]). As described in [32], a sequential test of a hypothesis means any statistical test that gives a specific rule, at any stage of the experiment for making one of the three decisions: (1) to accept the null hypothesis H_0, (2) to reject H_0, (3) to continue the experiment by making additional observation. Consequently, the test is carried out *sequentially*. An objective for the analysis of such test is to minimize the number of observations required to make a decision subject to a given tolerance level described as Type I and Type II errors. A related problem is *change point detection*, where the goal is to test hypotheses concerning the parameters of the observation stream in order to detect the point at which these parameters shift significantly. This dissertation addresses these two problems in specific cases where the underlying processes are assumed to be Lévy processes.

Clearly, there are an abundance of applications of these theories, including commodity price analysis, earthquake modeling [10], and even human weight fluctuations [27]. Basically, any sequential data could be a suitable target application. Analyzing the price of crude oil as it varies in time is the primary application of this dissertation, but hopefully, the theorems and methods here can be used to aid in a variety of other goals yet to come.

There are a couple of primary approaches to this problem of developing sequential hypothesis tests, e.g., the *Bayesian* and the *min-max*. For the first approach, each hypothesis is assigned with an a priori probability. For the second approach, no such assumption is made and the optimal solution is known to be given by the sequential probability ratio test (see [11]). The sequential probability ratio test is revisited and improved in various works. For a sequential decision problem, it is assumed that the amount of available information is increasing with time. But it is often difficult to handle all the data as represented by a σ-algebra as the actual amount may be very large. In [18] a reduction method is proposed that takes into account the underlying statistical structure.

An approach of improving the sequential hypothesis testing in the Bayesian case is presented in [15]. The sequential testing of more than two hypotheses has many important applications. In

[7] a sequential test (termed as MSPRT), which is a generalization of the sequential probability ratio test, is studied. It is shown that, under Bayesian assumptions, the MSPRT approximates optimal tests that are more intricate when error probabilities are small and expected stopping times are large. In [13], a sequential hypothesis test is conducted when there are finitely many simple hypotheses about the unknown arrival rate and mark distribution of a compound Poisson process, where exactly one is correct. This problem is formulated in a Bayesian framework when the objective is to determine the correct hypothesis with minimal error probability. A solution of this problem is presented in that paper. In the paper [8], an improved min-max approach to both sequential testing of many composite hypotheses and multi-decision change-point detection for composite alternatives is proposed. New performance measures for methods of hypothesis testing and change-point detection are introduced, and theoretical lower bounds for these performance measures are proved that do not depend on methods of sequential testing and detection. Minimax tests are proposed for which these lower bounds are attained asymptotically as decision thresholds tend to infinity.

In particular, we further present in this dissertation sequential hypothesis tests for the detection of the distribution of jump size in Lévy processes. Infinitesimal generators for the corresponding log-likelihood ratios are presented and analyzed. Bounds for infinitesimal generators in terms of super-solutions and sub-solutions are computed. This is shown to be implementable in relation to various classification problems for a crude oil price data set. Machine and deep learning algorithms are implemented to extract a specific deterministic component from the data set, and the deterministic component is implemented to improve the BN-S model, a commonly used stochastic model for derivative and commodity market analysis.

The organization of the dissertation is as follows. The next chapter introduces the foundational concepts pertaining to the rest of the dissertation. Chapter 3, which is based on my previously publish paper [25], motivates and advances the concept of sequential hypothesis tests with two streams of data. In Section 3.2, we generalize the result in [9] to a certain class of underlying Lévy processes, and then propose a decision rule to determine whether there is drift in the underlying processes. In Section 3.3, we present the infinitesimal generators for the hypothesis testing of *big* versus *small* fluctuations. This is conducted by considering two Lévy processes, where one has more *jump intensity* than the other. The existence of a viscosity solution to the related

partial-integro differential equation is shown, and its super- and sub-solutions are analyzed in that section. Applying the ideas of Chapter 3 to a data series of oil prices is the goal of Chapter 4, which is based on my published article [26]. In Section 4.2, a refined BN-S model with some of its properties is presented. In Section 4.3, we provide a general one-dimensional jump size detection analysis based on these hypothesis tests. In Section 4.4, an overview of the data set is provided, and then two procedures in the predictive classification problem are introduced. Concrete numerical results are shown in Section 4.5, and finally, Chapter 5 is a conclusion with potential future research goals and opportunities.

2. PRELIMINARY BACKGROUND

2.1. Brownian Motion

There are many phenomena in nature that are seemingly random. Ranging from the diffusion of organisms into habitable land, to the price of a stock or commodity, to the behavioral patterns of humans, stochastic processes have an abundance of applications. One of the most classic examples is the Brownian motion. The motivation for such a process is a continuous symmetric random walk.

Paraphrasing Chapter 3 of the book [31], let $\omega = \{\omega_1, \omega_2, \ldots\}$ be a sequence of results of fair coin tosses. That is, ω_i is the outcome of the ith toss, H or T. Define

$$X_i = \begin{cases} 1 \text{ if } \omega_i = H \\ \\ -1 \text{ if } \omega_i = T \end{cases},$$

and $M_n = \sum_{i=1}^{n} X_i$. Then $\{0, M_1, M_2, \ldots\}$ is a *symmetric random walk*. In particular, symmetric random walks are martingales that have independent increments. A *martingale* is a stochastic process that has expected value equal to the given present value for all times in the future. To have *independent increments* means for each $n \in \mathbb{N}$ and each $0 \leq t_1 \leq t_2 \leq \ldots < t_{n+1} < \infty$, the random variables $(M(t_{j+1}) - M(t_j), 1 \leq j \leq n)$ are independent.

Another crucial property of a symmetric random walk is its non-zero quadratic variation. In general, the *quadratic variation* of a discrete stochastic process M is

$$[M, M]_k := \sum_{j=1}^{k} (M_j - M_{j-1})^2,$$

which simplifies quite conveniently in our case to k. Note that while $[M, M]_k = \text{Var}(M_k) = k$ for a symmetric random walk, this is not true in general. One varies dramatically for changes in the probabilities of each coin toss, while the other, the quadratic variation, remains constant.

Continuing toward the goal of a continuous random walk, define the scaled symmetric random walk by

$$W_t^{(n)} = \frac{1}{\sqrt{n}} M_{nt},$$

where $nt \in \mathbb{Z}$. Otherwise, define $W_t^{(n)}$ by a linear interpolation of its values for the closest integers. This new process is similarly a martingale with independent increments and quadratic variation, for $nt \in \mathbb{Z}$,

$$[W^{(n)}, W^{(n)}]_t = t.$$

Finally, we obtain a standard Brownian motion as the limit of this sequence of scaled random walks. This inspires the following definition:

Definition 2.1.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. For each $\omega \in \Omega$, suppose there is a continuous function $W : [0, \infty) \to \mathbb{R}$ that satisfies $W(0) = 0$ and that depends on ω. Then $\{W(t), t \geq 0\}$ is a* Brownian motion *if for all $0 = t_0 < t_1 < \ldots < t_m$, the increments*

$$W(t_1) - W(t_0), W(t_2) - W(t_1), \ldots, W(t_m) - W(t_{m-1})$$

are independent and each is normally distributed with

$$E[W(t_{i+1}) - W(t_i)] = 0,$$

$$\text{Var}[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i.$$

2.2. Lévy Processes

While Brownian motions are a classic and powerful tool for modeling a wide range of phenomena, sometimes the processes take on a more sudden nature, and a process with discontinuous capabilities is more adequate. Lévy processes are a general class of such processes. In [1], we have the following definition of a Lévy process:

Definition 2.2.1. *Let $(X(t), t \geq 0)$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$. We say that it has* stationary increments *if each $X(t_{j+1}) - X(t_j) \stackrel{d}{=} X(t_{j+1} - t_j) - X(0)$.*

We say that $X(t)$ is a Lévy process *if $X(0) = 0$ (a.s.); X has independent and stationary increments; and X is* stochastically continuous*; i.e., for all $a > 0$ and for all $s \geq 0$,*

$$\lim_{t \to s} P(|X(t) - X(s)| > a) = 0.$$

Having a quick way of classifying Lévy processes is crucial to the remainder of this dissertation. To do so, we use the following definition and theorem from [1]:

Definition 2.2.2. *Let ν be a Borel measure defined on $\mathbb{R}^d \setminus \{0\}$. We say that it is a* Lévy measure *if*

$$\int_{\mathbb{R}^d \setminus \{0\}} (|y|^2 \wedge 1)\nu(dy) < \infty,$$

where $a \wedge b := \min\{a, b\}$ for any $a, b \in \mathbb{R}$.

Theorem 2.2.3. (Lévy-Itô decomposition) *Let $(X_t)_{t \geq 0}$ be a Lévy process on $\mathbb{R}$ and ν its Lévy measure. Then*

1. *ν is a random measure on $\mathbb{R} \setminus \{0\}$ and verifies: $\int_{|x| \leq 1} |x|^2 \nu(dx) < \infty$, and $\int_{|x| \geq 1} \nu(dx) < \infty$.*
2. *The jump measure of X, denoted by J_X, is a Poisson random measure on $[0, \infty) \times \mathbb{R}$ with intensity measure $\nu(dx)dt$.*
3. *There exist $\gamma, \sigma \in \mathbb{R}$, with $\sigma > 0$, and a Brownian motion $(W_t)_{t \geq 0}$ such that*

$$X_t = \gamma t + \sigma W_t + X_t^l + \lim_{\epsilon \to 0} \tilde{X}_t^\epsilon, \tag{2.1}$$

   where $X_t^l = \int_{|x| \geq 1, s \in [0,t]} x J_X(ds \times dx)$, and $\tilde{X}_t^\epsilon = \int_{\epsilon \leq |x| < 1, s \in [0,t]} x \left(J_X(ds \times dx) - \nu(dx)ds\right)$.
4. *The terms in (2.1) are independent and the convergence in $\tilde{X}_t^\epsilon$ is almost sure and uniform in $t \in [0, T]$.*

In particular, every Lévy process is *uniquely* determined by its characteristic triplet (γ, σ, ν) in the decomposition above. Many of the novel theorems in this dissertation rely on manipulations of these characteristic triplets.

There are a host of familiar processes that can be represented with these characteristics. The following are examples, along with representative sample paths. These show that Lévy processes

are suited to model a wide variety of phenomena, including the prices of commodities, as considered in Chapter 3.

- **Example 1** Brownian motion: $\gamma = 0$, $\sigma = 1$, $\nu(dx) = 0$.

- **Example 2** Poisson Process: $\gamma = 0$, $\sigma = 0$, $\nu(dx) = \lambda\delta_1(dx)$.

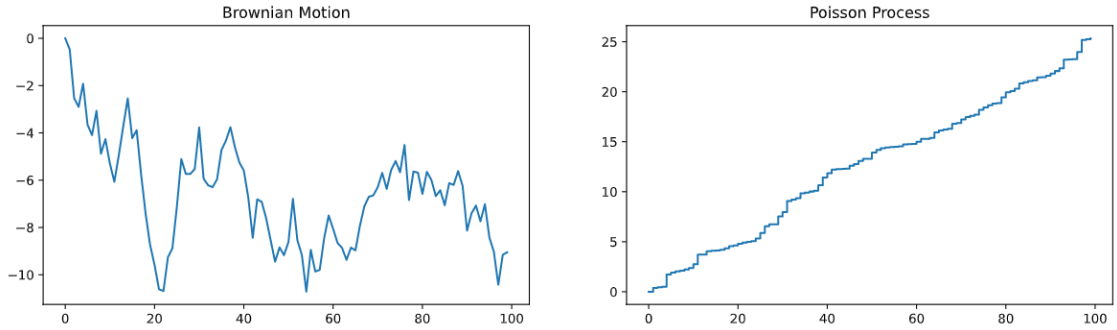

Figure 2.1. Sample paths of a standard Brownian motion and of a Poisson process with $\lambda = 1/4$.

- **Example 3** Gamma Process: $\gamma = -\int_0^1 x\nu(dx)$, $\sigma = 0$, $\nu(dx) = \beta x^{-1}e^{-\alpha x}1_{x\geq 0}\,dx$.

- **Example 4** Cauchy Process: $\gamma = 0$, $\sigma = 0$, $\nu(dx) = f(x)dx$, where $f(x) = |x|^{-2}$, $x \neq 0$.

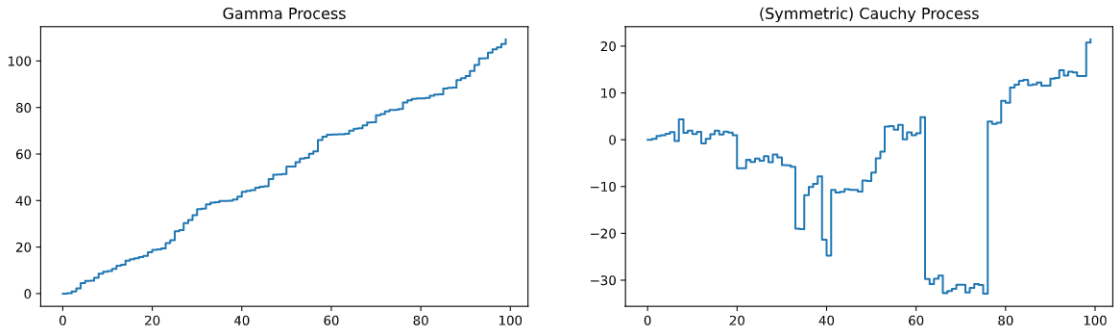

Figure 2.2. Sample paths of a Gamma process with mean and variance 1 and of a Cauchy process.

- **Example 5** Wiener Process: $\gamma = m$, $\sigma = s$, $\nu(dx) = 0$.

- **Example 6** Subordinator Jump Process: $\gamma = 0$, $\sigma = 0$, $\nu(dx) = f(x)dx$, where $f(x) \geq 0$: $x > 0$ and $f(x) = 0$: $x < 0$.

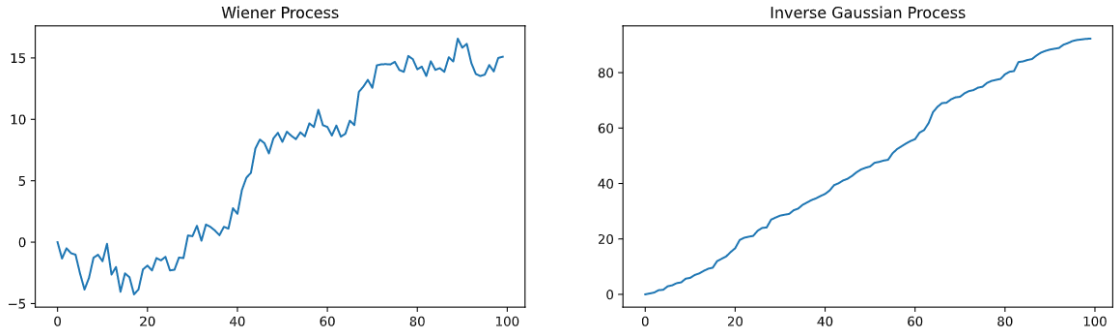

Figure 2.3. Sample paths of a Wiener process with $\gamma = 0.2$ and $\sigma = 1$ and of a subordinator process, an inverse Gaussian process with mean 1.

2.3. Itô Calculus

The rest of the dissertation uses multiple different concepts of integration. In particular, we often integrate with respect to some stochastic process. As such, it is important to understand the following definition from [12]:

Definition 2.3.1. *Let W_t be a Brownian motion and ϕ be a* simple cádlág (right-continuous with left limits) process *with partition $\pi = (0 = T_0, T_1, ..., T_{n+1} = T)$; i.e.*

$$\phi_t = \phi_0 1_{t=0} + \sum_{i=0}^{n} \phi_i 1_{[T_i, T_{i+1})}.$$

Then the Brownian stochastic integral $\int \phi dW$ is defined as

$$\int_0^T \phi_t dW_t = \sum_{i=0}^{n} \phi_i (W_{T_{i+1}} - W_{T_i}).$$

The above definition can, as usual in analysis, be extended to the closure of the set of simple processes. Of note, the set of square-integrable cádlág processes adapted to the same filtration as the Brownian motion, is in the closure.

This definition gives rise to another definition of a class of processes called Itô processes, by which we will define another integral.

Definition 2.3.2. ([31]) *Let $W(t), t \geq 0$ be a Brownian motion and $\mathcal{F}(t), t \geq 0$ be an associated filtration. An* Itô process *is a stochastic process of the form*

$$X(t) = X(0) + \int_0^t \Delta(t)dW(u) + \int_0^t \Theta(u)du,$$

where $X(0)$ is nonrandom, and Δ and Θ are adapted stochastic processes.

Theorem 2.3.3. *In particular, Itô processes have quadratic variation*

$$[X, X](t) = \int_0^t \Delta(u)^2 du.$$

Naturally, the previous definition inspires the integral

$$\int_0^t \Gamma(u)dX(u) := \int_0^t \Gamma(u)\Delta(u)dW(u) + \int_0^t \Gamma(u)\Theta(u)du,$$

for an adapted process Γ and Itô process X.

Finally, we can state the following:

Theorem 2.3.4. (Itô formula) *Let $X(t), t \geq 0$ be an Itô process and let $f(t, x)$ define a function for which partial derivatives f_t, f_x, and f_{xx} are defined and continuous. Then for every $T \geq 0$,*

$$\begin{aligned} f(T, X(T)) =& f(0, X(0)) + \int_0^T f_t(t, X(t))dt + \int_0^T f_x(t, X(t))\Delta(t)dW(t) \\ &+ \int_0^T f_x(t, X(t))\Theta(t)dt + \frac{1}{2}\int_0^T f_{xx}(t, X(t))\Delta(t)^2 dt, \end{aligned}$$

which may be written, for convenience,

$$df(t, X(t)) = f_t(t, X(t))dt + f_x(t, X(t))dX(t) + \frac{1}{2}f_{xx}(t, X(t))dX(t)dX(t).$$

This is often the case for the remainder of this dissertation: technical integrals are written in differential notation for convenience.

The Itô formula permits us to solve a large number of stochastic differential equations and is crucial in a thorough understanding of the BN-S model, which is investigated in further sections,

but for now, let us state a uniqueness and existence theorem for stochastic differential equations from [20]:

Theorem 2.3.5. *The system*

$$dX(t) = \alpha(t, X(t))dt + \sigma(t, X(t))dW(t) + \int_0^t \int_{\mathbb{R}^n} \gamma(s, X(s^-), z)\,(J_X(ds \times dx) - \nu(dx)ds)$$

with $X(0) = x_0 \in \mathbb{R}^n$ *and where* $\alpha : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$, *and* $\gamma : [0,T] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n \times l}$ *and* $\nu = \nu_1 \times \ldots \times \nu_n$ *satisfy the conditions:*

1. *There exists a constant* $C_1 < \infty$ *such that*

$$\|\sigma(t,x)\|^2 + |\alpha(t,x)|^2 + \int_{\mathbb{R}} \sum_{k=1}^{l} |\gamma_k(t,x,z)|^2 \nu_k(dz_k) \leq C_1(1 + |x|^2)$$

   for all $x \in \mathbb{R}^n$.

2. *There exists a constant* $C_2 < \infty$ *such that*

$$\|\sigma(t,x) - \sigma(t,y)\|^2 + |\alpha(t,x) - \alpha(t,y)|^2$$

$$+ \sum_{k=1}^{l} \int_{\mathbb{R}} |\gamma^{(k)}(t,x,z_k) - \gamma^{(k)}(t,y,z_k)|^2 \nu_k(dz_k) \leq C_2|x-y|^2$$

   for all $x, y \in \mathbb{R}^n$,

has a unique cádlág adapted solution $X(t)$ *such that*

$$E[|X(t)|^2] < \infty \text{ for all } t.$$

In the time homogeneous case, when the coefficients do not depend on t, the solutions are called *jump diffusions.*

2.4. Infinitesimal Generators

Jump diffusions sometimes have what are called infinitesimal generators.

Definition 2.4.1. *Let $X(t) \in \mathbb{R}^n$ be a jump diffusion. Then the* infinitesimal generator *A of X is defined on functions $f : \mathbb{R}^n \to \mathbb{R}$ by*

$$Af(x) = \lim_{t \to 0^+} \frac{1}{t} \left(E^x[f(X(t))] - f(x) \right)$$

when the limit exists, and where $E^x[f(X(t))] = E[f(X^{(x)}(t))]$ and $X^{(x)}(0) = x$.

We care particularly about how to find infinitesimal generators when a certain Lévy triplet is known, and [21] gives a very efficient and powerful theorem, which arises from the Itô formula.

Theorem 2.4.2. *Let X be a d-dimensional Lévy process with characteristics (b, Σ, ν) then the infinitesimal generator A of X is, for any bounded and twice continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, and any $x \in \mathbb{R}$,*

$$Af(x) = \sum_i b_i f_{x_i}(x) - \frac{1}{2} \sum_{i,j} f_{x_i x_j}(x) + \int_{\mathbb{R}^d} \left(f(x+y) - f(x) - \sum_i \frac{y_i f_{x_i}(x)}{1 + \|y\|} \right) \nu(dy).$$

2.5. Statistical Background

A crucial concept in statistics is error. Every practical model should allow for variation in its predictions. The errors of type I and type II specifically address errors in classification.

Definition 2.5.1. *We say a particular classification is a* type I error *if it is a rejection of a true null hypothesis; alternatively, a* type II error *is a non-rejection of a false null hypothesis.*

Thus, one might intuitively consider a type I error to be doing a bad thing, while a type II error is failing to do a good thing. In general, each type of error has its own cost for each specific scenario, but often, hypothesis tests are framed in a way to make type I error worse.

Another tool used in this dissertation is the *likelihood function* and its more computationally convenient counterpart the *log-likelihood function*. A likelihood function is a function that measures the goodness of fit of a model with varying parameters to a set of data points. The likelihood function is a function on the space of parameters in the model. Its peaks represent optimal choices of the parameters to fit a model to the data set. On the other hand, the log-likelihood function is simply the natural logarithm of the likelihood function. Often the likelihood function is a ratio, and so, the log-likelihood becomes a difference, which is computationally simpler. Thus, in this

dissertation, we formulate our decision rules based on the log-likelihood functions resulting from our underlying processes.

Finally, we have the idea of *sequential hypothesis tests*. Normally, hypothesis tests have fixed sample size and duration, say, when testing a drug on the effectiveness of curing an illness. Perhaps the study would have a large sample of people with the illness, and the researchers would be interested in the proportion of participants who end up cured after a period of a few months. This sample of people creates a stochastic process, the proportion of people cured. The idea of testing the hypotheses sequentially is that if a significant proportion is cured before the few months is over, the test can be concluded, and the medication can be more quickly distributed to the rest of the population. In particular, one might consider the log-likelihood function of the distributions induced by the hypotheses that the medication does nothing and that the medication works. Then, depending on the error probabilities, which side of an interval the log-likelihood process exits determines the outcome of the sequential hypothesis test. In essence, while a normal hypothesis test only concludes when a predetermined time period has elapsed, a sequential hypothesis test considers new information as it is available and evaluates the hypotheses dynamically to reduce the time requirements of the study. One might call the test optimal if the time requirements are minimized with respect to desired error probabilities. In the next chapter, we construct decision rules for sequential hypothesis tests based on the Bayesian approach. However, due to certain difficulties, we only have optimality results for a few cases.

3. TWO SENSOR HYPOTHESIS TESTING USING INFINITESIMAL GENERATORS

3.1. Sequential Decision Making with Underlying Wiener Processes

The initial inspiration for the majority of this section is [9]. It seeks to devise a two-sensor hypothesis test based upon a two-dimensional Wiener process $z = (z_1, z_2)$ where

$$dz_t^{(k)} = \sigma_k dW_t^{(k)} + \mu_k dt, \quad k = 1, 2,$$

where $W_t^{(k)}$ are Brownian motions with correlation ρ. The hypotheses

$$\begin{aligned} &H_{00} : \mu_1 = 0, \quad \mu_2 = 0, && H_{10} : \mu_1 = m_1 \neq 0, \quad \mu_2 = 0, \\ &H_{01} : \mu_1 = 0, \quad \mu_2 = m_2 \neq 0, && H_{11} : \mu_1 = m_1 \neq 0, \quad \mu_2 = m_2 \neq 0 \end{aligned} \tag{3.1}$$

are tested. Consequently, the analysis presented is applicable for sequential decision making on the state of a two-sensor system with uncorrelated noise. Each sensor receives or does not receive a signal obstructed by said noise. This gives rise to four possibilities, viz. ⟨noise, noise⟩ (denoted by 00), ⟨signal, noise⟩ (denoted by 10), ⟨noise, signal⟩ (denoted by 01), and ⟨signal, signal⟩ (denoted by 11).

The two-dimensional Wiener process generates a filtration, which will be denoted $\mathcal{F}_t$, along with marginal filtrations $\mathcal{F}_t^{(1)}$ and $\mathcal{F}_t^{(2)}$. Further, the hypotheses and diffusion correlation ρ induce probability measures $P_{ij,\rho}$ and marginal probability measures $P_i^{(k)}$. Further, [9] seeks to create an optimal decision rule (τ, δ_τ), where τ is a stopping rule with respect to $\mathcal{F}_t$, and δ_τ is a random variable taking values in the index set $\{00, 01, 10, 11\}$, representing which of the hypotheses to accept. Optimality will be based on minimizing the observation time required for given error probabilities $\alpha_{ij,\rho} := P_{ij,\rho}(\delta_\tau \neq ij)$.

Let the log-likelihood function be

$$u_t^{(i,k)} = \log \frac{dP_i^{(k)}}{dP_{1-i}^{(k)}}. \tag{3.2}$$

We define a rectangle $R := [l_1, r_1] \times [l_2, r_2] \subset \mathbb{R}^2$ and denote

$$\begin{aligned} \tau_k &= \inf\{t \geq 0 : u_t^{(i,k)} \notin [l_k, r_k]\}, \\ \delta_{\tau_k}^{(i,k)} &= 1 - i, \text{ if } u_{\tau_k}^{(i,k)} \leq l_k, \\ \delta_{\tau_k}^{(i,k)} &= i, \text{ if } u_{\tau_k}^{(i,k)} \geq r_k. \end{aligned} \tag{3.3}$$

The decision rule for the two-dimensional test is defined as

$$\begin{aligned} \tau &= \tau_1 \vee \tau_2 \\ \delta_{\tau}^{(i,j)} &= \delta_{\tau_1}^{(i,1)} \delta_{\tau_2}^{(j,2)}, \end{aligned} \tag{3.4}$$

where $\tau_1 \vee \tau_2 = \max\{\tau_1, \tau_2\}$.

The following theorem concerns the infinitesimal generator of the processes $u_t^{(i,k)}$:

Theorem 3.1.1. *Assuming the Brownian motions of the two one-dimensional processes z_1 and z_2 are uncorrelated and with $u_t^{(i,k)}$ defined as in* (3.2)*, we have two-dimensional infinitesimal generators, for the processes $u_t^{i,j} = (u_t^{(i,1)}, u_t^{(j,2)})$,*

$$\mathcal{L}_{ij,\rho} = \frac{m_1^2}{2\sigma_1^2}\left(\partial_{x_1x_1} + (-1)^{i+1}\partial_{x_1}\right) + \frac{m_2^2}{2\sigma_2^2}\left(\partial_{x_2x_2} + (-1)^{j+1}\partial_{x_2}\right) + \rho\frac{m_1 m_2}{\sigma_1\sigma_2}\partial_{x_1x_2}.$$

The proof of this theorem is classically known and does not rely on Lévy characteristics or the powerful Theorem 2.4.2, but of course, using them gives consistent results. While the theorem permits $\rho \neq 0$, the special case $\rho = 0$ is further investigated in [9], and by using the symmetry of the solutions of the partial differential equations $\mathcal{L}_{ij,0}\xi_{ij,0} = 0$ based on the infinitesimal generators, the bounds of the decision rule rectangle are obtained, completing the task of developing a well-defined decision rule. Finally, the optimality of that rule based on minimizing the time required to observe the underlying Wiener process is proved.

Before moving on, we must strictly define what type of optimality we are considering.

Definition 3.1.2. *A decision rule of the form* (3.4) *has* optimality of order 3 *if*

$$E_{ij,0}(\tau_1 \vee \tau_2) - E_{ij,0}(\tau_1) = o(1), \tag{3.5}$$

as the error probabilities $\alpha_{ij,0} \to 0$.

With this, the following theorem summarizes the main results from [9]:

Theorem 3.1.3. *The decision rule* (3.4) *when applied to hypotheses* (3.1) *has optimality of order* 3*, and choosing three of the four values for Type I errors* $\alpha_{ij,0}$ *induces the value of the fourth by*

$$(1-\alpha_{00,0})(1-\alpha_{11,0}) = (1-\alpha_{01,0})(1-\alpha_{10,0}).$$

Then, setting the value of l_1 *will determine the other three, fully defining the decision rule through the system*

$$\ln\left(\frac{\alpha_{10,0}-\alpha_{00,0}}{1-\alpha_{00,0}}\right) < l_1 < \ln\left(\frac{\alpha_{10,0}}{1-\alpha_{00,0}}\right),$$

$$r_1 = -\ln\left(1-\frac{1-e^{l_1}}{(1-\alpha_{10,0})/(1-\alpha_{00,0})}\right),$$

$$e^{r_1} = \frac{(1-\alpha_{10,0})/(1-\alpha_{00,0})}{\frac{e^{r_2}}{(1-\alpha_{01,0})+(1-\alpha_{00,0}e^{r_2})}-1},$$

$$r_2 = -\ln\left(1-\frac{1-e^{l_2}}{(1-\alpha_{01,0})/(1-\alpha_{00,0})}\right).$$

The paper [9] further analyzes briefly the cases when $\rho \neq 0$, relating these cases to the uncorrelated case by considering

$$\Delta\xi_{ij,\rho} := \xi_{ij,\rho} - \xi_{ij,0}.$$

In particular, it is shown that if $\rho > 0$, with $ij \in \{00, 11\}$ or $\rho < 0$, with $ij \in \{01, 10\}$, we have that $\Delta\xi_{ij,\rho}(u^{(1)}_{t\wedge\tau_1}, u^{(2)}_{t\wedge\tau_1})$ is a supermartingale, and in the other cases, it is a submartingale.

This ultimately gives us a result that can be condensed into a theorem:

Theorem 3.1.4. *When the correlation between the Brownian motions* ρ *is not* 0*, the stopping times defined in* (3.3) *have upper bounds based on the stopping times in the cases where* $|\rho| = 1$*. The* $\{00\}$ *case with* $\rho = 1$ *and* $\{01\}$ *case with* $\rho = -1$ *have expected stopping times at most* $E_0^{(x=0)(1)}(\tau_1)$ *while the* $\{11\}$ *case with* $\rho = 1$ *and* $\{10\}$ *case with* $\rho = -1$ *have expected stopping times at most* $E_0^{(x=0)(1)}(\tau_1)$*. These are bounds also in the* $|\rho| < 1$ *cases.*

While Theorem 3.1.3 provides an optimal decision rule for Wiener processes, often we experience processes with jump terms as well. We seek to generalize the result in [9] to a certain class of underlying Lévy processes.

3.2. Drift Test Generalization

For various financial time series data, *jumps* play an important role. Jumps in a stochastic model are typically captured by a Lévy process. In this section, we generalize the analysis presented in [9]. We present the analysis for the case when the signals are driven not only by a Brownian motion, but by a generalized Lévy process.

Consider $z = (z_1, z_2)$ a two-dimensional Lévy process defined by Lévy triplet (μ, Σ, ν^*), where $\mu = [\mu_1, \mu_2]$ is the two-dimensional drift, $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}$ is a symmetric non-negative definite matrix representing the diffusion, and ν^* is a two-dimensional Lévy measure defined by a product of two identical one-dimensional Lévy measures ν. Under this setting, we wish to test the hypotheses

$$\begin{aligned} &H_{00} : \mu_1 = 0, \quad \mu_2 = 0, && H_{10} : \mu_1 = m_1 \neq 0, \quad \mu_2 = 0, \\ &H_{01} : \mu_1 = 0, \quad \mu_2 = m_2 \neq 0, && H_{11} : \mu_1 = m_1 \neq 0, \quad \mu_2 = m_2 \neq 0. \end{aligned} \tag{3.6}$$

Note that these hypotheses strictly address the drift terms of the Lévy processes. The primary difference is that we include a Lévy measure, despite its not changing based on the hypotheses.

The Lévy process generates a filtration, which will be denoted $\mathcal{F}_t$, along with marginal filtrations $\mathcal{F}_t^{(1)}$ and $\mathcal{F}_t^{(2)}$. Further, the hypotheses and diffusion correlation ρ induce probability measures $P_{ij,\rho}$ and marginal probability measures $P_i^{(k)}$. We seek to create an optimal decision rule (τ, δ_τ), where τ is a stopping rule with respect to $\mathcal{F}_t$ and δ_τ is a random variable taking values in the index set $\{00, 01, 10, 11\}$. Optimality will be based on minimizing the observation time required for given error probabilities $\alpha_{ij,\rho} := P_{ij,\rho}(\delta_\tau \neq ij)$.

Using the same notation as in (3.2), (3.3), and (3.4), of course with different underlying processes, we have the following theorem concerning the infinitesimal generator of $u_t^{(i,k)}$:

Theorem 3.2.1. *Assuming the Brownian motions of the two one-dimensional processes z_1 and z_2 are uncorrelated, we have two-dimensional infinitesimal generators for the processes $u_t^{i,j} =$*

$(u_t^{(i,1)}, u_t^{(j,2)})$,

$$\mathcal{L}_{ij,0} = \frac{m_1^2}{2\sigma_1^2}\left(\partial_{x_1x_1} + (-1)^{i+1}\partial_{x_1}\right) + \frac{m_2^2}{2\sigma_2^2}\left(\partial_{x_2x_2} + (-1)^{j+1}\partial_{x_2}\right).$$

Proof. Since z_k is a Lévy process with characteristics $(0, \sigma_k^2, \nu)$ under $P_0^{(k)}$ and characteristics (m_k, σ_k^2, ν) under $P_1^{(k)}$, we can apply a generalized Girsanov's Theorem (see [33] Theorem 1.20). Using $\beta = (-1)^{i+1}m_k/\sigma_k$, we obtain

$$\frac{dP_i^{(k)}}{dP_{1-i}^{(k)}} = \mathcal{E}\left((-1)^{i+1}\frac{m_k}{\sigma_k}W.\right)_t,$$

where W is a standard Brownian motion. Further, by [12] (Proposition 8), we obtain characteristics $((-1)^i\frac{m_k^2}{2\sigma_k^2}, \frac{m_k^2}{\sigma_k^2}, 0)$ for $u_t^{(i,k)}$. Then the process $\left(u_t^{(i,1)}, u_t^{(j,2)}\right)$ is well-known to have generator

$$\mathcal{L}_{ij,0} = (-1)^{i+1}\frac{m_1^2}{2\sigma_1^2}\partial_{x_1} + (-1)^{j+1}\frac{m_2^2}{2\sigma_2^2}\partial_{x_2} + \frac{m_1^2}{2\sigma_1^2}\partial_{x_1x_1} + \frac{m_2^2}{2\sigma_2^2}\partial_{x_2x_2},$$

as claimed. □

These infinitesimal generators are then used to determine the bounds of the rectangle by applying them to the likelihood functions $\psi_{ij} : R \to [0, 1]$ which represents the probability of being in world ij at any position inside the rectangle. When in the correct world, the likelihood function does not change with respect to that world's generator: i.e., $\mathcal{L}_{ij,0}\psi_{ij} = 0$.

Next we present a theorem of optimality. The proof of this theorem follows directly from the method in [9].

Theorem 3.2.2. *The decision rule for* (3.6) *defined in* (3.4) *has asymptotic optimality of order-3.*

Proof. The key tool in this proof is the exponential killing trick; i.e., for any nonnegative random variable Y with no point mass at zero,

$$E(e^{-\lambda Y}) = E(F_Y(X_\lambda)),$$

where F_Y is the cumulative distribution function for Y and X_λ is an exponential random variable with parameter λ. We use this on τ_1 and τ_2 from (3.4):

$$\begin{aligned} E_{ij,0}(e^{-\lambda\tau_1}) &= E_{ij,0}(F_{\tau_1}(X_\lambda)), \\ E_{ij,0}\left(e^{-\lambda(\tau_1\vee\tau_2)}\right) &= E_{ij,0}(F_{\tau_1}(X_\lambda)F_{\tau_2}(X_\lambda)), \end{aligned}$$

as τ_1 and τ_2 are independent when $\rho = 0$. Applying Laplace transforms, simplifying, and applying exponential killing in (3.5), we find

$$E_{ij,0}(\tau_1 \vee \tau_2) - E_{ij}(\tau_1) = \lim_{\lambda\to 0} \frac{E_{ij,0}(F_{\tau_1}(X_\lambda)(1 - F_{\tau_2}(X_\lambda)))}{\lambda}. \tag{3.7}$$

Because our log-likelihood processes are Brownian motions with drift, and so their exit times are finite, this difference on the left is finite. Define

$$0 < A_k := e^{l_k} < 1 < B_k := e^{r_k} < \infty,$$

and

$$C_1 := \frac{1 - \alpha_{10,0}}{1 - \alpha_{00,0}} \to 1 \text{ as } \alpha_{ij,0} \to 0,$$

$$C_2 := \frac{1 - \alpha_{11,0}}{1 - \alpha_{10,0}} \to 1 \text{ as } \alpha_{ij,0} \to 0.$$

Due to this and 3.1.3, $l_k = -r_k$ as $\alpha_{ij,0} \to 0$.

Further, also by 3.1.3,

$$\frac{C_1 + B_1}{B_1} \cdot \frac{C_2 + B_2}{B_2} = \frac{1}{1 - \alpha_{00,0}} \to 1 \text{ as } \alpha_{ij,0} \to 0.$$

Because $B_k > 1$, we see that $r_k \to \infty$ as $\alpha_{ij,0} \to 0$, and so, $l_k \to -\infty$. Hence, $F_{\tau_k} \to 0$ as well. Finally, applying the dominated convergence theorem,

$$\lim_{\alpha_{ij,0}\to 0} \lim_{\lambda\to 0} \frac{E_{ij,0}(F_{\tau_1}(X_\lambda)(1 - F_{\tau_2}(X_\lambda)))}{\lambda} = \lim_{\lambda\to 0} \lim_{\alpha_{ij,0}\to 0} \frac{E_{ij,0}(F_{\tau_1}(X_\lambda)(1 - F_{\tau_2}(X_\lambda)))}{\lambda} = 0,$$

which yields the desired asymptotic optimality limit. □

3.3. Hypothesis Tests on the Lévy Measure

In this section, we expand the idea presented in the last section. Before presenting the analysis, we briefly introduce a possible application of this work. A commonly used stochastic model for the derivative market analysis is the BN-S model (see [3, 6, 5, 4, 19, 30]). The BN-S model is also implemented in the commodity market (see [30, 29]). Though this model is very efficient and simple to use, it suffers from the absence of a long-range dependence and many other issues. Mathematically, for the BN-S model, the stock or commodity price $S = (S_t)_{t\geq 0}$ on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ is modeled by

$$S_t = S_0 \exp(X_t), \tag{3.8}$$

$$dX_t = (\mu + \beta\sigma_t^2)\, dt + \sigma_t\, dW_t + \rho\, dZ_{\lambda t}, \tag{3.9}$$

$$d\sigma_t^2 = -\lambda\sigma_t^2\, dt + dZ_{\lambda t}, \quad \sigma_0^2 > 0, \tag{3.10}$$

where the parameters $\mu, \beta, \rho, \lambda \in \mathbb{R}$ with $\lambda > 0$ and $\rho \leq 0$ and r is the risk-free interest rate where a stock or commodity is traded up to a fixed horizon date T. In this model, W_t is a Brownian motion, and the process Z_t is a subordinator. Also W_t and Z_t are assumed to be independent, and $(\mathcal{F}_t)$ is assumed to be the usual augmentation of the filtration generated by the pair (W_t, Z_t).

In a recent work [29], it is shown that for various derivative and commodity price dynamics, the jump is *not* completely stochastic. On the contrary, there is a *deterministic* element in crude oil price that can be implemented in the existing models for an extended period of time. It may be shown that the dynamics of X_t in (3.9) can be more accurately written when we use a convex combination of two independent subordinators, Z and $Z^{(b)}$ as:

$$dX_t = (\mu + \beta\sigma_t^2)\, dt + \sigma_t\, dW_t + \rho\left(\theta\, dZ_{\lambda t} + (1-\theta) dZ_{\lambda t}^{(b)}\right), \tag{3.11}$$

where $\theta \in [0,1]$ is a *deterministic* parameter. The process $Z^{(b)}$ in (3.11) is a subordinator that has greater intensity than the subordinator Z. In this case (3.10) will be given by

$$d\sigma_t^2 = -\lambda\sigma_t^2\, dt + \theta' dZ_{\lambda t} + (1-\theta')dZ_{\lambda t}^{(b)}, \quad \sigma_0^2 > 0, \tag{3.12}$$

where, as before, $\theta' \in [0,1]$ is *deterministic*.

We observe that even for commonly implemented stochastic models, it is important to detect when a "smaller" fluctuation (Z) turns into a "larger" fluctuation ($Z^{(b)}$). Consequently, it is important to determine a method for sequential testing of the jump size distribution. The advantages of the dynamics given by the refined BN-S model, given by (3.8), (3.11), and (3.12), over existing models are significant. This minor change in the model incorporates *long-range dependence* without actually changing the model.

With this in mind, we consider $z = (z_1, z_2)$ a two-dimensional Lévy process defined by Lévy triplet (μ, Σ, ν^*), where $\mu = [\mu_1, \mu_2]$ is the two-dimensional drift, $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}$ is a symmetric non-negative definite matrix representing the diffusion, and ν^* is a two-dimensional Lévy measure defined by a product of two one-dimensional Lévy measures ν_1 and ν_2 with densities $\nu_k(dx) = (1+\alpha_k x)\nu(dx)$ for some Lévy measure ν defined on $\mathbb{R}^+$.

We wish to test the hypotheses

$$\begin{aligned} &H_{00} : \alpha_1 = 0, \quad \alpha_2 = 0, && H_{10} : \alpha_1 = a_1 > 0, \quad a_2 = 0, \\ &H_{01} : a_1 = 0, \quad \alpha_1 = a_2 > 0, && H_{11} : \alpha_1 = a_1 > 0, \quad \alpha_1 = a_2 > 0. \end{aligned} \tag{3.13}$$

These now address the size of the jumps in the Lévy process.

Similar to the previous section, the Lévy process generates a filtration, which will be denoted $\mathcal{F}_t$, along with marginal filtrations $\mathcal{F}_t^{(1)}$ and $\mathcal{F}_t^{(2)}$. Further, the hypotheses and diffusion correlation ρ induce probability measures $P_{ij,\rho}$ and marginal probability measures $P_i^{(k)}$. We seek to create optimal decision rules (τ, δ_τ), where τ is a stopping rule with respect to $\mathcal{F}_t$ and δ_τ is a random variable taking values in the index set $\{00, 01, 10, 11\}$.

Let $u_t^{(i,k)}$ be defined as in (3.2) and still consider a rectangle $[l_1, r_1] \times [l_2, r_2] \subset \mathbb{R}^2$. The decision rules for the non-correlated one-dimensional cases are as in (3.3), with the combined decision rule in (3.4).

It is known that if $(X_t)_{t\geq 0}$ is a Lévy process then there exists a unique cádlág process $(Z_t)_{t\geq 0}$ such that

$$dZ_t = Z_{t-}\, dX_t, \quad Z_0 = 1.$$

Z is called the stochastic exponential or Doléans-Dade exponential of X and is denoted by $Z = \mathcal{E}(X)$. We can now derive the infinitesimal generators:

Theorem 3.3.1. *With the process $u_t^{(i,k)}$ defined as in (3.2), we have two-dimensional infinitesimal generators, for the process $u_t^{i,j} = (u_t^{(i,1)}, u_t^{(j,2)})$, defined by*

$$\mathcal{L}_{ij,\rho}\xi(x) = (-1)^{i+1}\gamma_1\xi_{x_1}(x) + (-1)^{j+1}\gamma_2\xi_{x_2}(x) + \frac{1}{2}\beta_1^2\xi_{x_1x_1}(x) + \frac{1}{2}\beta_2^2\xi_{x_2x_2}(x) + \rho\beta_1\beta_2\xi_{x_1x_2}(x)$$

$$+(-1)^{i+j}\int_{\mathbb{R}_+^2}\left(\xi(x+y) - \xi(x) - \frac{y_1\xi_{x_1}(x) + y_2\xi_{x_2}(x)}{1+\|y\|}\right)K_1(dy_1)K_2(dy_2),$$

for any suitable ξ, where

$$x = (x_1, x_2), \quad y = (y_1, y_2)$$

$$\beta_k = -a_k\int_{x_k>0}(1\wedge x_k)\sigma_k^{-1}x_k\nu(dx_k) \tag{3.14}$$

$$m_k = a_k\int_{x_k>1}x_k\nu(dx_k) \tag{3.15}$$

$$\gamma_k = m_k - \frac{\beta_k^2}{2} + \int_0^1(\log(1+x_k)^2 - x_k)a_k\nu(dx_k). \tag{3.16}$$

$$K_k = a_k\log(1+x_k)\nu_k. \tag{3.17}$$

Proof. Since z_k is a Lévy process with characteristics $(\mu_k, \sigma_k^2, \nu_k)$ under $P_0^{(k)}$ and characteristics $(\mu_k, \sigma_k^2, (1+a_kx_k)\nu_k)$ under $P_1^{(k)}$, we apply the generalized Girsanov's Theorem. Using β_k as in (3.14), we obtain

$$\frac{dP_i^{(k)}}{dP_{1-i}^{(k)}} = \mathcal{E}\left((-1)^{i+1}N^{(k)}.\right)_t,$$

where

$$N_t^{(k)} = \beta_k W_t + \int_0^t \int_{x_k>0} a_k x_k (J_k - \nu_k)(ds, dx_k),$$

$a_k J_k$ is the jump measure for N, W is a standard Brownian motion, and $\mathcal{E}$ is the Doléans-Dade exponential. This gives that $N_t^{(k)}$ is a Lévy process with characteristics

$$((-1)^i m_k, \beta_k^2, (-1)^{i+1} a_k \nu_k).$$

Then, by [12] (Proposition 8), we obtain characteristics

$$((-1)^i \gamma_k, \beta_k^2, (-1)^{i+1} K_k)$$

for $u_t^{(i,k)}$. Finally, by [12] and Theorem 2.4.2, the process $\left(u_t^{(i,1)}, u_t^{(j,2)}\right)$ has the stated generator. □

Assign $\xi_{ij,\rho}$ to be the probability of a correct decision in world ij. Then we have the partial integro-differential equation $\mathcal{L}_{ij,\rho}\xi_{ij,\rho} = 0$ with boundary conditions

$$\begin{aligned}
\xi_{00,\rho}(l_1, l_2) &= 1, & \xi_{01,\rho}(l_1, r_2) &= 1, \\
\xi_{00,\rho}(r_1, y) &= 0, & \xi_{01,\rho}(x, l_2) &= 0, \\
\xi_{00,\rho}(x, r_2) &= 0, & \xi_{01,\rho}(r_1, y) &= 0, \\
\xi_{10,\rho}(r_1, l_2) &= 1, & \xi_{11,\rho}(r_1, r_2) &= 1, \\
\xi_{10,\rho}(l_1, y) &= 0, & \xi_{11,\rho}(l_1, y) &= 0, \\
\xi_{10,\rho}(x, r_2) &= 0, & \xi_{11,\rho}(x, l_2) &= 0,
\end{aligned} \tag{3.18}$$

for $x \in [l_1, r_1]$ and $y \in [l_2, r_2]$. Further, we have $\xi_{ij,\rho} > 0$ inside $R = (l_1, r_1) \times (l_2, r_2)$.

Before proceeding to prove the existence of a solution to such a boundary value problem, we state the following definitions and a theorem from [2] that will be used.

Definition 3.3.2. *An upper semicontinuous function $l : \mathbb{R}^2 \to \mathbb{R}$ is a* subsolution *of*

$$F(0, \xi, D\xi, D\xi^2, \mathcal{I}[\xi](x)) = 0$$

subject to boundary conditions (4.14) *if for any test function* $\phi \in C^2(\mathbb{R}^2)$*, at each maximum point* $x_0 \in \bar{R}$ *of* $l - \phi$ *in* $B_\delta(x_0)$*, we have*

$$E(l, \phi, x_0) := F(x_0, l(x_0), D\phi(x_0), D^2\phi(x_0), I_\delta^1[\phi](x_0) + I_\delta^2[l](x_0)) \leq 0 \text{ if } x_0 \in R$$

or

$$\min(E(l, \phi, x_0); u(x_0) - g(x_0)) \leq 0 \text{ if } x_0 \in \partial R,$$

where

$$I_\delta^1[\phi](x_0) = \int_{|z|<\delta} (\phi(x_0 + z) - \phi(x_0) - (D\phi(x_0) \cdot z)\mathbf{1}_B(z))\, d\mu_{x_0}(z),$$
$$I_\delta^2[u](x_0) = \int_{|z|\geq\delta} (u(x_0 + z) - u(x_0) - (D\phi(x_0) \cdot z)\mathbf{1}_B(z))\, d\mu_{x_0}(z).$$

Similarly, a lower semicontinuous function $u : \mathbb{R}^2 \to \mathbb{R}$ *is a* supersolution *of the same boundary value problem if for any test function* $\phi \in C^2(\mathbb{R}^2)$*, at each minimum point* $x_0 \in \bar{R}$ *of* $u - \phi$ *in* $B_\delta(x_0)$*, we have*

$$E(u, \phi, x_0) \geq 0 \text{ if } x_0 \in R$$

or

$$\max(E(l, \phi, x_0); u(x_0) - g(x_0)) \leq 0 \text{ if } x_0 \in \partial R.$$

Finally, a viscosity solution *is a function whose upper and lower semicontinuous envelopes are respectively a* sub-solution *and a* super-solution*.*

Theorem 3.3.3. *If* $F : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}^2 \times \mathbb{S}^2 \times \mathbb{R} \to \mathbb{R}$*, where* $\mathbb{S}^n$ *is the space of* $n \times n$ *symmetric matrices, and*

(A1) $F(x, u, p, X, i_1) \leq F(x, u, p, Y, i_2)$ *if* $X \geq Y$ *and* $i_1 \geq i_2$*,*

(A2) there exists $\gamma > 0$ *such that for any* $x \in \mathbb{R}^2$*,* $u, v \in \mathbb{R}$*,* $p \in \mathbb{R}^2$*,* $X \in \mathbb{S}^2$*, and* $i \in \mathbb{R}$*,*

$$F(x, u, p, X, i) - F(x, v, p, X, i) \geq \gamma(u - v) \text{ if } u \geq v,$$

for some $\epsilon > 0$ *and* $r(\beta) \to 0$ *as* $\beta \to 0$*, we have*

$$F(y, v, \epsilon^{-1}(x-y), Y, i) - F(x, v, \epsilon^{-1}(x-y), X, i) \leq \omega_R(\epsilon^{-1}|x-y|^2 + |x-y| + r(\beta)),$$

(A3) F *is uniformly continuous with respect to all arguments,*

(A4) $\sup_{x\in\mathbb{R}} |F(x, 0, 0, 0, 0)| < \infty$*,*

(A5) $K = K_1 \times K_2$ *is a Lévy-Itô measure,*

(A6) the inequalities in (3.19) *are strict,*

(A7) for any $R > 0$*, there exists a modulus of continuity* ω_R *such that, for any* $x, y \in \mathbb{R}^2$*,* $|v| \leq R$*,* $i \in \mathbb{R}$*, and for any* $X, Y \in \mathbb{S}^2$ *satisfying*

$$\begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \leq \frac{1}{\epsilon} \begin{bmatrix} I & -I \\ -I & I \end{bmatrix} + r(\beta) \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix},$$

then there is a unique solution to $F(0, \xi, D\xi, D^2\xi, \mathcal{I}[\xi](x)) = 0$ *between any pair of super-solution and sub-solutions, where*

$$\mathcal{I}[\xi](x) := \int_{\mathbb{R}^2_+} \left(\xi(x+y) - \frac{y_1\xi_{x_1}(x) + y_2\xi_{x_2}(x)}{1 + \|y\|} \right) K_1(dy_1)K_2(dy_2).$$

Lemma 3.3.4. *In particular, our function*

$$F(x, u, p, X, i) := Mu + \langle \gamma_1, \gamma_2 \rangle \cdot p - \mathrm{Tr}\left(\begin{bmatrix} \beta_1/2 & 0 \\ 0 & \beta_2/2 \end{bmatrix} X \right) - i$$

satisfies (A1)-(A4) and our measure K *satisfies (A5) in* (3.3.3)*, where*

$$M = \int_{\mathbb{R}^2_+} K_1(dy_1)K_2(dy_2).$$

Proof. First, consider (A1):

$$F(x,u,p,X,i_1) - F(x,u,p,Y,i_2) = \mathrm{Tr}\left(\begin{bmatrix} \beta_1/2 & 0 \\ 0 & \beta_2/2 \end{bmatrix}(Y-X)\right) + i_2 - i_1 \geq 0$$

if $i_2 \leq i_1$ and $Y \leq X$.

Next, $F(x,u,p,X,i) - F(x,v,p,X,i) = M(u-v)$, so choosing $\gamma = M > 0$, we have property (A2).

Property (A3) is satisfied because F is linear in each argument, and (A4) is satisfied because F does not depend on its first argument explicitly. Last, K is a Lévy-Itô measure by the assumptions of the underlying Lévy processes.

□

Note that the F above corresponds to case $ij = 00$. The other three cases can be similarly satisfied through manipulation of the signs in F. Before proceeding, we present a few formal definitions:

Definition 3.3.5. *We write that a function $f(x) = O(g(x))$ if we have some $M, \epsilon \in \mathbb{R}$ satisfying $|f(x)| \leq Mg(x)$ for all $x > \epsilon$. Similarly, we write that a function $f(x) = o(g(x))$ if for every $M > 0$, there exists $\epsilon \in \mathbb{R}$ satisfying $|f(x)| \leq Mg(x)$ for all $x > \epsilon$.*

The norm $\|f\|_\infty$ is defined as the essential supremum of the absolute value of f over Ω. It is the smallest number so that $\{x : |f(x)| \geq \|f\|_\infty\}$ has measure zero.

We now state the additional limit assumptions on F from [2]:

$$\begin{aligned} \liminf_{y\to x, y\in\bar{\Omega}, \eta\downarrow 0, d(y)\eta^{-1}\to 0} \left[\sup_{0<\delta\in[d(y),r)} \inf_{s\in[-R,R]} F(y,s,p_\eta(y),M_\eta(y),I_{\eta,\delta,r}(y))\right] < 0, \\ \limsup_{y\to x, y\in\bar{\Omega}, \eta\downarrow 0, d(y)\eta^{-1}\to 0} \left[\inf_{0<\delta\in[d(y),r)} \sup_{s\in[-R,R]} F(y,s,-p_\eta(y),-M_\eta(y),-I_{\eta,\delta,r}(y))\right] < 0, \end{aligned} \tag{3.19}$$

where

$$p_\eta(y) = O(\epsilon^{-1}) + \frac{k_1 + o(1)}{\eta} Dd(y),$$

$$M_\eta(y) = O(\epsilon^{-1}) + \frac{k_1 + o(1)}{\eta} D^2 d(y) - \frac{k_2 + o(1)}{\eta^2} Dd(y) \otimes Dd(y),$$

$$
\begin{aligned}
I_{\eta,\delta,r}(y) = & -\nu I^{\mathrm{ext},1}_{\delta,r}(y) + 2\|u\|_\infty I^{\mathrm{int},1}_{\beta(\nu),r}(y) \\
& - \frac{k_1 + o(1)}{\eta}\left(I^{\mathrm{tr}}(y) + I^{\mathrm{int},2}_{\beta(\eta),r}(y) + I^{\mathrm{ext},2}_{\delta,r}(y) - \|D^2 d\|_\infty I^4_{\delta,\beta(\eta),r}(y)\right) \\
& + O(\epsilon^{-1})\left(1 + o(1) I^{\mathrm{int},3}_{\beta(\eta),r}(y) + o(1) I^{\mathrm{ext},3}_{\delta,r}(y)\right),
\end{aligned}
$$

with $O(\epsilon^{-1})$ not depending on k_1 nor k_2, and

$$\mathcal{A}_{\delta,\beta,r}(x) := \{z \in B_r : -\delta \le d(x+z) - d(x) \le \beta\},$$

$$\mathcal{A}^{\mathrm{ext}}_{\delta,r}(x) := \{z \in B_r : d(x+z) - d(x) < -\delta\},$$

$$\mathcal{A}^{\mathrm{int}}_{\beta,r}(x) := \{z \in B_r : d(x+z) - d(x) > \beta\},$$

$$I^{\mathrm{ext},1}_{\delta,r}(x) := \int_{\mathcal{A}^{\mathrm{ext}}_{\delta,r}(x)} d\mu_x(z),$$

$$I^{\mathrm{ext},2}_{\delta,r}(x) := \int_{\mathcal{A}^{\mathrm{ext}}_{\delta,r}(x)} Dd(x) \cdot z d\mu_x(z),$$

$$I^{\mathrm{ext},3}_{\delta,r}(x) := \int_{\mathcal{A}^{\mathrm{ext}}_{\delta,r}(x)} |z| d\mu_x(z),$$

$$I^{\mathrm{int},1}_{\beta,r}(x) := \int_{\mathcal{A}^{\mathrm{ext}}_{\beta,r}(x)} d\mu_x(z),$$

$$I^{\mathrm{int},2}_{\beta,r}(x) := \int_{\mathcal{A}^{\mathrm{ext}}_{\beta,r}(x)} Dd(x) \cdot z d\mu_x(z),$$

$$I^{\mathrm{int},3}_{\beta,r}(x) := \int_{\mathcal{A}^{\mathrm{ext}}_{\beta,r}(x)} |z| d\mu_x(z),$$

$$I^4_{\delta,\beta,r}(x) := \frac{1}{2}\int_{\mathcal{A}_{\delta,\beta,r}(x)} |z|^2 d\mu_x(z),$$

$$I^{\mathrm{tr}}(x) := \int_{r<|z|<1} Dd(x) \cdot z d\mu_x(z).$$

Theorem 3.3.6. *The partial integro-differential equation $\mathcal{L}_{ij,0}\xi_{ij,0} = 0$ subject to boundary conditions* (4.14) *and $0 < \xi_{ij,\rho}$ has a viscosity solution between sub-solution and super-solution*

$$L_{ij}(x,y) = E_{ij}\frac{\sinh\left(A_{(1,2-i)}\sqrt{B^2_{ij(1)}+L}\right)\sinh\left(A_{(2,2-j)}\sqrt{B^2_{ij(2)}+L}\right)}{\sinh\left(\frac{r_1-l_1}{\beta_1}\sqrt{B^2_{ij(1)}+L}\right)\sinh\left(\frac{r_2-l_2}{\beta_2}\sqrt{B^2_{ij(2)}+L}\right)},$$
$$U_{ij}(x,y) = E_{ij}\frac{\sinh\left(A_{(1,2-i)}\sqrt{B^2_{ij(1)}-L}\right)\sinh\left(A_{(2,2-j)}\sqrt{B^2_{ij(2)}-L}\right)}{\sinh\left(\frac{r_1-l_1}{\beta_1}\sqrt{B^2_{ij(1)}-L}\right)\sinh\left(\frac{r_2-l_2}{\beta_2}\sqrt{B^2_{ij(2)}-L}\right)},$$

where

$$A = \begin{bmatrix} \frac{x-l_1}{\beta_1} & \frac{r_1-x}{\beta_1} \\ \frac{y-l_2}{\beta_2} & \frac{r_2-y}{\beta_2} \end{bmatrix},$$
$$B_{ij} = \begin{bmatrix} \frac{\gamma_1+(-1)^jC}{\beta_1} & \frac{\gamma_2+(-1)^iC}{\beta_2} \end{bmatrix},$$
$$C = \int_{\mathbb{R}^2_+}\frac{y_i}{1+\|y\|}K_1(dy_1)K_2(dy_2),$$
$$E_{ij} = \exp\left(A_{(1,i+1)}B_{ij(1)} + A_{(2,j+1)}B_{ij(2)}\right),$$

and L is a positive constant, provided (A6) and (A7) are satisfied.

Proof. We define

$$H(x) = \int_{\mathbb{R}^2_+}\xi(x+y)K_1(dy_1)K_2(dy_2),$$
$$\xi(x) = f(x_1)g(x_2).$$

Consequently,

$$\begin{aligned} 0 = \mathcal{L}_{ij,0}\xi(x) =& (-1)^{i+1}\gamma_1\xi_{x_1}(x) + (-1)^{j+1}\gamma_2\xi_{x_2}(x) + \frac{1}{2}\beta_1^2\xi_{x_1x_1}(x) + \frac{1}{2}\beta_2^2\xi_{x_2x_2}(x) \\ &+ (-1)^{i+j}\int_{\mathbb{R}^2_+}\left(\xi(x+y) - \xi(x) - \frac{y_1\xi_{x_1}(x) + y_2\xi_{x_2}(x)}{1+\|y\|}\right)K_1(dy_1)K_2(dy_2) \end{aligned}$$

can be rewritten as

$$\begin{aligned}0 =&(-1)^{i+1}\gamma_1 f'(x_1)g(x_2) + (-1)^{j+1}\gamma_2 f(x_1)g'(x_2) + \frac{1}{2}\beta_1^2 f''(x_1)g(x_2) + \frac{1}{2}\beta_2^2 f(x_1)g''(x_2)\\ &+ (-1)^{i+j}H(x) + (-1)^{i+j+1}Mf(x_1)g(x_2)\\ &+ (-1)^{i+j+1}Cf'(x_1)g(x_2) + (-1)^{i+j+1}Cf(x_1)g'(x_2).\end{aligned}$$

When $ij \in \{00, 11\}$, the sign on H is positive; therefore, we have sub-solution equations, through separation of variables,

$$\begin{aligned}\frac{1}{2}\beta_1^2\frac{f''(x_1)}{f(x_2)} + \left((-1)^{i+j+1}C + (-1)^{i+1}\gamma_1\right)\frac{f'(x_1)}{f(x_1)} &= -\lambda_{1,ij} + (-1)^{i+j}M,\\ \frac{1}{2}\beta_2^2\frac{g''(x_2)}{g(x_2)} + \left((-1)^{i+j+1}C + (-1)^{j+1}\gamma_2\right)\frac{g'(x_2)}{g(x_2)} &= \lambda_{1,ij}.\end{aligned}$$

Alternatively, when $ij \in \{01, 10\}$, we have these as super-solution equations.

On the other hand, since $\xi > 0$ inside R, there exists some $K > 0$ so that

$$\xi(x+y) - \xi(x) \le K\xi(x) \iff H(x) - \int \xi(x)K_1(dy_1)K_2(dy_2) \le KMf(x_1)g(x_2).$$

Using this, in cases $ij \in \{00, 11\}$, we have super-solution equations

$$\begin{aligned}\frac{1}{2}\beta_1^2\frac{f''(x_1)}{f(x_2)} + \left((-1)^{i+j+1}C + (-1)^{i+1}\gamma_1\right)\frac{f'(x_1)}{f(x_1)} &= -\lambda_{2,ij} + (-1)^{i+j+1}KM,\\ \frac{1}{2}\beta_2^2\frac{g''(x_2)}{g(x_2)} + \left((-1)^{i+j+1}C + (-1)^{j+1}\gamma_2\right)\frac{g'(x_2)}{g(x_2)} &= \lambda_{2,ij}.\end{aligned}$$

When $ij \in \{01, 10\}$, we have these as sub-solution equations instead.

Now, choosing $L = \max\{KM, M\}$ and $\lambda_{k,ij}$ to yield $\pm L/2$ on the right-hand sides, the boundary-value problem gives the super-solution and sub-solutions claimed. Due to the monotonicity of sinh and exp, we see that the super-solution and sub-solutions are ordered so that $U_{ij} \ge L_{ij} \ge 0$. Finally, by [2], we have existence of a viscosity solution to $\mathcal{L}_{ij,\rho}\xi_{ij,\rho} = 0$ with boundary conditions (4.14). □

Remark 3.3.7. *Note that due to the structure of U_{ij} and $L_{i,j}$, as $L/B_{ij}^2 \to 0$, the super-solution and sub-solutions tend toward each other. While this cannot occur precisely, it grants a particularly interesting condition that can reduce the size of the rectangle used in the decision rule.*

In the following figures (Figures 1, 2, and 3), we plot a super-solution and a sub-solution for a special case. Figure 1 depicts the monotonic nature of U_{00} and L_{00}, and shows the boundary conditions are met. Note that the sides of the rectangle, and therefore the bounds of the domain of x here, would be chosen in such a way as to have $1-\alpha_{00}$ between the graphs along the line $x = 0$ in Figure 2. Finally, Figure 3 fully shows the super-solution and sub-solution of a viscosity solution.

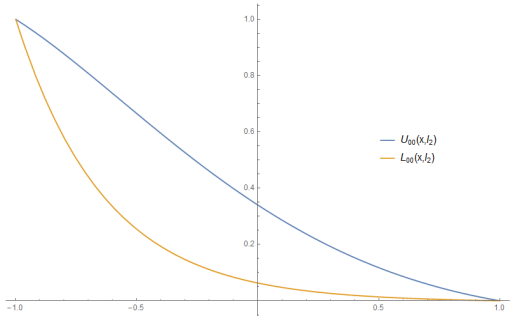

Figure 3.1. The $y = l_2 = -1$ cross-section, with boundary conditions met.

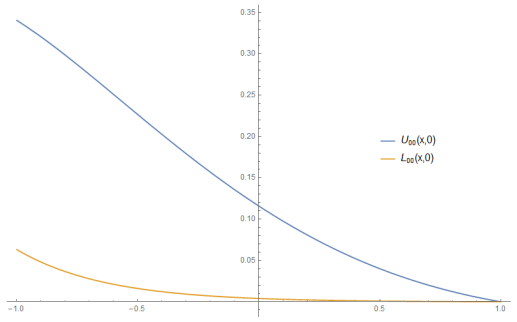

Figure 3.2. The $y = 0$ cross-section, showing the region that determines the actual bounds of R.

The super-solution and sub-solutions depend on the hypothesis parameters a_k. With regards to conducting the hypothesis test, one option is to bound $a_k < \mathcal{A}$. After doing so, U_{ij} and L_{ij} create two sets of inequalities

$$\{U_{ij}(0,0) \geq 1 - \alpha_{ij} : ij \in \{00, 01, 10, 11\}\},$$

$$\{L_{ij}(0,0) \leq 1 - \alpha_{ij} : ij \in \{00, 01, 10, 11\}\}.$$

These can potentially be solved for upper and lower estimates for the sides of the rectangle $(l_1, r_1) \times (l_2, r_2)$. While no longer optimal, these estimates can be used to conduct hypothesis tests with Type I error probability at most α_{ij}.

In summary, we have studied a sequential decision making problem in connection to the Lévy processes. Sequential decision making describes a situation where the decision maker makes

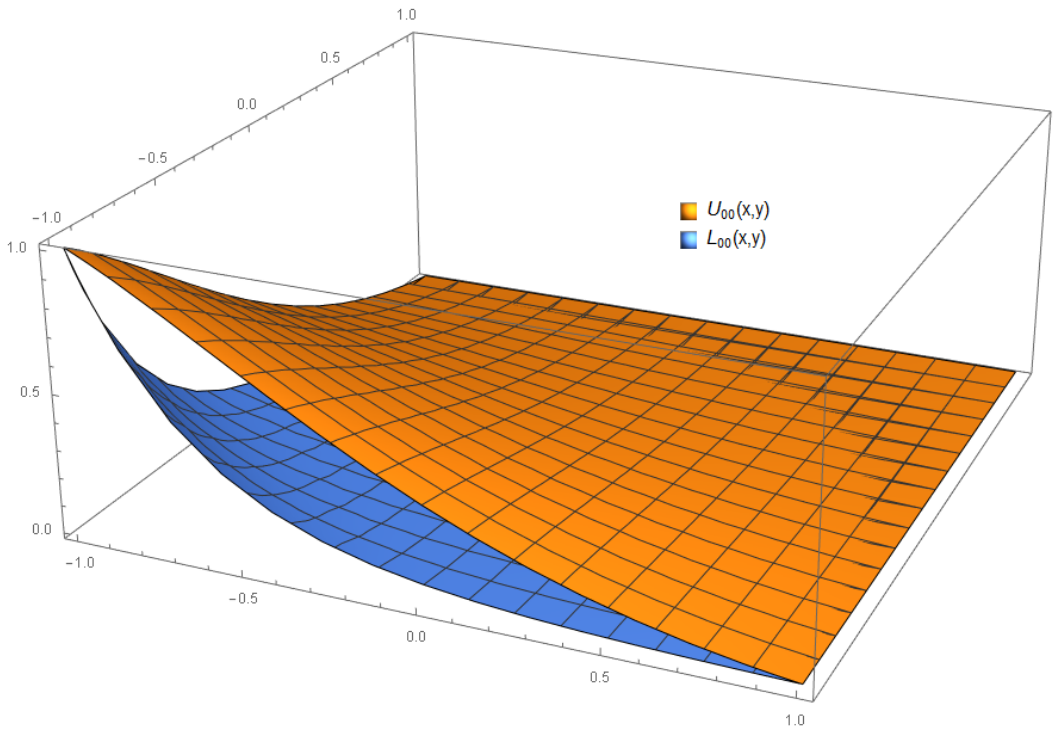

Figure 3.3. The functions U_{00} and L_{00}, which envelope the viscosity solution ξ_{00}.

successive observations of a process before a final decision is made. The procedure to decide when to stop taking observations and when to continue is called the *stopping rule*. This problem can be implemented for financial derivative or commodity markets. A single stochastic model may not appropriately represent derivative or commodity market dynamics. However, the procedure presented in this chapter can be incorporated to determine the fluctuations in the *jump term* of the Lévy processes. Consequently, the *jump term* can be replaced or modified. Thus with a minor adjustment, the original model becomes more effective. This modification also enables *long-range dependence* in the new model without significantly changing the model. The next chapter is dedicated to implementing these ideas in a concrete application, focused on oil prices.

4. ONE DIMENSIONAL APPLICATION TO OIL PRICE DATA

4.1. Oil Price Motivation

Various existing hedging algorithms and insurance risks depend on the underlying statistical model of the commodity market. Consequently, an improvement in the underlying model directly improves the hedging strategies and the understanding of insurance risks. In this chapter, we develop a novel statistical methodology for the refinement of stochastic models using various machine and deep learning algorithms.

As availability of information to the public through alternative data sources increases, machine learning is necessary for adequate analysis. Currently, 97% of North American businesses are using machine learning capabilities to analyze and apply data sources to their trading platforms and analytic focused activities (see [24]). The advent of these technologies allows participants to train, test, and project models using data that have historically been inaccessible. "Any innovation that makes better use of data, and enables data scientists to combine disparate sources of data in a meaningful fashion, offers the potential to gain competitive advantage" (see [24]). Trading capabilities, scale, scope, and speeds have increased exponentially with advancements in applications of Artificial Intelligence and Algorithmic trading.

As stated in the previous chapter, a commonly used stochastic model for derivative and commodity market analysis is the BN-S model (see [3, 6, 5, 4, 17, 19, 22]). In [30], the BN-S model is implemented to find an optimal hedging strategy for the oil commodity from the Bakken, a new region of oil extraction that is benefiting from fracking technology. In [34], the BN-S model is used in this way, in the presence of quantity risk for oil produced in that region. In the recent paper [29], a machine learning-based improvement of the BN-S model is proposed. It is shown that this refined BN-S model is more efficient and has fewer parameters than other models which are used in practice as improvements of the BN-S model. Machine learning-based techniques are implemented for extracting a *deterministic component* (θ) out of processes that are usually considered to be completely stochastic. Equipped with the aforementioned θ, the obtained refined BN-S stochastic

model can be implemented to incorporate *long-range dependence* without actually changing the model.

It is clear that the real challenge is to obtain an estimation of the value of the *deterministic component* for an empirical data set. In [29], a naïve way to find this value for crude oil price is proposed. The method proposed in that paper provides an algorithm to form a classification problem for the data set. After that, various machine and deep learning techniques are implemented for that classification problem.

The primary motivation for this chapter is the fact that the refined BN-S model can be successfully implemented for the analysis of crude oil price. In addition, it seems reasonable that some parameters of the refined BN-S model can be estimated by using various machine/deep learning algorithms. Consequently, it opens up the scope of an abundance of financial applications of the model to the commodity markets. With this, we investigate the problem from the perspective of sequential hypothesis testing and change point detection. Subsequently, machine and deep learning algorithms can be implemented to extract a *deterministic component* from a financial data set, improving on the results from [34].

In Section 4.2, a refined BN-S model from [34] is presented, along with some of its properties. In Section 4.3, we provide a general one-dimensional jump size detection analysis based on the sequential testing of hypotheses. In Section 4.4, an overview of the data set is provided, and then two procedures in the predictive classification problem are introduced. Finally, concrete numerical results are shown in Section 4.5.

4.2. A Refined Barndorff-Nielsen & Shephard Model

Many models in recent literature try to capture the stochastic behavior of time series. As an example, for the BN-S model, the stock or commodity with price $S = (S_t)_{t\geq 0}$ on some filtered probability space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{0\leq t\leq T}, \mathbb{P})$ is modeled by (3.8), (3.9), and (3.10), where the parameters $\mu, \beta, \rho, \lambda \in \mathbb{R}$ with $\lambda > 0$ and $\rho \leq 0$ and r is the risk free interest rate where a stock or commodity is traded up to a fixed horizon date T. In the above model W_t is a Brownian motion, and the process $Z_{\lambda t}$ is a subordinator. Also W and Z are assumed to be independent, and $(\mathcal{G}_t)$ is assumed to be the usual augmentation of the filtration generated by the pair (W, Z).

However, the results and theoretical framework are far from being satisfactory. The BN-S model does not incorporate the *long-range dependence* property. As such, the model fails signifi-

cantly for longer ranges of time. To incorporate *long-range dependence*, a class of superpositions of Ornstein-Uhlenbeck (OU)-type processes is constructed in literature in terms of integrals with respect to independently scattered random measures (see [3, 16]). With appropriate conditions, the resulting processes are incorporated with *long-range dependence*. A limiting procedure results in processes that are second-order self-similar with stationary increments. Other resulting limiting processes are stable and self-similar with stationary increments. However, it is statistically unappealing to fit a large number of OU processes, at least by any formal likelihood-based method. To address this issue, in [29] a new method is developed.

As proposed in [29], $S = (S_t)_{t\geq 0}$ on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$, is given by (3.8), where the dynamics of X_t in (3.9) is given by (3.11), where Z and $Z^{(b)}$ are two independent subordinators, and $\theta \in [0, 1]$ is a deterministic parameter. Machine learning algorithms can be implemented to determine the value of θ. The process $Z^{(b)}$ in (3.11) is a subordinator that has greater intensity than the subordinator Z. Also, W, Z and $Z^{(b)}$ are assumed to be independent, and $(\mathcal{F}_t)$ is assumed to be the usual augmentation of the filtration generated by $(W, Z, Z^{(b)})$.

In this case (3.10) will be given by (3.12) where, as before, $\theta' \in [0, 1]$ is deterministic. It is worth noting that when $\theta = 0$, (3.11) reduces to (3.9). Similarly, when $\theta' = 0$, (3.12) reduces to (3.10).

We conclude this section with some properties of this new model. Note that $(1-\mu)\, dZ_{\lambda t} + \mu dZ^{(b)}_{\lambda t}$, where $\mu \in [0, 1]$, is also a Lévy subordinator that is positively correlated with both Z and $Z^{(b)}$. Note that the solution of (3.12) can be explicitly written as

$$\sigma_t^2 = e^{-\lambda t}\sigma_0^2 + \int_0^t e^{-\lambda(t-s)} \left((1-\theta')dZ_{\lambda t} + \theta' dZ^{(b)}_{\lambda t}\right). \tag{4.1}$$

The *integrated variance* over the time period $[t, T]$ is given by $\sigma_I^2 = \int_t^T \sigma_s^2\, ds$, and a straight-forward calculation shows

$$\sigma_I^2 = \epsilon(t, T)\sigma_t^2 + \int_t^T \epsilon(s, T) \left((1-\theta')dZ_{\lambda t} + \theta' dZ^{(b)}_{\lambda t}\right), \tag{4.2}$$

where

$$\epsilon(s, T) = \left(1 - \exp(-\lambda(T-s))\right)/\lambda, \quad t \leq s \leq T. \tag{4.3}$$

We derive a general expression for the characteristic function of the conditional distribution of the log-asset price process appearing in the BN-S model given by equations (3.8), (3.11) and (3.12). For simplicity, we assume

$$\theta = \theta'.$$

As shown in [29], the advantages of the dynamics given by (3.8), (3.11), and (3.12) over the existing models are significant. The following theorem is proved in [29]. From this result, it is clear that as θ is constantly adjusted, for a fixed s, the value of t always has an upper limit. Consequently, $\text{Corr}(X_t, X_s)$ never becomes very small, and thus *long-range dependence* is incorporated in the model.

Theorem 4.2.1. *If the jump measures associated with the subordinators Z and $Z^{(b)}$ are J_Z and $J_Z^{(b)}$ respectively, and $J(s) = \int_0^s \int_{\mathbb{R}^+} J_Z(\lambda d\tau, dy)$, $J^{(b)}(s) = \int_0^s \int_{\mathbb{R}^+} J_Z^{(b)}(\lambda d\tau, dy)$; then for the log-return of the improved BN-S model given by* (3.8), (3.11), *and* (3.12),

$$Corr(X_t, X_s) = \frac{\int_0^s \sigma_\tau^2 d\tau + \rho^2(1-\theta)^2 J(s) + \rho^2\theta^2 J^{(b)}(s)}{\sqrt{\alpha(t)\alpha(s)}}, \tag{4.4}$$

for $t > s$, where $\alpha(\nu) = \int_0^\nu \sigma_\tau^2 d\tau + \nu\rho^2\lambda((1-\theta)^2\, Var(Z_1) + \theta^2\, Var(Z_1^{(b)}))$.

We implement the above analysis to empirical data sets. For example, we consider the West Texas Intermediate (WTI or NYMEX) crude oil prices data set for the period June 1, 2009 to May 30, 2019 (Figure 1). In the recent paper [29], the appropriateness of modeling such data with a BN-S type stochastic volatility model is discussed. It is clear that such a process is dependent on random shocks, and thus an implementation of the classical model is argued in [29]. However, in Figure 2, we provide the autocorrelation function of the given data set. It is clear that the long-range dependence criteria must be incorporated in the stochastic model. This justifies the implementation of the refined BN-S model presented in this section. We will discuss a detailed data analysis in Section 4.4. The implementation of the refined BN-S model in lieu of the classical BN-S model comes with the price of the estimation of θ as described earlier. In the later sections, this serves as a motivation to apply sequential hypothesis testing combined with various machine/deep learning algorithms. This leads to some novel numerical results related to the present data set.

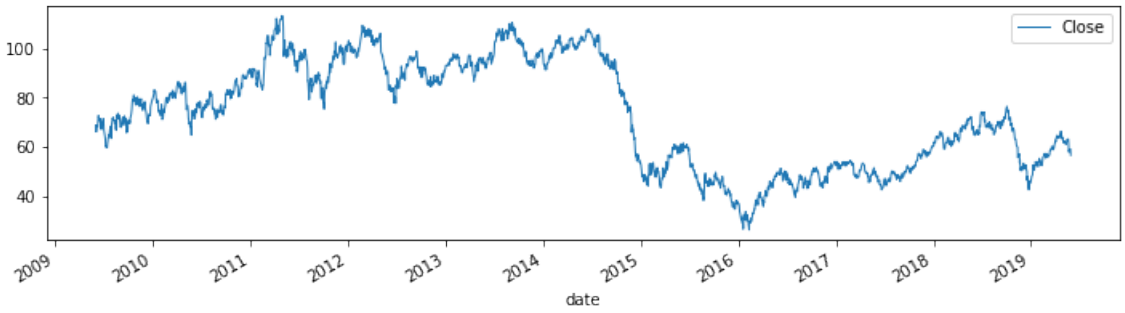

Figure 4.1. Crude oil close price.

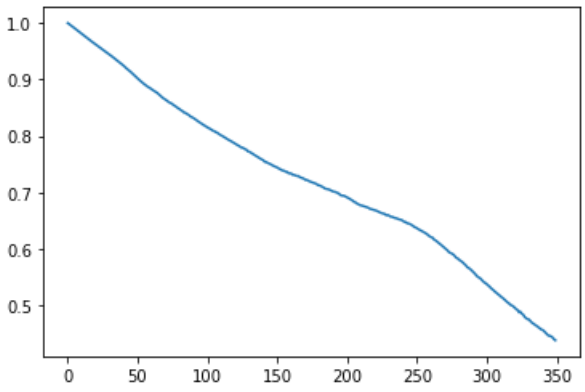

Figure 4.2. Autocorrelation in crude oil close price.

We denote $Z^{(e)} = (1-\theta)Z + \theta Z^{(b)}$. Note that $Z^{(e)}$ is also a subordinator. We call this the *effective* subordinator. We denote the cumulant transforms as $\kappa^{(e)}(\theta) = \log E^{\mathbb{P}}[e^{\theta Z_1^{(e)}}]$. In this work, we make the following assumption similar to [23, 28].

Assumption 4.2.2. *Assume that* $\hat{\theta}^{(e)} = \sup\{\theta \in \mathbb{R} : \kappa^{(e)}(\theta) < +\infty\} > 0$.

We state the following well-known result from [23, 28] and denote the real part and imaginary part of $z \in \mathbb{C}$ as $\Re(z)$ and $\Im(z)$, respectively.

Theorem 4.2.3. *Let* Z *be a subordinator with cumulant transform* κ*, and let* $f : \mathbb{R}_+ \to \mathbb{C}$ *be a complex-valued, left continuous function such that* $\Re(f) \leq 0$*. Then*

$$E\left[\exp\left(\int_0^t f(s)\, dZ_{\lambda s}\right)\right] = \exp\left(\lambda \int_0^t \kappa(f(s))\, ds\right). \tag{4.5}$$

The above formula still holds if $Z = Z^{(e)}$ *satisfies Assumption 4.2.2 and* f *is such that* $\Re(f) \leq \frac{\hat{\theta}^{(e)}}{(1+\epsilon)}$*, for* $\epsilon > 0$.

The *Laplace transform* of $X_{T|t}$, the conditional distribution of X_T given the information up to time $t \leq T$, is given by $\phi(z) = E^{\mathbb{P}}[\exp(zX_T)|\mathcal{F}_t]$, for $z \in \mathbb{C}$ such that the expectation is well-defined.

Theorem 4.2.4. *In the case of the general BN-S model described in equations* (3.8), (3.11) *and* (3.12), *the Laplace transform* $\phi(z) = E[\exp(zX_T)|\mathcal{F}_t]$ *of* $X_{T|t}$ *is given by*

$$\phi(z) = \exp\left(z(X_t + \mu(T-t)) + \frac{1}{2}(z^2 + 2\beta z)\epsilon(t,T)\sigma_t^2 + \lambda \int_t^T G(s,z)\, ds\right), \tag{4.6}$$

where $G(s,z) = \kappa^{(e)}\left(\rho z + \frac{1}{2}(z^2 + 2\beta z)\epsilon(s,T)\right)$.

The transform $\phi(z)$ *is well-defined in the open strip* $\mathcal{S} = \{z \in \mathbb{C} : \Re(z) \in (\theta_-, \theta_+)\}$, *where*

$$\theta_- = \sup_{t \leq s \leq T}\{-\beta - \frac{\rho}{\epsilon(s,T)} - \sqrt{\Delta_1}\},$$

and

$$\theta_+ = \inf_{t \leq s \leq T}\{-\beta - \frac{\rho}{\epsilon(s,T)} + \sqrt{\Delta_1}\},$$

where $\Delta_1 = (\beta + \frac{\rho}{\epsilon(s,T)})^2 + 2\frac{\hat{\theta}^{(e)}}{\epsilon(s,T)}$.

Proof. We obtain from equation (3.11)

$$X_T = \zeta + \beta\sigma_I^2 + \int_t^T \sigma_s\, dW_s + \rho \int_t^T dZ_{\lambda s}^{(e)},$$

where $\zeta = X_t + \mu(T-t)$. Let $\mathcal{G}$ denote the σ-algebra generated by $Z^{(e)}$ up to time T and by $\mathcal{F}_t$. Then, proceeding by iterated conditional expectations, we obtain

$$\begin{aligned}
\phi(z) &= E^{\mathbb{P}}[\exp(zX_T)|\mathcal{F}_t] \\
&= E^{\mathbb{P}}\left[E^{\mathbb{P}}\left[\exp(z(\zeta + \beta\sigma_I^2 + \int_t^T \sigma_s\, dW_s + \rho\int_t^T dZ^{(e)}))|\mathcal{G}\right]|\mathcal{F}_t\right] \\
&= E^{\mathbb{P}}\left[\exp(z(\zeta + \beta\sigma_I^2 + \rho\int_t^T dZ_{\lambda s}^{(e)}))E^{\mathbb{P}}\left[\exp(z\int_t^T \sigma_s\, dW_s)|\mathcal{G}\right]|\mathcal{F}_t\right] \\
&= E^{\mathbb{P}}\left[\exp\left(z(\zeta + \beta\sigma_I^2 + \rho\int_t^T dZ_{\lambda s}^{(e)}) + \frac{1}{2}\sigma_I^2 z^2\right)|\mathcal{F}_t\right].
\end{aligned}$$

Using (4.2) we obtain

$$\phi(z) = \exp\left(\zeta z + \frac{1}{2}\epsilon(t,T)\sigma_t^2(z^2+2\beta z)\right) E^{\mathbb{P}}\left[\exp\left(\int_t^T \left(\rho z + \frac{1}{2}(z^2+2\beta z)\epsilon(s,T)\right) dZ_{\lambda s}^{(e)}\right)\right].$$

Clearly if $z \in \mathcal{S}$, then $\Re(\rho z + \frac{1}{2}(z^2+2\beta z) < \hat{\theta}$. Thus the result follows from (4.5). □

4.3. Jump Size Change Point Detection Based on Hypothesis Tests

In Section 4.2, it is observed that the refined BN-S model can be successfully implementable only when θ can be successfully computed for (3.11) and (3.12) (with $\theta = \theta'$). Note that, as discussed in the previous section, the value of θ is in the interval $0 \leq \theta \leq 1$. However, in order to simplify the subsequent analysis, θ is rounded to either 0 or 1. This is motivated by the simplistic assumption that the jumps are either from one distribution or another. Also, in this case it is easier to interpret the confusion matrix corresponding to a related classification algorithm. To find θ, in [29], a machine learning based empirical analysis is implemented. However, the procedure implemented in that paper does not incorporate any hypothesis testing for θ. In this section, we provide a more theoretical jump size detection analysis based on the sequential test of a hypothesis.

4.3.1. Theoretical results

We consider a Lévy process Z defined by Lévy triplet (μ, σ^2, ν^*), where μ is the drift, σ is the diffusion, and $\nu^*(dx) = (1+\alpha x)\nu(dx)$ for some Lévy measure ν defined on $\mathbb{R}^+$. We are interested in detecting a *significant jump* in the process. Consequently, we wish to test the hypotheses

$$H_0 : \alpha = 0, \qquad H_1 : \alpha = a > 0, \tag{4.7}$$

which clearly address the size of the jumps in the Lévy process.

The Lévy process generates a filtration, which will be denoted $\mathcal{F}_t^{(i)}$, $i = 0, 1$. Further, the hypotheses induce probability measures P_i, $i = 0, 1$. We seek to create a decision rule (τ, δ_τ), where τ is a stopping rule with respect to $\mathcal{F}_t$, and δ_τ is a random variable taking values in the index set $\{0, 1\}$, denoting which hypothesis to reject.

Let the log-likelihood ratio of the marginal density be given by

$$u_t = \log \frac{dP_0}{dP_1}, \tag{4.8}$$

and consider an interval $[l, r] \subset \mathbb{R}$ with $l < 0 < r$. We define the decision rules to be

$$\begin{aligned} \tau &= \inf\{t \geq 0 : u_t \notin [l, r]\}, \\ \delta_\tau &= 1, \text{ if } u_\tau \leq l, \\ \delta_\tau &= 0, \text{ if } u_\tau \geq r. \end{aligned} \tag{4.9}$$

Theorem 4.3.1. *With the process u_t defined as in* (4.8)*, we have infinitesimal generators, given by*

$$\mathcal{L}\xi(x) := -\gamma\xi'(x) + \frac{1}{2}\beta^2\xi''(x) + \int_{\mathbb{R}_+} \left(\xi(x+y) - \xi(x) - \frac{y\xi'(x)}{1+|y|}\right) K(dy),$$

for any suitable ξ, where

$$\beta = -a \int_{x>0} (1 \wedge x)\sigma^{-1} x\nu(dx), \tag{4.10}$$

$$m = a \int_{x>1} x\nu(dx), \tag{4.11}$$

$$\gamma = m - \frac{\beta^2}{2} + \int_0^1 (\log(1+x)^2 - x)a\nu(dx), \tag{4.12}$$

$$K = a\log(1+x)\nu. \tag{4.13}$$

Proof. Since z is a Lévy process with characteristics (μ, σ^2, ν) under P_0 and characteristics $(\mu, \sigma^2, (1+ax)\nu)$ under P_1, we apply the generalized Girsanov's Theorem. Using β as in (4.10), we obtain

$$\frac{dP_0}{dP_1} = \mathcal{E}\left(-N.\right)_t,$$

where

$$N_t = \beta W_t + \int_0^t \int_{x>0} ax(J - \nu)(ds, dx),$$

aJ is the jump measure for N, W is a standard Brownian motion, and $\mathcal{E}$ is the Doléans-Dade exponential, defined previously. This gives that N_t is a Lévy process with characteristics

$$(m, \beta^2, -a\nu).$$

Then, by [12] (Proposition 8), we obtain characteristics

$$(\gamma, \beta^2, -K),$$

for u_t. Finally, by [12] and Theorem 2.4.2, the process has the stated generator. □

Assign ξ to be the probability of a correct decision given H_0. Then we have the partial integro-differential equation $\mathcal{L}\xi = 0$ with boundary conditions

$$\xi(l) = 1, \qquad \xi(r) = 0. \tag{4.14}$$

Further, we have $\xi > 0$ inside $R = (l, r)$.

The following are one-dimensional adaptations of 3.3.3 and 3.3.4:

Theorem 4.3.2. *If $F : \mathbb{R}^5 \to \mathbb{R}$, and*

(A1) $F(x, u, p, X, i_1) \leq F(x, u, p, Y, i_2)$ if $X \geq Y$ and $i_1 \geq i_2$,

(A2) there exists $\gamma > 0$ such that for any $x, u, v, p, X, i \in \mathbb{R}$,

$$F(x, u, p, X, i) - F(x, v, p, X, i) \geq \gamma(u - v) \text{ if } u \geq v,$$

for some $\epsilon > 0$ and $r(\beta) \to 0$ as $\beta \to 0$, we have

$$F(y, v, \epsilon^{-1}(x - y), Y, i) - F(x, v, \epsilon^{-1}(x - y), X, i) \leq \omega_R(\epsilon^{-1}|x - y|^2 + |x - y| + r(\beta)),$$

(A3) F is uniformly continuous with respect to all arguments,

(A4) $\sup_{x \in \mathbb{R}} |F(x, 0, 0, 0, 0)| < \infty$,

(A5) K is a Lévy-Itô measure,

(A6) the inequalities in (3.19) *are strict,*

(A7) for any $R > 0$, there exists a modulus of continuity ω_R such that, for any $x, y \in \mathbb{R}$, $|v| \leq R$, $i \in \mathbb{R}$, and for any $X, Y \in \mathbb{R}$ satisfying

$$\begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \leq \frac{1}{\epsilon} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + r(\beta) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

then there is a unique solution to $F(0, \xi, D\xi, D^2\xi, \mathcal{I}[\xi](x)) = 0$ between any pair of super-solution and sub-solutions, where

$$\mathcal{I}[\xi](x) := \int_{\mathbb{R}_+} \left(\xi(x+y) - \frac{y\xi'(x)}{1+|y|} \right) K(dy).$$

Lemma 4.3.3. *In particular, the function*

$$F(x, u, p, X, i) := Mu + \gamma p - \frac{\beta}{2} X - i$$

satisfies (A1)-(A4) and our measure K satisfies (A5) in (4.3.2)*, where*

$$M = \int_{\mathbb{R}_+} K(dy).$$

Proof. First, consider (A1):

$$F(x, u, p, X, i_1) - F(x, u, p, Y, i_2) = \frac{\beta}{2}(Y - X) + i_2 - i_1 \geq 0$$

if $i_2 \leq i_1$ and $Y \leq X$.

Next, $F(x, u, p, X, i) - F(x, v, p, X, i) = M(u - v)$, so choosing $\gamma = M > 0$, we have property (A2).

Property (A3) is satisfied because F is linear in each argument, and (A4) is satisfied because F does not depend on its first argument explicitly. Last, K is a Lévy-Itô measure by the assumptions of the underlying Lévy process.

□

We now use the infinitesimal generators to prove the existence of a viscosity solution. Using all of the previous, we can finally state the existence theorem.

Theorem 4.3.4. *If ξ is monotonic, then the partial integro-differential equation $\mathcal{L}\xi = 0$, subject to boundary conditions* (4.14) *and $\xi > 0$ has a viscosity solution between sub-solution and super-solution*

$$g(x) = \exp(B(x-l))\frac{\sinh\left(\frac{r-x}{\beta}\sqrt{2M+B^2}\right)}{\sinh\left(\frac{r-l}{\beta}\sqrt{2M+B^2}\right)},$$
$$f(x) = \frac{\exp(2Br) - \exp(2Bx)}{\exp(2Br) - \exp(2Bl)},$$

where

$$C = \int_0^\infty \frac{y}{1+|y|}K(dy),$$
$$B = \frac{2(C+\gamma)}{\beta^2},$$
$$M = \int_0^\infty K(dy).$$

Proof. We define

$$H(x) = \int_0^\infty \xi(x+y)K(dy),$$
$$M = \int_0^\infty K(dy).$$

Consequently,

$$0 = \mathcal{L}\xi(x) = -\gamma\xi'(x) + \frac{1}{2}\beta^2\xi''(x) + \int_{\mathbb{R}_+}\left(\xi(x+y) - \xi(x) - \frac{y\xi'(x)}{1+|y|}\right)K(dy)$$

can be rewritten as

$$0 = -\gamma\xi'(x) + \frac{1}{2}\beta^2\xi''(x) + H(x) - M\xi(x) - C\xi'(x).$$

The sign on H is positive; therefore, we have sub-solution equation

$$0 = \frac{1}{2}\beta^2\xi''(x) - (C+\gamma)\xi'(x) - M\xi(x).$$

On the other hand, since ξ is monotonic and positive inside R,

$$\xi(x+y) - \xi(x) \leq 0 \iff H(x) - M\xi(x) \leq 0.$$

Using this, we have super-solution equation

$$0 = -\gamma\xi'(x) + \frac{1}{2}\beta^2\xi''(x) - C\xi'(x).$$

Finally, applying the previous theorem 4.3.2, we have the existence of a viscosity solution.

□

Remark 4.3.5. *The monotonicity assumption yields a tighter super- and sub-solution envelope and is here to make the application of this theorem to time series data more effective, but is not necessary for the proof of a viscosity solution's existence.*

4.3.2. Jump size detection algorithm

We will use the prior super- and sub-solutions as envelopes to approximate an important parameter in the following algorithm that uses the previous hypothesis test to classify Lévy processes as having small or large jumps.

Given oil price close values in length-n work day periods, we do the following:

1. An inverse Gaussian density $\nu(dx)$ is fit to the distribution of negative percent daily jumps for the entire (training) data set.

2. We then fit the density of the Lévy measure from 4.3.1, $\nu^*(dx) = (1+ax)\nu(dx)$, to the distribution of the negative percent daily jumps for the n-length period. This gives a test statistic a for the parameter in the hypothesis test.

3. We calculate the standard deviation σ of all daily percent changes for the n-length period.

4. Using the density ν^* and standard deviation σ, we calculate γ, β, and C from 4.3.1 and 4.3.4.

5. The left side of the interval is chosen to be -1, then using a, σ, β, γ, and C in the super- and sub-solution equations in 4.3.4, we can solve for the right side of the interval using $f(0) = 1-\alpha_0$ and $g(0) = 1-\alpha_0$, and take the average of the two. The parameter α_0 is chosen to be the maximum desired probability of a type I error.

6. Simulations of the log-likelihood process with drift γ, volatility β, and jumps represented by an inverse Gaussian process with expected value $-t\int_0^\infty xK(dx)$ at time t, are run. We record the frequency of exits out of the right-side of the interval to get a number that represents, relatively, the size of the jumps. We call this number the right-exit frequency.

7. Time periods whose right-exit frequencies are at or above a certain threshold p^* are then classified as having large jumps, while the others are classified as having small jumps.

To demonstrate the capacity of the hypothesis testing algorithm in distinguishing between processes with small and large jumps, we run it on simulated data. Multiple classes of Lévy processes are simulated, all of which start initially at 100:

1. a training data time series with drift 1, diffusion 0.5, and jumps that follow an inverse Gaussian distribution with mean 1 and scale factor 1, which gives a Lévy measure ν,

2. a control data set of 100 processes with parameters identical to the training data,

3. a data set of 100 processes with obvious large negative jumps: the parameters are the same as the training set except the Lévy measure is now represented by $(1 + x)\nu(dx)$, and

4. a data set of 100 processes with subtle large negative jumps: the drift is increased to 3 to compensate for the previous increase in jump size.

The training time series is run for 500 time periods, and the other three are run for 30 each, with representatives shown in Figure 4.3.

The hypothesis test algorithm with $p^* = 8$ and $\alpha_0 = 0.1$ is run on each data set. (The parameter p^* is chosen here to give desirable results and will be used in the application in the next section.) For the control, 79 out of the 100 processes are correctly identified as coming from the distribution with small jumps. This is to be expected because some number of processes would randomly have significantly larger jumps just by chance. All 100 from the obvious large jumps set

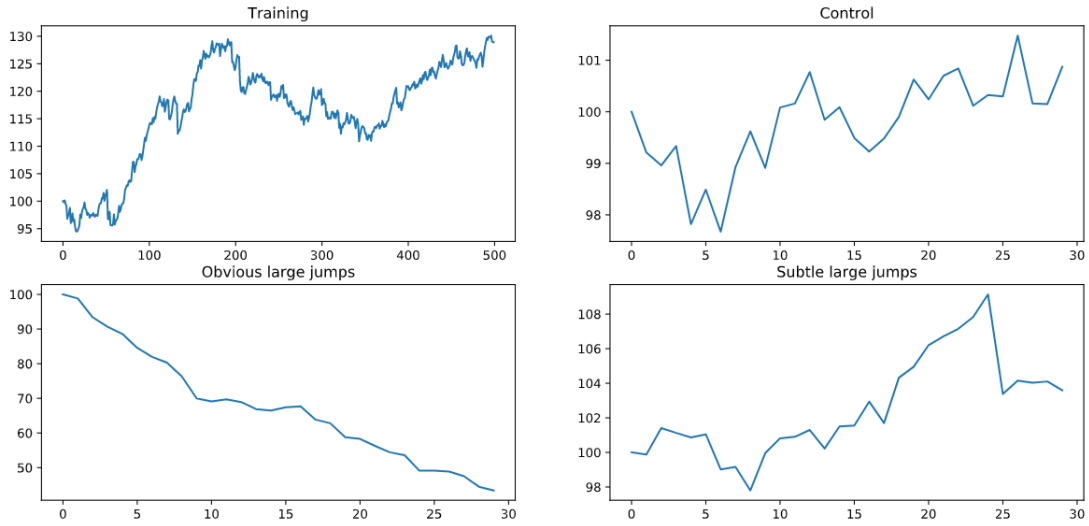

Figure 4.3. The training data and a representative from each other data set.

are identified as having large jumps, and 85 out of the 100 processes in the subtle large jumps set are correctly identified.

Alternatively, a naïve approach of simply classifying each 30 day period based on comparing only the mean jump size relative to the training data's mean jump size results in only 28 of the 100 control processes being correctly identified; although it correctly identifies all but 3 of the large jump simulations. Because of the significant potential for type I error in this naïve approach, the hypothesis test algorithm has evident advantages.

4.4. Prediction Method

We briefly discussed the data set in Section 4.2. In this section, we present an overview of the data set in its entirety, and then develop two procedures used in the predictive classification problem. As discussed in Section 4.2, we consider the West Texas Intermediate (WTI or NYMEX) crude oil prices data set for the period June 1, 2009 to May 30, 2019. West Texas Intermediate crude oil is described as light sweet oil traded and delivered at Cushing, Oklahoma. WTI usually refers to the price of the New York Mercantile Exchange (NYMEX) WTI Crude Oil futures contract. For WTI, spot and futures prices are used as a benchmark in oil pricing. The WTI crude oil futures contract specifies the deliverable asset for the contract to be a blend of crude oil, as long as it is of acceptable lightness and sweetness. The data set is available online in [14]. We index the available dates from 0 (for June 1, 2009) to 2529 (for May 30, 2019).

The following table summarizes various estimates for the data set.

Table 4.1. Properties of the empirical data set.

	Daily Price Change	Daily Price Change %
Mean	-0.0047	0.01370 %
Median	0.04399	0.06521 %
Maximum	7.62	12.32 %
Minimum	-8.90	-10.53 %

In Figure 4.4 the distribution plot for close oil price is provided. Histograms for daily change in close oil price and daily change percentage in close oil price are provided in Figure 4.5 and Figure 4.6, respectively, for exploratory purposes.

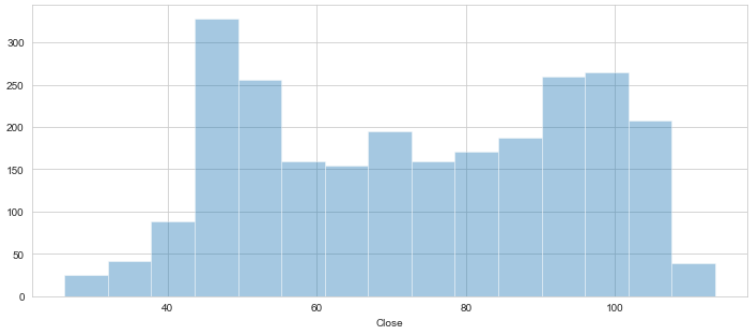

Figure 4.4. Distribution plot for close oil price.

In the following subsections, two procedures are described for constructing the related classification problem. The procedures differ in the features used for the analysis: percent daily changes and right-exit frequencies. In each, the algorithm at the end of Section 4.3 is used to determine whether an individual time period has large or small jumps, represented by the right-exit frequency of that time period. The machine learning algorithms are then used to predict whether the right-exit frequency of the next time period will be large or small. Consequently, before truncation, the resulting probabilities of large jumps from each machine learning algorithm can be used to update θ from the refined BN-S model in Section 4.2 each period.

4.4.1. Percent daily changes as features

We implement the following procedure to create a machine learning classification problem:

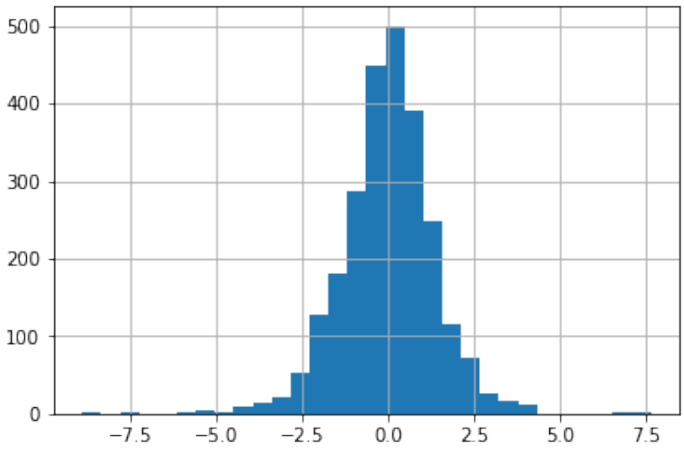

Figure 4.5. Histogram for daily change in close oil price.

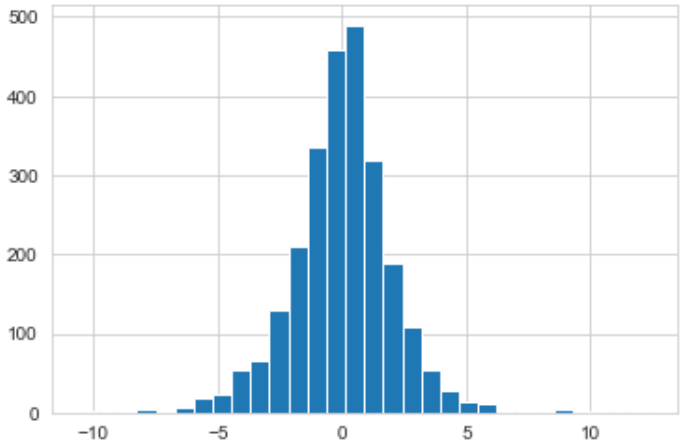

Figure 4.6. Histogram for daily change percentage in close oil price.

1. We consider the percent daily changes for the historical oil price data and create a new data-frame from the old where the columns will be n consecutive daily change percentages. For example, if the changes are

$$a_1, a_2, a_3, ...,$$

then the first row of the data set will be

$$a_1, a_2, ..., a_n,$$

and the second row will be

$$a_2, a_3, ..., a_{n+1},$$

and so forth.

2. We create a target column that is 0 if the right-exit frequency of the next disjoint n days is less than some threshold p^*, and is 1 otherwise. For example, if the time period

$$a_{1+n}, a_{2+n}, ..., a_{2n-1}$$

has a significant frequency of right-exits, then the time period

$$a_1, a_2, ..., a_n$$

is given a target value 1.

3. We run various classification algorithms where the input is a list of n consecutive close prices, and the output is a 1 to represent large jumps or 0 to represent small jumps of the next n consecutive close prices. Classification reports and confusion matrices are evaluated for each algorithm.

4.4.2. Right-exit frequencies as features

We implement the following procedure to create a machine learning classification problem:

1. Similar to the previous, we consider close prices for the historical oil price data and create a new data-frame exactly as before.

2. A new column is created that holds the right-exit frequencies for each consecutive set of n days, say,

$$b_1, b_2, b_3, ...$$

These represent how large the jumps in close prices are for the previous n days.

3. From this column, a new staggered data-frame is created, similar to before.

4. Finally, a target column is created: if the row is

$$b_{30}, b_{31}, ..., b_{30+n-1},$$

then the entry in the target column will be $b_{30+2n-1}$. This is the right-exit frequency of the next disjoint n-day period.

5. We run various classification algorithms where the input is a list of n consecutive right-exit frequencies, and the output is 1 to represent large jumps or 0 to represent small jumps of the next n consecutive close prices. Classification reports and confusion matrices are evaluated for each algorithm.

4.5. Numerical results

Now we apply the procedures described in the previous section to specific cases. For this section, the period length $n = 30$. Further, α, the parameter representing an approximation for the type I error of the test is chosen to be $\alpha = 0.9$, and because it worked optimally in the simulation study, the cut off for significant right-exit frequencies is chosen to be $p^* = 8$. Two different time periods are used for training, and two are used for testing. The time periods are

- T_1: *training data (index)*: October 21, 2009 (100) to May 17, 2013 (1000); and *testing data (index)*: April 21, 2017 (2000) to April 10, 2019 (2500);
- T_2: *training data (index)*: August 11, 2009 (50) to May 13, 2013 (1500); and *testing data (index)*: October 5, 2015 (1600) to January 29, 2019 (2450).

Because the data is significantly imbalanced in favor of small jump time periods, random small jump periods from the training data are removed while performing algorithms 4.4.1 and 4.4.2. This allows the neural nets and other algorithms to isolate the attributes of large and small jump periods without becoming distracted by the imbalanced frequency of small jump periods. Without doing so, the algorithms often predict all time periods to be small jump periods – simply because those are more prevalent. The results of the machine learning algorithms using the time periods above are recorded in the following tables (Tables 4.2-4.5). Those used are linear regression (LR), decision trees (DT), random forests (RF), and three different types of neural nets, (A) a standard net, (B) a long-short term memory net, and (C) a LSTM net with a batch normalizer.

Most of the machine learning algorithms perform better than how one might expect from guessing uniformly whether the next time period would have big jumps. Some perform notably

poorly, however, particularly the LSTM neural nets without a batch normalizer. However, the neural nets with a batch normalizer consistently perform quite well.

Figure 4.7 provides a histogram showing the distribution of right-exit frequencies for period lengths of 30 business days in the T_2 testing data. For each set of 30 consecutive days, 10 simulations are run, and the frequency of right-exits is recorded. The x-axis in the figure is the number of simulated processes that exit to the right of the testing interval for a given period, while the y-axis is the number of 30 day periods with that frequency of right-exits.

Once the value of θ is estimated, this can be implemented in the refined BN-S model (3.11) (and (3.12), with $\theta = \theta'$). Equipped with θ, as described in [29] and as shown in Theorem 4.2.1, the refined BN-S stochastic model can be used to incorporates *long-range dependence* without actually changing the model. In addition, this shows a real-time application of data science for extracting a *deterministic component* out of processes that are thus far considered to be completely stochastic. By the *deterministic component*, it is meant that θ is a deterministic signal. This is deterministic in the sense that its value is extracted from the data before the model is implemented. Once the value of θ is obtained, it is kept constant for a certain period of time. For the computational effectiveness of θ, the results in Tables 4.2-4.5 show better estimation compared to the benchmark study in [29].

Table 4.2. Various estimations for T_1, using daily percent changes as features.

	LR	DT	RF	Neural Network (A)	LSTM (B)	BN (C)
precision $\theta = 0$	0.92	0.89	0.89	0.88	0.83	0.93
recall $\theta = 0$	0.61	0.56	0.64	0.54	0.77	0.88
f1-score $\theta = 0$	0.74	0.69	0.74	0.67	0.80	0.90
support $\theta = 0$	340	340	340	340	340	340
precision $\theta = 1$	0.29	0.24	0.25	0.22	0.19	0.53
recall $\theta = 1$	0.76	0.67	0.60	0.64	0.26	0.66
f1-score $\theta = 1$	0.42	0.35	0.36	0.33	0.22	0.59
support $\theta = 1$	70	70	70	70	70	70

Table 4.3. Various estimations for T_1, using right-exit frequencies as features.

	LR	DT	RF	Neural Network (A)	LSTM (B)	BN (C)
precision $\theta = 0$	0.89	0.92	0.87	0.83	0.88	0.85
recall $\theta = 0$	0.66	0.67	0.69	0.56	0.21	0.71
f1-score $\theta = 0$	0.76	0.77	0.77	0.67	0.34	0.77
support $\theta = 0$	340	340	340	340	340	340
precision $\theta = 1$	0.24	0.30	0.25	0.17	0.18	0.21
recall $\theta = 1$	0.61	0.70	0.50	0.43	0.86	0.39
f1-score $\theta = 1$	0.37	0.42	0.33	0.24	0.30	0.27
support $\theta = 1$	70	70	70	70	70	70

Table 4.4. Various estimations for T_2, using daily percent changes as features.

	LR	DT	RF	Neural Network (A)	LSTM (B)	BN (C)
precision $\theta = 0$	0.79	0.74	0.79	0.80	0.77	0.75
recall $\theta = 0$	0.82	0.50	0.57	0.66	0.58	0.91
f1-score $\theta = 0$	0.80	0.59	0.66	0.72	0.66	0.82
support $\theta = 0$	519	519	519	519	519	519
precision $\theta = 1$	0.57	0.37	0.42	0.47	0.41	0.65
recall $\theta = 1$	0.53	0.63	0.68	0.63	0.63	0.37
f1-score $\theta = 1$	0.55	0.46	0.52	0.54	0.50	0.47
support $\theta = 1$	241	241	241	241	241	241

Table 4.5. Various estimations for T_2, using right-exit frequencies as features.

	LR	DT	RF	Neural Network (A)	LSTM (B)	BN (C)
precision $\theta = 0$	0.80	0.76	0.79	0.76	0.80	0.75
recall $\theta = 0$	0.54	0.56	0.58	0.54	0.16	0.63
f1-score $\theta = 0$	0.65	0.64	0.67	0.63	0.27	0.68
support $\theta = 0$	519	519	519	519	519	519
precision $\theta = 1$	0.42	0.39	0.42	0.39	0.34	0.40
recall $\theta = 1$	0.70	0.62	0.66	0.64	0.91	0.54
f1-score $\theta = 1$	0.52	0.48	0.52	0.49	0.49	0.46
support $\theta = 1$	241	241	241	241	241	241

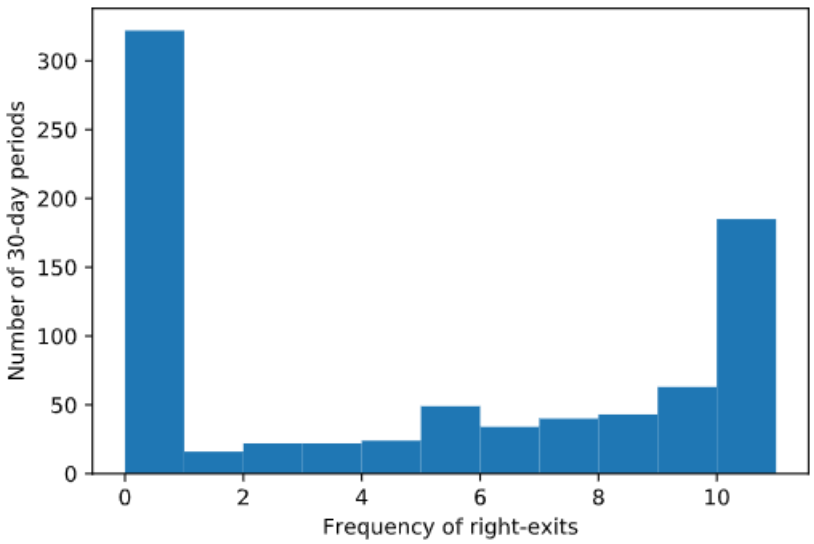

Figure 4.7. Histogram for daily (previous 30 days) right-exit frequencies.

5. CONCLUSION

Motivated by the fact that the refined BN-S model can be successfully implemented to the analysis of crude oil price, and that the parameters of the refined BN-S model can be estimated by using various machine/deep learning algorithms, in this dissertation we study the refined BN-S model from the sequential hypothesis testing perspective, with an application to the oil market.

Mathematical modeling of oil price data is directly inspired by various stochastic models. Thorough understanding and theoretical development of appropriate stochastic models contribute to a better understanding of the risk-management problem of various commodities, and various existing algorithms in a financial market depend on the underlying statistical model. Consequently, an improvement in the underlying model directly improves the existing algorithms. A sequential decision making problem in connection to Lévy processes is studied to analyze the jump size distribution. This is coupled with various machine and deep learning techniques to improve the existing stochastic models. Consequently, the analysis presented in this dissertation provides a necessary mathematical framework for an appropriate generalization of various stochastic models.

Future works related to this topic should definitely include seeking to find a more adequate approximation for the right side of the decision rule intervals. This would greatly increase the sensitivity of the algorithms in classifying large-jump time periods, thereby requiring less computational power for even better results. Applications to other data sets more independent of exogenous forces, and constructing decision rules for hypothesis tests on other parameters in the underlying processes could open up this type of analysis to more generalized scenarios, as well, and studying applications across multiple streams of data using methods similar to those in Chapter 3, should also be explored. For this dissertation, only the case $\rho = 0$ is considered. The situation becomes much more involved when $\rho \neq 0$. This will be considered as a follow-up work of this dissertation. Further, it is worth investigating whether some method exists to determine analytically exact bounds of the rectangle used in these decision rules based on super-solutions or sub-solutions. Even more generally, additional hypothesis tests could be developed. One of such could be a test on the Lévy measures while keeping a constant diffusion coefficient for each underlying Lévy process. Another could be a test on the diffusion terms with no drift terms in either process. Last, it is worth exploring the

one-dimensional test more, which could yield a solution useful for finding final boundary conditions for the two-dimensional test, giving uniqueness of the likelihood function solution.

REFERENCES

[1] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2009.

[2] G. Barles, E. Chasseigne, and C. Imbert. On the Dirichlet problem for second-order elliptic integro-differential equations. *Indiana Univ. Math. J.*, 57(1):213–246, 2008.

[3] O. E. Barndorff-Nielsen. Superposition of Ornstein-Uhlenbeck type processes. *Theory Probab. Appl.*, 45:175–194, 2001.

[4] O. E. Barndorff-Nielsen, J. L. Jensen, and M. Sørensen. Some stationary processes in discrete and continuous time. *Adv. in Appl. Probab.*, 30:989–1007, 1998.

[5] O. E. Barndorff-Nielsen and N. Shephard. *Lévy Processes: Theory and Applications*, chapter Modelling by Lévy Processes for Financial Econometrics, pages 283–318. Birkhäuser, 2001.

[6] O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63:167–241, 2001.

[7] C. Baum and V. Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6):1994–1997, 1994.

[8] B. Brodsky and B. Darkhovsky. Minimax methods for multihypothesis sequential testing and change-point detection problems. *Sequential Analysis*, 27(2):141–173, 2008.

[9] M. Carlisle and O. Hadjiliadis. Sequential decision making in two-dimensional hypothesis testing. In *52nd IEEE Conference on Decision and Control*, pages 6508–6515, 2013.

[10] K. P. Chen, Y. B. Tsai, D. Amorese, and W. Y. Chang. Incorporating change-point detection updates of frequency-magnitude distributions within the taiwan earthquake. *Terr. Atmos. Ocean. Sci.*, 22:261–269, 2011.

[11] Y. S. Chow, H. Robbins, and D. Siegmund. *The theory of optimal stopping*. Dover Publications, 1991.

[12] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall / CRC Press, 2003.

[13] S. Dayanik, V. Poor, and S. Sezer. Sequential multi-hypothesis testing for compound Poisson processes. *Stochastics*, 80(1):19–50, 2008.

[14] Federal Reserve Bank of St. Louis FRED. U.S. Energy Information Administration, Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma [DCOILWTICO]. Available at `https://fred.stlouisfed.org/series/DCOILWTICO` (2020/9/23).

[15] G. K. Golubev and R.Z. Khas'minski. Sequential testing for several signals in Gaussian white noise. *Theory of Probability and its Applications*, 28:573–584, 1983.

[16] S. Habtemicael, M. Ghebremichael, and I. SenGupta. Volatility and variance swap using superposition of the Barndorff-Nielsen and Shephard type Lévy processes. *Sankhya B*, 81:75–92, 2019.

[17] S. Habtemicael and I. SenGupta. Pricing variance and volatility swaps for Barndorff-Nielsen and Shephard process driven financial markets. *International Journal of Financial Engineering*, 03(04):1650027, 2016.

[18] A. Irle. Transitivity in problems of optimal stopping. *Annals of Probability*, 9:642–647, 1981.

[19] A. Issaka and I. SenGupta. Analysis of variance based instruments for Ornstein–Uhlenbeck type models: swap and price index. *Annals of Finance*, 13(4):401–434, 2017.

[20] B. Øksendal and A. Sulem. *Applied Stochastic Control of Jump Diffusions*. Springer, 2005.

[21] G. Lowther. Lévy processes. Available at `https://almostsure.wordpress.com/2010/11/23/levy-processes/` (2021/4/1).

[22] T. Miljkovic and I. SenGupta. A new analysis of VIX using mixture of regressions: examination and short-term forecasting for the S&P 500 market. *High Frequency*, 1(1):53–65, 2018.

[23] E. Nicolato and E. Venardos. Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Math. Finance*, 13:445–466, 2003.

[24] Refinitiv. Available at `https://www.refinitiv.com/en/resources/special-report/refinitiv-2019-artificial-intelligence-machine-learning-global-study` (2019).

[25] M. Roberts and I. SenGupta. Infinitesimal generators for two-dimensional Lévy process-driven hypothesis testing. *Annals of Finance*, 16(1):121–139, 2020.

[26] M. Roberts and I. SenGupta. Sequential hypothesis testing in machine learning, and crude oil price jump size detection. *Applied Mathematical Finance*, 27(5):374–395, 2020.

[27] F. D. Schönbrodt, E.-J. Wagenmakers, M. Zehetleitner, and M. Perugini. Sequential hypothesis testing with bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2):322–339, 2017.

[28] I. SenGupta. Generalized BN-S stochastic volatility model for option pricing. *International Journal of Theoretical and Applied Finance*, 19(02):1650014, 2016.

[29] I. SenGupta, W. Nganje, and E. Hanson. Refinements of Barndorff-Nielsen and Shephard model: an analysis of crude oil price with machine learning. *Annals of Data Science*, 8:39–55, 2021.

[30] I. SenGupta, W. Wilson, and W. Nganje. Barndorff-Nielsen and Shephard model: oil hedging with variance swap and option. *Mathematics and Financial Economics*, 13(2):209–226, 2019.

[31] S. Shreve. *Stochastic Calculus for Finance II*. Springer, 2004.

[32] A. Wald. *Sequential Analysis*. Wiley, New York, 1947.

[33] J. Wannenwetsch. *Lévy Processes in Finance: The Change of Measure and Non-Linear Dependence*. PhD thesis, Rhenish Friedrich Wilhelm University of Bonn, 2005.

[34] W. Wilson, W. Nganje, S. Gebresilasie, and I. SenGupta. Barndorff-Nielsen and Shephard model for hedging energy with quantity risk. *High Frequency*, 2(3-4):202–214, 2019.