Signal processing with Lévy information

Dorje C. Brody[1], Lane P. Hughston[2], Xun Yang[3,4]

[1]*Mathematical Sciences, Brunel University, Uxbridge UB8 3PH, UK*
[2]*Department of Mathematics, University College London, London WC1E 6BT, UK*
[3]*Department of Mathematics, Imperial College London, London SW7 2BZ, UK*
[4]*Shell International Limited, Shell Centre, London SE1 7NA, UK*

Lévy processes, which have stationary independent increments, are ideal for modelling the various types of noise that can arise in communication channels. If a Lévy process admits exponential moments, then there exists a parametric family of measure changes called Esscher transformations. If the parameter is replaced with an independent random variable, the true value of which represents a "message", then under the transformed measure the original Lévy process takes on the character of an "information process". In this paper we develop a theory of such Lévy information processes. The underlying Lévy process, which we call the fiducial process, represents the "noise type". Each such noise type is capable of carrying a message of a certain specification. A number of examples are worked out in detail, including information processes of the Brownian, Poisson, gamma, variance gamma, negative binomial, inverse Gaussian, and normal inverse Gaussian type. Although in general there is no additive decomposition of information into signal and noise, one is led nevertheless for each noise type to a well-defined scheme for signal detection and enhancement relevant to a variety of practical situations.

Key Words: Signal processing; Lévy process; Esscher transformation; nonlinear filtering; innovations process; information process; cybernetics.

I. INTRODUCTION

The idea of filtering the noise out of a noisy message as a way of increasing its information content is illustrated by Norbert Wiener in his book *Cybernetics* (Wiener 1948) by means of the following example. The true message is represented by a variable X which has a known probability distribution. An agent wishes to determine as best as possible the value of X, but due to the presence of noise the agent can only observe a noisy version of the message of the form $\xi = X + \epsilon$, where ϵ is independent of X. Wiener shows how, given the observed value of the noisy message ξ, the original distribution of X can be transformed into an improved *a posteriori* distribution that has a higher information content. The *a posteriori* distribution can then be used to determine a best estimate for the value of X.

The theory of filtering was developed in the 1940s when the inefficiency of anti-aircraft fire made it imperative to introduce effective filtering-based devices (Wiener 1949, 1954). A breakthrough came with the work of Kalman, who reformulated the theory in a manner more well-suited for dynamical state-estimation problems (Kailath 1974, Davis 1977). This period coincided with the emergence of the modern control theory of Bellman and Pontryagin (Bellman 1961, Pontryagin *et al.* 1962). Owing to the importance of its applications, much work has been carried out since then. According to an estimate of Kalman (1994),

over 200,000 articles and monographs had been published on applications of the Kalman filter alone. The theory of stochastic filtering, in its modern form, is not much different conceptually from the elementary example described by Wiener in the 1940s. The message, instead of being represented by a single variable, in the general setup can take the form of a time series (the "signal" or "message" process). The information made available to the agent also takes the form of a time series (the "observation" or "information" process), typically given by the sum of two terms, the first being a functional of the signal process, and the second being a noise process. The nature of the signal process can be rather general, but in most applications the noise is chosen to be a Wiener process (see, e.g., Liptser & Shiryaev 2000, Xiong 2008, Bain & Crisan 2010). There is no reason *a priori*, however, why an information process should be "additive", or even why it should be given as a functional of a signal process and a noise process. From a mathematical perspective, it seems that the often proposed ansatz of an additive decomposition of the observation process is well-adapted to the situation where the noise is Gaussian, but is not so natural when the noise is discontinuous. Thus while a good deal of recent research has been carried out on the problem of filtering noisy information containing jumps (see, e.g., Rutkowski 1994, Ahn & Feldman 1999, Meyer-Brandis & Proske 2004, Poklukar 2006, Popa & Sritharan 2009, Grigelionis & Mikulevicius 2011, and references cited therein), such work has usually been pursued under the assumption of an additive relation between signal and noise, and it is not unreasonable to ask whether a more systematic treatment of the problem might be available that involves no presumption of additivity and that is more naturally adapted to the mathematics of the situation.

The purpose of the present paper is to introduce a broad class of information processes suitable for modelling situations involving discontinuous signals, discontinuous noise, and discontinuous information. No assumption is made to the effect that information can be expressed as a function of signal and noise. Instead, information processes are classified according to their "noise type". Information processes of the same noise type are then distinguished from one another by the messages that they carry. Each noise type is associated to a Lévy process, which we call the *fiducial process*. The fiducial process is the information process that results for a given noise type in the case of a null message, and can be thought of as a "pure noise" process of that noise type. Information processes can then be classified by the characteristics of the associated fiducial processes. To keep the discussion elementary, we consider the case of a one-dimension fiducial process and examine the situation where the message is represented by a single random variable. The goal is to construct the optimal filter for the class of information processes that we consider in the form of a map that takes the *a priori* distribution of the message to an *a posteriori* distribution that depends on the information that has been made available. A number of examples will be presented. The results vary remarkably in detail and character for the different types of filters considered, and yet there is an overriding unity in the general scheme, which allows for the construction of a multitude of examples and applications.

A synopsis of the main ideas, which we develop more fully in the remainder of the paper, can be presented as follows. We recall the idea of the Esscher transform as a change of probability measure on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_0)$ that supports a Lévy process $\{\xi_t\}_{t \geq 0}$ that possesses $\mathbb{P}_0$-exponential moments. The space of admissible moments is the set $A = \{w \in \mathbb{R} : \mathbb{E}^{\mathbb{P}_0}[\exp(w\xi_t)] < \infty\}$. The associated Lévy exponent $\psi(\alpha) = t^{-1} \ln \mathbb{E}^{\mathbb{P}_0}[\exp(\alpha\xi_t)]$ then exists for all $\alpha \in A_{\mathbb{C}} := \{w \in \mathbb{C} : \operatorname{Re} w \in A\}$, and does not depend on t. A parametric family of measure changes $\mathbb{P}_0 \to \mathbb{P}_\lambda$ commonly called Esscher transformations

can be constructed by use of the exponential martingale family $\{\rho_t^\lambda\}_{t\geq 0}$, defined for each $\lambda \in A$ by $\rho_t^\lambda = \exp(\lambda \xi_t - \psi(\lambda)t)$. If $\{\xi_t\}$ is a $\mathbb{P}_0$-Brownian motion, then $\{\xi_t\}$ is $\mathbb{P}_\lambda$-Brownian with drift λ; if $\{\xi_t\}$ is a $\mathbb{P}_0$-Poisson process with intensity m, then $\{\xi_t\}$ is $\mathbb{P}_\lambda$-Poisson with intensity $\mathrm{e}^\lambda m$; if $\{\xi_t\}$ is a $\mathbb{P}_0$-gamma process with rate parameter m and scale parameter κ, then $\{\xi_t\}$ is $\mathbb{P}_\lambda$-gamma with rate parameter m and scale parameter $\kappa/(1-\lambda)$. Each case is different in character. A natural generalisation of the Esscher transform results when the parameter λ in the measure change is replaced by a random variable X. From the perspective of the new measure $\mathbb{P}_X$, the process $\{\xi_t\}$ retains the "noisy" character of its $\mathbb{P}_0$-Lévy origin, but also carries information about X. In particular, if one assumes that X and $\{\xi_t\}$ are $\mathbb{P}_0$-independent, and that the support of X lies in A, then we say that $\{\xi_t\}$ defines a *Lévy information process* under $\mathbb{P}_X$ carrying the message X. Thus, the change of measure inextricably intertwines signal and noise. More abstractly, we say that on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ a random process $\{\xi_t\}$ is a Lévy information process with message (or "signal") X and noise type (or "fiducial exponent") $\psi_0(\alpha)$ if $\{\xi_t\}$ is *conditionally* a $\mathbb{P}$-Lévy given X, with Lévy exponent $\psi_0(\alpha+X) - \psi_0(X)$ for $\alpha \in \mathbb{C}^\mathrm{I} := \{w \in \mathbb{C} : \mathrm{Re}\, w = 0\}$. We are thus able to classify Lévy information processes by their noise type, and for each noise type we can specify the class of random variables that are admissible as signals that can be carried in the environment of such noise. We consider a number of different noise types, and construct explicit representations of the associated information processes. We also derive an expression for the optimal filter in the general situation, which transforms the *a priori* distribution of the signal to the improved *a posteriori* distribution that can be inferred on the basis of received information.

The plan of the paper is as follows. In Section II, after recalling some facts about processes with stationary and independent increments, we define Lévy information, and in Proposition 1 we show that the signal carried by a Lévy information process is effectively "revealed" after the passage of sufficient time. In Section III we present in Proposition 2 an explicit construction using a change of measure technique that ensures the existence of Lévy information processes, and in Proposition 3 we prove a converse to the effect that any Lévy information process can be obtained in this way. In Proposition 4 we construct the optimal filter for general Lévy information processes, and in Proposition 5 we show that such processes have the Markov property. In Proposition 6 we establish a result that indicates in more detail how the information content of the signal is coded into the structure of an information process. Then in Proposition 7 we present a general construction of the so-called innovations process associated with Lévy information. Finally in Section IV we proceed to examine a number of specific examples of Lévy information processes, for which explicit representations are constructed in Propositions 8–15.

II.   LÉVY INFORMATION

We assume that the reader is familiar with the theory of Lévy processes (Bingham 1975, Sato 1999, Appelbaum 2004, Bertoin 2004, Protter 2005, Kyprianou 2006). For an overview of some of the specific Lévy processes considered later in this paper we refer the reader to Schoutens (2003). A real-valued process $\{\xi_t\}_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a Lévy process if: (i) $\mathbb{P}(\xi_0 = 0) = 1$, (ii) $\{\xi_t\}$ has stationary and independent increments, (iii) $\lim_{t\to s} \mathbb{P}(|\xi_t - \xi_s| > \epsilon) = 0$, and (iv) $\{\xi_t\}$ is almost surely càdlàg. For a Lévy process $\{\xi_t\}$ to give rise to a class of information processes, we require that it should possess exponential

moments. Let us consider the set defined for some (equivalently for all) $t > 0$ by

$$A = \left\{ w \in \mathbb{R} \; : \; \mathbb{E}^{\mathbb{P}}[\exp(w\xi_t)] < \infty \right\}. \tag{1}$$

If A contains points other than $w = 0$, then we say that $\{\xi_t\}$ possesses exponential moments. We define a function $\psi : A \to \mathbb{R}$ called the Lévy exponent (or cumulant function), such that

$$\mathbb{E}^{\mathbb{P}} \left[\exp(\alpha\, \xi_t) \right] = \exp(\psi(\alpha)\, t) \tag{2}$$

for $\alpha \in A$. If a Lévy process possesses exponential moments, then an exercise shows that $\psi(\alpha)$ is convex on A, that the mean and variance of ξ_t are given respectively by $\psi'(0)\, t$ and $\psi''(0)\, t$, and that as a consequence of the convexity of $\psi(\alpha)$ the marginal exponent $\psi'(\alpha)$ possesses a unique inverse $I(y)$ such that $I(\psi'(\alpha)) = \alpha$ for $\alpha \in A$. The Lévy exponent extends to a function $\psi : A_{\mathbb{C}} \to \mathbb{C}$ where $A_{\mathbb{C}} = \{w \in \mathbb{C} \; : \; \operatorname{Re} w \in A\}$, and it can be shown (Sato 1999, Theorem 25.17) that $\psi(\alpha)$ admits a Lévy-Khintchine representation of the form

$$\psi(\alpha) = p\alpha + \frac{1}{2}q\alpha^2 + \int_{\mathbb{R}\backslash\{0\}} (\mathrm{e}^{\alpha z} - 1 - \alpha z \mathbb{1}\{|z| < 1\})\nu(\mathrm{d}z) \tag{3}$$

with the property that (2) holds for for all $\alpha \in A_{\mathbb{C}}$. Here $\mathbb{1}\{\cdot\}$ denotes the indicator function, $p \in \mathbb{R}$ and $q \geq 0$ are constants, and the so-called Lévy measure $\nu(\mathrm{d}z)$ is a positive measure defined on $\mathbb{R}\backslash\{0\}$ satisfying

$$\int_{\mathbb{R}\backslash\{0\}} (1 \wedge z^2)\nu(\mathrm{d}z) < \infty. \tag{4}$$

If the Lévy process possesses exponential moments, then for $\alpha \in A$ we also have

$$\int_{\mathbb{R}\backslash\{0\}} \mathrm{e}^{\alpha z} \mathbb{1}\{|z| \geq 1\}\, \nu(\mathrm{d}z) < \infty. \tag{5}$$

The Lévy measure has the following interpretation: if B is a measurable subset of $\mathbb{R}\backslash\{0\}$, then $\nu(B)$ is the rate at which jumps arrive for which the jump size lies in B. Consider the sets defined for $n \in \mathbb{N}$ by $B_n = \{z \in \mathbb{R} \,|\, 1/n \leq |z| \leq 1\}$. If $\nu(B_n)$ tends to infinity for large n we say that $\{\xi_t\}$ is a process of infinite activity, meaning that the rate of arrival of small jumps is unbounded. If $\nu(\mathbb{R}\backslash\{0\}) < \infty$ one says that $\{\xi_t\}$ has finite activity. We refer to the data $K = (p, q, \nu)$ as the characteristic triplet (or "characteristic") of the associated Lévy process. Thus we can classify a Lévy process abstractly by its characteristic K, or, equivalently, its exponent $\psi(\alpha)$. This means one can speak of a "type" of Lévy noise by reference to the associated characteristic or exponent.

Now suppose we fix a measure $\mathbb{P}_0$ on a measurable space $(\Omega, \mathcal{F})$, and let $\{\xi_t\}$ be $\mathbb{P}_0$-Lévy, with exponent $\psi_0(\alpha)$. There exists a parametric family of probability measures $\{\mathbb{P}_\lambda\}_{\lambda \in A}$ on $(\Omega, \mathcal{F})$ such that for each choice of λ the process $\{\xi_t\}$ is $\mathbb{P}_\lambda$-Lévy. The changes of measure arising in this way are called Esscher transformations (Esscher 1932, Gerber & Shiu 1994, Chan 1999, Kallsen & Shiryaev 2002, Hubalek & Sgarra 2006). Under an Esscher transformation the characteristics of a Lévy process are transformed from one type to another, and one can speak of a "family" of Lévy processes interrelated by Esscher transformations. The relevant change of measure can be specified by use of the process $\{\rho_t^\lambda\}$ defined for $\lambda \in A$ by

$$\rho_t^\lambda := \frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}_0}\bigg|_{\mathcal{F}_t} = \exp\left(\lambda\xi_t - \psi_0(\lambda)t\right), \tag{6}$$

where $\mathcal{F}_t = \sigma\left[\{\xi_s\}_{0 \le s \le t}\right]$. One can check that $\{\rho_t^\lambda\}$ is an $(\{\mathcal{F}_t\}, \mathbb{P}_0)$-martingale: indeed, as a consequence of the fact that $\{\xi_t\}$ has stationary and independent increments we have

$$\mathbb{E}_s^{\mathbb{P}_0}[\rho_t^\lambda] = \mathbb{E}_s^{\mathbb{P}_0}[\mathrm{e}^{\lambda(\xi_t - \xi_s)}]\,\mathrm{e}^{\lambda \xi_s - t\psi_0(\lambda)} = \rho_s^\lambda \tag{7}$$

for $s \le t$, where $\mathbb{E}_t^{\mathbb{P}_0}[\,\cdot\,]$ denotes conditional expectation under $\mathbb{P}_0$ with respect to $\mathcal{F}_t$. It is straightforward to show that $\{\xi_t\}$ has $\mathbb{P}_\lambda$-stationary and independent increments, and that the $\mathbb{P}_\lambda$-exponent of $\{\xi_t\}$, which is defined on the set $A_{\mathbb{C}}^\lambda := \{w \in \mathbb{C} \,|\, \mathrm{Re}\, w + \lambda \in A\}$, is given by

$$\psi_\lambda(\alpha) := t^{-1} \ln \mathbb{E}^{\mathbb{P}_\lambda}[\exp(\alpha \xi_t)] = \psi_0(\alpha + \lambda) - \psi_0(\lambda), \tag{8}$$

from which by use of the Lévy-Khintchine representation (3) one can work out the characteristic triplet K_λ of $\{\xi_t\}$ under $\mathbb{P}_\lambda$. We observe that if the Esscher martingale (6) is expanded as a power series in λ, then the resulting coefficients, which are given by polynomials in ξ_t and t, form a so-called Sheffer set (Schoutens & Teugels 1998), each element of which defines an $(\{\mathcal{F}_t\}, \mathbb{P}_0)$-martingale. The first three of these polynomials take the form $Q^1(x, t) = x - \psi' t$, $Q^2(x, t) = \frac{1}{2}[(x - \psi' t)^2 - \psi'' t]$, and $Q^3(x, t) = \frac{1}{6}[(x - \psi' t)^3 - 3\psi'' t(x - \psi' t) - \psi''' t]$, where $\psi' = \psi_0'(0)$, $\psi'' = \psi_0''(0)$, and $\psi''' = \psi_0'''(0)$. The corresponding polynomial Lévy-Sheffer martingales are given by $Q_t^1 = Q^1(\xi_t, t)$, $Q_t^2 = Q^2(\xi_t, t)$, and $Q_t^3 = Q^3(\xi_t, t)$.

In what follows we use the terms "signal" and "message" interchangeably. We write $\mathbb{C}^{\mathrm{I}} = \{w \in \mathbb{C} : \mathrm{Re}\, w = 0\}$. For any random variable Z on $(\Omega, \mathcal{F}, \mathbb{P})$ we write $\mathcal{F}^Z = \sigma[Z]$, and when it is convenient we write $\mathbb{E}^{\mathbb{P}}[\,\cdot\,|Z]$ for $\mathbb{E}^{\mathbb{P}}[\,\cdot\,|\mathcal{F}^Z]$. For processes we use both of the notations $\{Z_t\}$ and $\{Z(t)\}$, depending on the context.

With these background remarks in mind, we are in a position to define a *Lévy information process*. We confine the discussion to the case of a "simple" message, represented by a random variable X. In the situation when the noise is Brownian motion, the information admits a linear decomposition into signal and noise. In the general situation the relation between signal and noise is more subtle, and has the character of a fibre space, where one thinks of the points of the base space as representing the different noise types, and the points of the fibres as corresponding to the different information processes that one can construct in association with a given noise type. Alternatively, one can think of the base as being the convex space of Lévy characteristics, and the fibre over a given point of the base as the convex space of messages that are compatible with the associated noise type.

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and an Esscher family of Lévy characteristics K_λ, $\lambda \in A$, with associated Lévy exponents $\psi_\lambda(\alpha)$, $\alpha \in A_{\mathbb{C}}^\lambda$. We refer to K_0 as the fiducial characteristic, and $\psi_0(\alpha)$ as the fiducial exponent. The intuition here is that the abstract Lévy process of characteristic K_0 and exponent $\psi_0(\alpha)$, which we call the "fiducial" process, represents the noise type of the associated information process. Thus we can use K_0, or equivalently $\psi_0(\alpha)$, to label the noise type.

Definition 1 *By a Lévy information process with fiducial characteristic K_0, carrying the message X, we mean a random process $\{\xi_t\}$, together with a random variable X, such that $\{\xi_t\}$ is conditionally K_X-Lévy given $\mathcal{F}^X$.*

Thus, given $\mathcal{F}^X$ we require $\{\xi_t\}$ to have conditionally independent and stationary increments under $\mathbb{P}$, and to possess a conditional exponent of the form

$$\psi_X(\alpha) := t^{-1} \ln \mathbb{E}^{\mathbb{P}}[\exp(\alpha \xi_t) \,|\, \mathcal{F}^X] = \psi_0(\alpha + X) - \psi_0(X) \tag{9}$$

for $\alpha \in \mathbb{C}^{\mathrm{I}}$, where $\psi_0(\alpha)$ is the fiducial exponent of the specified noise type. It is implicit in the statement of Definition 1 that a certain compatibility condition holds between the message and the noise type. For any random variable X we define its support S_X to be the smallest closed set F with the property that $\mathbb{P}(X \in F) = 1$. Then we say that X is compatible with the fiducial exponent $\psi_0(\alpha)$ if $S_X \subset A$. Intuitively speaking, the compatibility condition ensures that we can use X to make a random Esscher transformation. In the theory of signal processing, it is advantageous to require that the variables to be estimated should be square integrable. This condition ensures that the conditional expectation exists and admits the interpretation as a best estimate in the sense of least squares. For our purpose it will suffice to assume throughout the paper that the information process is square integrable under $\mathbb{P}$. This in turn implies that $\psi'(X)$ is square integrable, and that $\psi''(X)$ is integrable. Note that we do not require that the Lévy information process should possess exponential moments under $\mathbb{P}$, but a sufficient condition for this to be the case is that there should exist a nonvanishing real number ϵ such that $\lambda + \epsilon \in A$ for all $\lambda \in S_X$.

To gain a better understanding of the sense in which the information process $\{\xi_t\}$ actually "carries" the message X, it will be useful to investigate its asymptotic behaviour. We write $I_0(y)$ for the inverse marginal fiducial exponent.

Proposition 1 *Let $\{\xi_t\}$ be a Lévy information process with fiducial exponent $\psi_0(\alpha)$ and message X. Then for every $\epsilon > 0$ we have*

$$\lim_{t \to \infty} \mathbb{P}\big[|I_0(t^{-1}\xi_t) - X| \geq \epsilon\big] = 0. \tag{10}$$

Proof. It follows from (9) that $\psi'_X(0) = \psi'_0(X)$, and hence that at any time t the conditional mean of the random variable $t^{-1}\xi_t$ is given by

$$\mathbb{E}^{\mathbb{P}}\big[t^{-1}\xi_t \,|\, \mathcal{F}^X\big] = \psi'_0(X). \tag{11}$$

A calculation then shows that the conditional variance of $t^{-1}\xi_t$ takes the form

$$\mathrm{Var}^{\mathbb{P}}\big[t^{-1}\xi_t \,|\, \mathcal{F}^X\big] := \mathbb{E}^{\mathbb{P}}\left[\left(t^{-1}\xi_t - \psi'_0(X)\right)^2 \,\Big|\, \mathcal{F}^X\right] = \frac{1}{t}\,\psi''_0(X), \tag{12}$$

which allows us to conclude that

$$\mathbb{E}^{\mathbb{P}}\left[\left(t^{-1}\xi_t - \psi'_0(X)\right)^2\right] = \frac{1}{t}\,\mathbb{E}^{\mathbb{P}}\left[\psi''_0(X)\right], \tag{13}$$

and hence that

$$\lim_{t \to \infty} \mathbb{E}^{\mathbb{P}}\left[\left(t^{-1}\xi_t - \psi'_0(X)\right)^2\right] = 0. \tag{14}$$

On the other hand for all $\epsilon > 0$ we have

$$\mathbb{P}[\,|t^{-1}\xi_t - \psi'_0(X)| \geq \epsilon\,] \leq \frac{1}{\epsilon^2}\,\mathbb{E}^{\mathbb{P}}\left[\left(t^{-1}\xi_t - \psi'_0(X)\right)^2\right] \tag{15}$$

by Chebychev's inequality, from which we deduce that

$$\lim_{t \to \infty} \mathbb{P}[\,|t^{-1}\xi_t - \psi'_0(X)| \geq \epsilon\,] = 0, \tag{16}$$

and it follows that $I_0(t^{-1}\xi_t)$ converges to X in probability. $\qquad\square$

Thus we see that the information process does indeed carry information about the message, and in the long run "reveals" it. The intuition here is that as more information is gained we improve our estimate of X to the point that the value of X eventually becomes known with near certainty.

III.   PROPERTIES OF LÉVY INFORMATION

It will be useful if we present a construction that ensures the existence of Lévy information processes. First we select a noise type by specification of a fiducial characteristic K_0. Next we introduce a probability space $(\Omega, \mathcal{F}, \mathbb{P}_0)$ that supports the existence of a $\mathbb{P}_0$-Lévy process $\{\xi_t\}$ with the given fiducial characteristic, together with an independent random variable X that is compatible with K_0.

Write $\{\mathcal{F}_t\}$ for the filtration generated by $\{\xi_t\}$, and $\{\mathcal{G}_t\}$ for the filtration generated by $\{\xi_t\}$ and X jointly: $\mathcal{G}_t = \sigma[\{\xi_t\}_{0 \leq s \leq t}, X]$. Let $\psi_0(\alpha)$ be the fiducial exponent associated with K_0. One can check that the process $\{\rho_t^X\}$ defined by

$$\rho_t^X = \exp\left(X\xi_t - \psi_0(X)\,t\right) \tag{17}$$

is a $(\{\mathcal{G}_t\}, \mathbb{P}_0)$-martingale. We are thus able to introduce a change of measure $\mathbb{P}_0 \to \mathbb{P}_X$ on $(\Omega, \mathcal{F}, \mathbb{P}_0)$ by setting

$$\left.\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\mathbb{P}_0}\right|_{\mathcal{G}_t} = \rho_t^X. \tag{18}$$

It should be evident that $\{\xi_t\}$ is conditionally $\mathbb{P}_X$-Lévy given $\mathcal{F}^X$, since for fixed X the measure change is an Esscher transformation. In particular, a calculation shows that the conditional exponent of ξ_t under $\mathbb{P}_X$ is given by

$$t^{-1}\ln\mathbb{E}^{\mathbb{P}_X}\left[\exp(\alpha\xi_t)\,|\,\mathcal{F}^X\right] = \psi_0(\alpha + X) - \psi_0(X) \tag{19}$$

for $\alpha \in \mathbb{C}^{\mathrm{I}}$, which shows that the conditions of Definition 1 are satisfied, allowing us to conclude the following:

Proposition 2 *The $\mathbb{P}_0$-Lévy process $\{\xi_t\}$ is a $\mathbb{P}_X$-Lévy information process, with message X and noise type $\psi_0(\alpha)$.*

In fact, the converse also holds: if we are given a Lévy information process, then by a change of measure we can find a Lévy process and an independent "message" variable. Here follows a more precise statement.

Proposition 3 *Let $\{\xi_t\}$ be a Lévy information process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with message X and noise type $\psi_0(\alpha)$. Then there exists a change of measure $\mathbb{P} \to \mathbb{P}_0$ such that $\{\xi_t\}$ and X are $\mathbb{P}_0$-independent, $\{\xi_t\}$ is $\mathbb{P}_0$-Lévy with exponent $\psi_0(\alpha)$, and the probability law of X under $\mathbb{P}_0$ is the same as probability law of X under $\mathbb{P}$.*

Proof. First we establish that the process $\{\tilde{\rho}_t^X\}$ defined by the expression $\tilde{\rho}_t^X = \exp(-X\xi_t + \psi_0(X)t)$ is a $(\{\mathcal{G}_t\}, \mathbb{P})$-martingale. We have

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}}[\tilde{\rho}_t^X | \mathcal{G}_s] &= \mathbb{E}^{\mathbb{P}}\left[\exp(-X\xi_t + \psi_0(X)t)\,|\,\mathcal{G}_s\right] \\
&= \mathbb{E}^{\mathbb{P}}[\exp(-X(\xi_t - \xi_s))|\mathcal{G}_s]\exp(-X\xi_s + \psi_0(X)t) \\
&= \exp(\psi_X(-X)(t-s))\exp(-X\xi_s + \psi_0(X)t)
\end{aligned} \tag{20}
$$

by virtue of the fact that $\{\xi_t\}$ is $\mathcal{F}^X$-conditionally Lévy under $\mathbb{P}$. By use of (9) we deduce that $\psi_X(-X) = -\psi_0(X)$, and hence that $\mathbb{E}^{\mathbb{P}}[\tilde{\rho}_t^X | \mathcal{G}_s] = \tilde{\rho}_s^X$, as required. Then we use $\{\tilde{\rho}_t^X\}$ to define a change of measure $\mathbb{P} \to \mathbb{P}_0$ on $(\Omega, \mathcal{F}, \mathbb{P})$ by setting

$$\left.\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{G}_t} = \tilde{\rho}_t^X. \tag{21}$$

To show that ξ_t and X are $\mathbb{P}_0$-independent for all t, it suffices to show that their joint characteristic function under $\mathbb{P}_0$ factorises. Letting $\alpha, \beta \in \mathbb{C}^I$, we have

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}_0}[\exp(\alpha\xi_t + \beta X)] &= \mathbb{E}^{\mathbb{P}}\left[\exp(-X\xi_t + \psi_0(X)t)\exp(\alpha\xi_t + \beta X)\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}[\exp((-X + \alpha)\xi_t + \psi_0(X)t + \beta X)|\mathcal{F}^X]\right] \\
&= \mathbb{E}^{\mathbb{P}}[\exp(\psi_X(-X + \alpha)t + \psi_0(X)t + \beta X)] \\
&= \exp(\psi_0(\alpha)t)\,\mathbb{E}^{\mathbb{P}}[\exp(\beta X)],
\end{aligned}
\tag{22}
$$

where the last step follows from (9). This argument can be extended to show that $\{\xi_t\}$ and X are $\mathbb{P}_0$-independent. Next we observe that

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}_0}&[\exp(\alpha(\xi_u - \xi_t) + \beta\xi_t)] \\
&= \mathbb{E}^{\mathbb{P}}\left[\exp(-X\xi_u + \psi_0(X)u + \alpha(\xi_u - \xi_t) + \beta\xi_t)\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}[\exp(-X\xi_u + \psi_0(X)u + \alpha(\xi_u - \xi_t) + \beta\xi_t)\,|\mathcal{F}^X]\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}[\exp(\psi_0(X)u + (\alpha - X)(\xi_u - \xi_t) + (\beta - X)\xi_t)\,|\mathcal{F}^X]\right] \\
&= \mathbb{E}^{\mathbb{P}}[\exp(\psi_0(X)u + \psi_X(\alpha - X)(u - t) + \psi_X(\beta - X)t)] \\
&= \exp(\psi_0(\alpha)(u - t))\,\exp(\psi_0(\beta)t)
\end{aligned}
\tag{23}
$$

for $u \geq t \geq 0$, and it follows that $\xi_u - \xi_t$ and ξ_t are independent. This argument can be extended to show that $\{\xi_t\}$ has $\mathbb{P}_0$-independent increments. Finally, if we set $\alpha = 0$ in (22) it follows that the probability laws of X under $\mathbb{P}_0$ and $\mathbb{P}$ are identical; if we set $\beta = 0$ in (22) it follows that the $\mathbb{P}_0$ exponent of $\{\xi_t\}$ is $\psi_0(\alpha)$; and if we set $\beta = 0$ in (23) it follows that $\{\xi_t\}$ is $\mathbb{P}_0$-stationary. $\qquad\square$

Going forward, we adopt the convention that $\mathbb{P}$ always denotes the "physical" measure in relation to which an information process with message X is defined, and that $\mathbb{P}_0$ denotes the transformed measure with respect to which the information process and the message decouple. Therefore, henceforth we write $\mathbb{P}$ rather than $\mathbb{P}_X$. In addition to establishing the existence of Lévy information processes, the results of Proposition 3 provide useful tools for calculations, allowing us to work out properties of information processes by referring the calculations back to $\mathbb{P}_0$. We consider as an example the problem of working out the $\mathcal{F}_t$-conditional expectation under $\mathbb{P}$ of a $\mathcal{G}_t$-measurable integrable random variable Z. The $\mathbb{P}$-expectation of Z can be written in terms of $\mathbb{P}_0$-expectations, and is given by a "generalised Bayes formula" (Kallianpur & Striebel 1968) of the form

$$
\mathbb{E}^{\mathbb{P}}[Z\,|\,\mathcal{F}_t] = \frac{\mathbb{E}^{\mathbb{P}_0}\left[\rho_t^X Z\,|\,\mathcal{F}_t\right]}{\mathbb{E}^{\mathbb{P}_0}[\rho_t^X\,|\,\mathcal{F}_t]}.
\tag{24}
$$

This formula can be used to obtain the $\mathcal{F}_t$-conditional probability distribution function for X, defined for $y \in \mathbb{R}$ by

$$
F_t^X(y) = \mathbb{P}(X \leq y\,|\,\mathcal{F}_t).
\tag{25}
$$

In the Bayes formula we set $Z = \mathbb{1}\{X \leq y\}$, and the result is

$$
F_t^X(y) = \frac{\int \mathbb{1}\{x \leq y\}\exp(x\xi_t - \psi_0(x)t)\,\mathrm{d}F^X(x)}{\int \exp(x\xi_t - \psi_0(x)t)\,\mathrm{d}F^X(x)},
\tag{26}
$$

where $F^X(y) = \mathbb{P}(X < y)$ is the *a priori* distribution function. It is useful for some purposes to work directly with the conditional probability measure $\pi_t(\mathrm{d}x)$ induced on $\mathbb{R}$ defined by $\mathrm{d}F_t^X(x) = \pi_t(\mathrm{d}x)$. In particular, when X is a continuous random variable with a density function $p(x)$ one can write $\pi_t(\mathrm{d}x) = p_t(x)\mathrm{d}x$, where $p_t(x)$ is the conditional density function.

Proposition 4 *Let $\{\xi_t\}$ be a Lévy information process under $\mathbb{P}$ with noise type $\psi_0(\alpha)$, and let the* a priori *distribution of the associated message X be $\pi(\mathrm{d}x)$. Then the $\mathcal{F}_t$-conditional* a posteriori *distribution of X is*

$$\pi_t(\mathrm{d}x) = \frac{\exp\left(x\xi_t - \psi_0(x)t\right)}{\int \exp\left(x\xi_t - \psi_0(x)t\right)\pi(\mathrm{d}x)}\,\pi(\mathrm{d}x). \tag{27}$$

It is straightforward to establish by use of a variational argument that for any function $f : \mathbb{R} \to \mathbb{R}$ such that the random variable $Y = f(X)$ is integrable, the best estimate for Y conditional on the information $\mathcal{F}_t$ is given by

$$\hat{Y}_t := \mathbb{E}^{\mathbb{P}}[Y \,|\, \mathcal{F}_t] = \int f(x)\,\pi_t(\mathrm{d}x). \tag{28}$$

By the "best estimate" for Y we mean the $\mathcal{F}_t$-measurable random variable $\hat{Y}_t$ that minimises the quadratic error $\mathbb{E}^{\mathbb{P}}[(Y - \hat{Y}_t)^2|\mathcal{F}_t]$.

It will be observed that at any given time t the best estimate can be expressed as a function of ξ_t and t, and does not involve values of the information process at times earlier than t. That this should be the case can be seen as a consequence of the following:

Proposition 5 *The Lévy information process $\{\xi_t\}$ has the Markov property.*

Proof. For the Markov property it suffices to establish that for $a \in \mathbb{R}$ we have

$$\mathbb{P}\left(\xi_t \leq a \,|\, \mathcal{F}_s\right) = \mathbb{P}\left(\xi_t \leq a \,|\, \mathcal{F}^{\xi_s}\right), \tag{29}$$

where $\mathcal{F}_t = \sigma[\{\xi_s\}_{0 \leq s \leq t}]$ and $\mathcal{F}^{\xi_t} = \sigma[\xi_t]$. We write

$$\Phi_t := \mathbb{E}^{\mathbb{P}_0}\left[\rho_t^X|\mathcal{F}_t\right] = \int \exp\left(x\xi_t - \psi_0(x)t\right)\pi(\mathrm{d}x), \tag{30}$$

where ρ_t^X is defined as in equation (17). It follows that

$$\mathbb{P}\left(\xi_t \leq a \,|\, \mathcal{F}_s\right) = \mathbb{E}^{\mathbb{P}}[\mathbb{1}\{\xi_t \leq a\} \,|\, \mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}_0}[\Phi_t \mathbb{1}\{\xi_t \leq a\} \,|\, \mathcal{F}_s]}{\mathbb{E}^{\mathbb{P}_0}[\Phi_t \,|\, \mathcal{F}_s]}$$

$$= \frac{\mathbb{E}^{\mathbb{P}_0}[\Phi_t \mathbb{1}\{\xi_t \leq a\} \,|\, \mathcal{F}^{\xi_s}]}{\mathbb{E}^{\mathbb{P}_0}[\Phi_t \,|\, \mathcal{F}^{\xi_s}]} = \mathbb{E}^{\mathbb{P}}[\mathbb{1}\{\xi_t \leq a\} \,|\, \mathcal{F}^{\xi_s}] = \mathbb{P}\left(\xi_t \leq a \,|\, \mathcal{F}^{\xi_s}\right), \tag{31}$$

since $\{\xi_t\}$ has the Markov property under the transformed measure $\mathbb{P}_0$. $\qquad\square$

We note that since X is $\mathcal{F}_\infty$-measurable, which follows from Proposition 1, the Markov property implies that if $Y = f(X)$ is integrable we have

$$\mathbb{E}^{\mathbb{P}}[Y|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[Y \,|\, \mathcal{F}^{\xi_t}]. \tag{32}$$

This identity allows one to work out the optimal filter for a Lévy information process by direct use of the Bayes formula. It should be apparent that simulation of the dynamics of the filter is readily approachable on account of this property.

We remark briefly on what might appropriately be called a "time consistency" property satisfied by Lévy information processes. It follows from (27) that, given the conditional distribution $\pi_s(\mathrm{d}x)$ at time $s \leq t$, we can express $\pi_t(\mathrm{d}x)$ in the form

$$\pi_t(\mathrm{d}x) = \frac{\exp\big(x(\xi_t - \xi_s) - \psi_0(x)(t-s)\big)}{\int \exp\big(x(\xi_t - \xi_s) - \psi_0(x)(t-s)\big)\pi_s(\mathrm{d}x)} \, \pi_s(\mathrm{d}x). \tag{33}$$

Then if for fixed $s \geq 0$ we introduce a new time variable $u := t - s$, and define $\eta_u = \xi_{u+s} - \xi_s$, we find that $\{\eta_u\}_{u \geq 0}$ is an information process with fiducial exponent $\psi_0(\alpha)$ and message X with *a priori* distribution $\pi_s(\mathrm{d}x)$. Thus given up-to-date information we can "re-start" the information process at that time to produce a new information process of the same type, with an adjusted message distribution.

Further insight into the nature of Lévy information can be gained by examination of expression (9) for the conditional exponent of an information process. In particular, as a consequence of the Lévy-Khintchine representation (3) we are able to deduce that

$$\psi_0(\alpha + X) - \psi_0(X) = \left(p + qX + \int_{\mathbb{R}\backslash\{0\}} z(\mathrm{e}^{Xz} - 1)\mathbb{1}\{|z| < 1\})\nu(\mathrm{d}z) \right)\alpha + \tfrac{1}{2}q\alpha^2$$
$$+ \int_{\mathbb{R}\backslash\{0\}} (\mathrm{e}^{\alpha z} - 1 - \alpha z\mathbb{1}\{|z| < 1\})\mathrm{e}^{Xz}\nu(\mathrm{d}z), \tag{34}$$

for $\alpha \in \mathbb{C}^{\mathrm{I}}$, which leads to the following:

Proposition 6 *The randomisation of the $\mathbb{P}_0$-Lévy process $\{\xi_t\}$ achieved through the change of measure generated by the randomised Esscher martingale $\rho_t = \exp(X\xi_t - \psi_0(X)t)$ induces two effects on the characteristics of the process:* (i) *a random shift in the drift term, given by*

$$p \to p + qX + \int_{\mathbb{R}\backslash\{0\}} z(\mathrm{e}^{Xz} - 1)\mathbb{1}\{|z| < 1\})\nu(\mathrm{d}z), \tag{35}$$

and (ii) *a random rescaling of the Lévy measure, given by* $\nu(\mathrm{d}z) \to \mathrm{e}^{Xz}\nu(\mathrm{d}z)$.

The integral appearing in the shift in the drift term is well defined since the term $z(\mathrm{e}^{Xz} - 1)$ vanishes to second order at the origin. It follows from Proposition 6 that in sampling an information process an agent is in effect trying to detect a random shift in the drift term, and a random "tilt" and change of scale in the Lévy measure, altering the overall rate as well as the relative rates at which jumps of various sizes occur. It is from these data, within which the message is encoded, that the agent attempts to estimate the value of X. It is interesting to note that randomised Esscher martingales arise in the construction of pricing kernels in the theory of finance (see, e.g., Shefrin 2008, Macrina & Parbhoo 2011).

We turn to examine the properties of certain martingales associated with Lévy information. We establish the existence of a so-called innovations representation for Lévy information. In the case of the Brownian filter the ideas involved are rather well understood (see, e.g., Liptser & Shiryaev 2000), and the matter has also been investigated in the case of Poisson information (Segall & Kailath 1975). These examples arise as special cases in the general theory of Lévy information. Throughout the discussion that follows we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Proposition 7 *Let $\{\xi_t\}$ be a Lévy information process with fiducial exponent $\psi_0(\alpha)$ and message X, let $\{\mathcal{F}_t\}$ denote the filtration generated by $\{\xi_t\}$, let $Y = \psi_0'(X)$, where $\psi_0'(\alpha)$ is the marginal fiducial exponent, and set $\hat{Y}_t = \mathbb{E}^{\mathbb{P}}[Y|\mathcal{F}_t]$. Then the process $\{M_t\}$ defined by*

$$\xi_t = \int_0^t \hat{Y}_u \,\mathrm{d}u + M_t \tag{36}$$

is an $(\{\mathcal{F}_t\}, \mathbb{P})$-martingale.

Proof. We recall that $\{\xi_t\}$ is by definition $\mathcal{F}^X$-conditionally $\mathbb{P}$-Lévy. It follows therefore from (11) that $\mathbb{E}^{\mathbb{P}}[\xi_t|X] = Yt$, where $Y = \psi_0'(X)$. As before we let $\{\mathcal{G}_t\}$ denote the filtration generated jointly by $\{\xi_t\}$ and X. First we observe that the process defined for $t \geq 0$ by $m_t = \xi_t - Yt$ is a $(\{\mathcal{G}_t\}, \mathbb{P})$-martingale. This assertion can be checked by consideration of the one-parameter family of $(\{\mathcal{G}_t\}, \mathbb{P}_0)$-martingales defined by

$$\rho_t^{X+\epsilon} = \exp\left((X + \varepsilon)\xi_t - \psi_0(X + \varepsilon)t\right) \tag{37}$$

for $\epsilon \in \mathbb{C}^{\mathrm{I}}$. Expanding this expression to first order in ϵ, we deduce that the process defined for $t \geq 0$ by $\rho_t^X(\xi_t - \psi_0'(X)t)$ is a $(\{\mathcal{G}_t\}, \mathbb{P}_0)$-martingale. Thus we have

$$\mathbb{E}^{\mathbb{P}_0}\left[\rho_t^X(\xi_t - \psi_0'(X)t) \,|\, \mathcal{G}_s\right] = \rho_s^X(\xi_s - \psi_0'(X)s). \tag{38}$$

Then using $\{\rho_t^X\}$ to make a change of measure from $\mathbb{P}_0$ to $\mathbb{P}$ we obtain

$$\mathbb{E}^{\mathbb{P}}\left[\xi_t - \psi_0'(X)t \,|\, \mathcal{G}_s\right] = \xi_s - \psi_0'(X)s, \tag{39}$$

and the result follows if we set $Y = \psi_0'(X)$. Next we introduce the "projected" process $\{\hat{m}_t\}$ defined by $\hat{m}_t = \mathbb{E}^{\mathbb{P}}[m_t \,|\, \mathcal{F}_t]$. We note that since $\{m_t\}$ is a $(\{\mathcal{G}_t\}, \mathbb{P})$-martingale we have

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}[\hat{m}_t|\mathcal{F}_s] &= \mathbb{E}^{\mathbb{P}}[\xi_t - Yt \,|\, \mathcal{F}_s] \\
&= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}[\xi_t - Yt \,|\, \mathcal{G}_s]|\mathcal{F}_s\right] \\
&= \mathbb{E}^{\mathbb{P}}[\xi_s - Ys \,|\, \mathcal{F}_s] \\
&= \hat{m}_s,
\end{aligned} \tag{40}$$

and thus $\{\hat{m}_t\}$ is an $(\{\mathcal{F}_t\}, \mathbb{P})$-martingale. Finally we observe that

$$\mathbb{E}^{\mathbb{P}}[M_t|\mathcal{F}_s] = \mathbb{E}^{\mathbb{P}}\left[\xi_t - \int_0^t \hat{Y}_u \,\mathrm{d}u \,\middle|\, \mathcal{F}_s\right] = \mathbb{E}^{\mathbb{P}}[\xi_t|\mathcal{F}_s] - \mathbb{E}^{\mathbb{P}}\left[\int_s^t \hat{Y}_u \,\mathrm{d}u \,\middle|\, \mathcal{F}_s\right] - \int_0^s \hat{Y}_u \,\mathrm{d}u, \tag{41}$$

where we have made use of the fact that the final term is $\mathcal{F}_s$-measurable. The fact that $\{\hat{m}_t\}$ and $\{\hat{Y}_t\}$ are both $(\mathcal{F}_t, \mathbb{P})$-martingales implies that

$$\mathbb{E}^{\mathbb{P}}[\xi_t|\mathcal{F}_s] - \xi_s = (t - s)\hat{Y}_s = \mathbb{E}^{\mathbb{P}}\left[\int_s^t \hat{Y}_u \,\mathrm{d}u \,\middle|\, \mathcal{F}_s\right], \tag{42}$$

from which it follows that $\mathbb{E}^{\mathbb{P}}\left[M_t \,\middle|\, \mathcal{F}_s\right] = M_s$, which is what we set out to prove. $\square$

Although the general information process does not admit an additive decomposition into signal and noise, it does admit a linear decomposition into terms representing (i) information already received and (ii) new information. The random variable Y entering via its conditional expectation into the first of these terms is itself in general a nonlinear function of the message variable X. It follows on account of the convexity of the fiducial exponent that the marginal fiducial exponent is invertible, which ensures that X can be expressed in terms of Y by the relation $X = I_0(Y)$, which is linear if and only if the information process is Brownian. Thus signal and noise are deeply intertwined in the case of general Lévy information. Vestiges of linearity remain, and these suffice to provide an overall element of tractability.

IV.   EXAMPLES OF LÉVY INFORMATION PROCESSES

In a number of situations one can construct explicit examples of information processes, categorised by noise type. The Brownian and Poisson constructions, which are familiar in other contexts, can be seen as belonging to a unified scheme that brings out their differences and similarities. We then proceed to construct information processes of the gamma, the variance gamma, the negative binomial, the inverse Gaussian, and the normal inverse Gaussian type. It is interesting to take note of the diverse nature of noise, and to observe the many different ways in which messages can be conveyed in a noisy environment.

Example 1: Brownian information. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{B_t\}$ be a Brownian motion, let X be an independent random variable, and set

$$\xi_t = Xt + B_t. \tag{43}$$

The random process $\{\xi_t\}$ thereby defined, which we call the Brownian information process, is $\mathcal{F}^X$-conditionally K_X-Lévy, with conditional characteristic $K_X = (X, 1, 0)$ and conditional exponent $\psi_X(\alpha) = X\alpha + \frac{1}{2}\alpha^2$. The fiducial characteristic is $K_0 = (0, 1, 0)$, the fiducial exponent is $\psi_0(\alpha) = \frac{1}{2}\alpha^2$, and the associated fiducial process or "noise type" is standard Brownian motion. In the case of Brownian information, there is a linear separation of the process into signal and noise. This model, considered by Wonham (1965), is perhaps the simplest continuous-time generalisation of the example described by Wiener (1948). The message is given by the value of X, but X can only be observed indirectly, through $\{\xi_t\}$. The observations of X are obscured by the noise represented by the Brownian motion $\{B_t\}$. Since the signal term grows linearly in time, whereas $|B_t| \sim \sqrt{t}$, it is intuitively plausible that observations of $\{\xi_t\}$ will asymptotically reveal the value of X, and a direct calculation using properties of the normal distribution function confirms that $t^{-1}\xi_t$ converges in probability to X; this is consistent with Proposition 1 if we note that $\psi_0'(\alpha) = \alpha$ and $I_0(y) = y$ in the Brownian case.

The best estimate for X conditional on $\mathcal{F}_t$ is given by (28), which can be derived by use of the generalised Bayes formula (24). In the Brownian case there is an elementary method leading to the same result, worth mentioning briefly since it is of interest. First we present an alternative proof of Proposition 5 in the Brownian case that uses a Brownian bridge argument.

We recall that if $s > s_1 > 0$ then B_s and $s^{-1}B_s - s_1^{-1}B_{s_1}$ are independent. More generally, we observe that if $s > s_1 > s_2$, then B_s, $s^{-1}B_s - s_1^{-1}B_{s_1}$, and $s_1^{-1}B_{s_1} - s_2^{-1}B_{s_2}$ are independent, and that $s^{-1}\xi_s - s_1^{-1}\xi_{s_1} = s^{-1}B_s - s_1^{-1}B_{s_1}$. Extending this line of reasoning, we see that for any $a \in \mathbb{R}$ we have

$$\mathbb{P}\left(\xi_t \le a \,|\, \xi_s, \xi_{s_1}, \ldots, \xi_{s_k}\right) = \mathbb{P}\left(\xi_t \le a \,\Big|\, \xi_s, \frac{\xi_s}{s} - \frac{\xi_{s_1}}{s_1}, \ldots, \frac{\xi_{s_{k-1}}}{s_{k-1}} - \frac{\xi_{s_k}}{s_k}\right)$$
$$= \mathbb{P}\left(\xi_t \le a \,\Big|\, \xi_s\right), \tag{44}$$

since ξ_t and ξ_s are independent of $s^{-1}B_s - s_1^{-1}B_{s_1}, \ldots, s_{k-1}^{-1}B_{s_{k-1}} - s_k^{-1}B_{s_k}$, and that gives us the Markov property (29). Since we have established that X is $\mathcal{F}_\infty$-measurable, it follows that (32) holds. As a consequence, the *a posteriori* distribution of X can be worked out by

use of the standard Bayes formula, and for the best estimate of X we obtain

$$\hat{X}_t = \frac{\int x \, \exp(x\xi_t - \frac{1}{2}x^2 t) \, \pi(\mathrm{d}x)}{\int \exp(x\xi_t - \frac{1}{2}x^2 t) \, \pi(\mathrm{d}x)}. \tag{45}$$

The innovations representation (36) in the case of a Brownian information process can be derived by the following argument. We observe that the $(\{\mathcal{F}_t\}, \mathbb{P}_0)$-martingale $\{\Phi_t\}$ defined in (30) is a "space-time" function of the form

$$\Phi_t := \mathbb{E}^{\mathbb{P}_0}[\rho_t \mid \mathcal{F}_t] = \int \exp\left(x\xi_t - \frac{1}{2}x^2 t\right) \pi(\mathrm{d}x). \tag{46}$$

By use of the Ito calculus together with (45), we deduce that $\mathrm{d}\Phi_t = \hat{X}_t \, \Phi_t \, \mathrm{d}\xi_t$, and thus by integration we obtain

$$\Phi_t = \exp\left(\int_0^t \hat{X}_s \mathrm{d}\xi_s - \frac{1}{2}\int_0^t \hat{X}_s^2 \mathrm{d}s\right). \tag{47}$$

Since $\{\xi_t\}$ is an $(\{\mathcal{F}_t\}, \mathbb{P}_0)$-Brownian motion, it follows from (47) by the Girsanov theorem that the process $\{M_t\}$ defined by

$$\xi_t = \int_0^t \hat{X}_s \, \mathrm{d}s + M_t \tag{48}$$

is an $(\{\mathcal{F}_t\}, \mathbb{P})$-Brownian motion, which we call the innovations process (see, e.g., Heunis 2011). The increments of $\{M_t\}$ represent the arrival of new information.

We conclude our discussion of Brownian information with the following remarks. In problems involving prediction and valuation, it is not uncommon that the message is revealed after the passage of a finite amount of time. This is often the case in applications to finance, where the message takes the form of a random cash flow at some future date, or, more generally, a random factor that affects such a cash flow. There are also numerous examples coming from the physical sciences, economics and operations research where the goal of an agent is to form a view concerning the outcome of a future event by monitoring the flow of information relating to it. How does one handle problems involving the revelation of information over finite time horizons?

One way of modelling finite time horizon scenarios in the present context is by use of a time change. If $\{\xi_t\}$ is a Lévy information process with message X and a specified fiducial exponent, then a generalisation of Proposition 1 shows that the process $\{\xi_{tT}\}$ defined over the time interval $0 \le t < T$ by

$$\xi_{tT} = \frac{T-t}{T}\xi\left(\frac{tT}{T-t}\right) \tag{49}$$

reveals the value of X in the limit as $t \to T$, and one can check that

$$\mathrm{Cov}\left[\xi_{sT}, \xi_{tT} \mid \mathcal{F}^X\right] = \frac{s(T-t)}{T} \, \psi_0''(X), \quad (0 \le s \le t < T). \tag{50}$$

In the case where $\{\xi_t\}$ is a Brownian information process represented as above in the form $\xi_t = Xt + B_t$, the time-changed process (49) takes the form $\xi_{tT} = Xt + \beta_{tT}$, where $\{\beta_{tT}\}$ is

a Brownian bridge over the interval $[0, T]$. Such processes have had applications in physics (Brody & Hughston 2005, 2006; see also Adler *et al.* 2001, Brody & Hughston 2002) and in finance (Brody *et al.* 2007, 2008a, Rutkowski & Yu 2007, Brody *et al.* 2009, Filipović *et al.* 2012). It seems reasonable to conjecture that time-changed Lévy information processes of the more general type proposed above may be similarly applicable.

Example 2: Poisson information. Consider a situation in which an agent observes a series of events taking place at a random rate, and the agent wishes to determine the rate as best as possible since its value conveys an important piece of information. One can model the information flow in this situation by a modulated Poisson process for which the jump rate is an independent random variable. Such a scenario arises in many real-world situations, and has been investigated in the literature (Segall & Kailath 1975, Segall *et al.* 1975, Brémaud 1981, Di Masi & Runggaldier 1983, Kailath & Poor 1998). The Segall-Kailath scheme can be seen to emerge naturally as an example of our general model for Lévy information.

As in the Brownian case, one can construct the relevant information process directly. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{N(t)\}_{t \geq 0}$ be a standard Poisson process with jump rate $m > 0$, let X be an independent random variable, and set

$$\xi_t = N(\mathrm{e}^X t). \tag{51}$$

Thus $\{\xi_t\}$ is a time-changed Poisson process, and the effect of the signal is to randomly modulate the rate at which the process jumps. It is evident that $\{\xi_t\}$ is $\mathcal{F}^X$-conditionally Lévy and satisfies the conditions of Definition 1. In particular,

$$\mathbb{E}\left[\exp\left(\alpha N(\mathrm{e}^X t)\right) \mid \mathcal{F}^X\right] = \exp\left(m\mathrm{e}^X(\mathrm{e}^\alpha - 1) t\right), \tag{52}$$

and for fixed X one obtains a Poisson process with rate $m\mathrm{e}^X$. It follows that (51) is an information process. The fiducial characteristic is given by $K_0 = (0, 0, m\delta_1(\mathrm{d}z))$, that of a Poisson process with unit jumps at the rate m, where $\delta_1(\mathrm{d}z)$ is the Dirac measure with unit mass at $z = 1$, and the fiducial exponent is $\psi_0(\alpha) = m(\mathrm{e}^\alpha - 1)$. A calculation using (9) shows that $K_X = (0, 0, m\mathrm{e}^X \delta_1(\mathrm{d}z))$, and that $\psi_X(\alpha) = m\mathrm{e}^X(\mathrm{e}^\alpha - 1)$. The relation between signal and noise in the case of Poisson information is rather subtle. The noise is associated with the random fluctuations of the inter-arrival times of the jumps, whereas the message determines the average rate at which the jumps occur.

It will be instructive in this example to work out the conditional distribution of X by elementary methods. Since X is $\mathcal{F}_\infty$-measurable and $\{\xi_t\}$ has the Markov property, we have

$$F_t^X(y) := \mathbb{P}(X \leq y \mid \mathcal{F}_t) = \mathbb{P}(X \leq y \mid \xi_t) \tag{53}$$

for $y \in \mathbb{R}$. It follows then from the Bayes law for an information process taking values in $\mathbb{N}_0$ that

$$\mathbb{P}(X \leq y \mid \xi_t = n) = \frac{\int \mathbb{1}\{x \leq y\} \mathbb{P}(\xi_t = n \mid X = x) \, \mathrm{d}F^X(x)}{\int \mathbb{P}(\xi_t = n \mid X = x) \, \mathrm{d}F^X(x)}. \tag{54}$$

In the case of Poisson information the relevant conditional distribution is

$$\mathbb{P}(\xi_t = n \mid X = x) = \exp(-mt\mathrm{e}^x) \frac{(mt\mathrm{e}^x)^n}{n!}. \tag{55}$$

After some cancellation we deduce that

$$\mathbb{P}(X \le y \,|\, \xi_t = n) = \frac{\int \mathbb{1}\{x \le y\} \exp(xn - m(\mathrm{e}^x - 1)t) \,\mathrm{d}F^X(x)}{\int \exp(xn - m(\mathrm{e}^x - 1)t) \,\mathrm{d}F^X(x)}, \tag{56}$$

and hence

$$F_t^X(y) = \frac{\int \mathbb{1}\{x \le y\} \exp(x\xi_t - m(\mathrm{e}^x - 1)t) \,\mathrm{d}F^X(x)}{\int \exp(x\xi_t - m(\mathrm{e}^x - 1)t) \,\mathrm{d}F^X(x)}, \tag{57}$$

and thus

$$\pi_t(\mathrm{d}x) = \frac{\exp(x\xi_t - m(\mathrm{e}^x - 1)t)}{\int \exp(x\xi_t - m(\mathrm{e}^x - 1)t) \,\pi(\mathrm{d}x)} \pi(\mathrm{d}x), \tag{58}$$

which we can see is consistent with (27) if we recall that in the case of noise of the Poisson type the fiducial exponent is given by $\psi_0(\alpha) = m(\mathrm{e}^\alpha - 1)$.

If a Geiger counter is monitored continuously in time, the sound that it produces provides a nice example of a Poisson information process. The crucial message (proximity to radioactivity) carried by the noisy sputter of the instrument is represented by the rate at which the clicks occur.

Example 3: Gamma information. It will be convenient first to recall a few definitions and conventions (cf. Yor 2007, Brody *et al.* 2008b, Brody *et al.* 2012). Let m and κ be positive numbers. By a gamma process with rate m and scale κ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we mean a Lévy process $\{\gamma_t\}_{t \ge 0}$ with exponent

$$t^{-1} \ln \mathbb{E}^{\mathbb{P}}\left[\exp(\alpha\gamma_t)\right] = -m \ln(1 - \kappa\alpha) \tag{59}$$

for $\alpha \in A_{\mathbb{C}} = \{w \in \mathbb{C} \,|\, \mathrm{Re}\, w < \kappa^{-1}\}$. The probability density for γ_t is

$$\mathbb{P}(\gamma_t \in \mathrm{d}x) = \mathbb{1}\{x > 0\} \frac{\kappa^{-mt} x^{mt-1} \exp\left(-x/\kappa\right)}{\Gamma[mt]} \,\mathrm{d}x, \tag{60}$$

where $\Gamma[a]$ is the gamma function. A short calculation making use of the functional equation $\Gamma[a + 1] = a\Gamma[a]$ shows that $\mathbb{E}^{\mathbb{P}}[\gamma_t] = m\kappa t$ and $\mathrm{Var}^{\mathbb{P}}[\gamma_t] = m\kappa^2 t$. Clearly, the mean and variance determine the rate and scale. If $\kappa = 1$ we say that $\{\gamma_t\}$ is a *standard* gamma process with rate m. If $\kappa \ne 1$ we say that $\{\gamma_t\}$ is a scaled gamma process. The Lévy measure associated with the gamma process is

$$\nu(\mathrm{d}z) = \mathbb{1}\{z > 0\} \, m \, z^{-1} \exp(-\kappa z) \,\mathrm{d}z. \tag{61}$$

It follows that $\nu(\mathbb{R}\backslash\{0\}) = \infty$ and hence that the gamma process has infinite activity. Now let $\{\xi_t\}$ be a standard gamma process with rate m on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_0)$, and let $\lambda \in \mathbb{R}$ satisfy $\lambda < 1$. Then the process $\{\rho_t^\lambda\}$ defined by

$$\rho_t^\lambda = (1 - \lambda)^{mt} \mathrm{e}^{\lambda\gamma_t} \tag{62}$$

is an $(\{\mathcal{F}_t\}, \mathbb{P}_0)$-martingale. If we let $\{\rho_t^\lambda\}$ act as a change of measure density for the transformation $\mathbb{P}_0 \to \mathbb{P}_\lambda$, then we find that $\{\gamma_t\}$ is a *scaled* gamma process under $\mathbb{P}_\lambda$, with rate m and scale $1/(1 - \lambda)$. Thus we see that the effect of an Esscher transformation on a gamma process is to alter its scale. With these facts in mind, one can establish the following:

Proposition 8 *Let $\{\gamma_t\}$ be a standard gamma process with rate m on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let the independent random variable X satisfy $X < 1$ almost surely. Then the process $\{\xi_t\}$ defined by*

$$\xi_t = \frac{1}{1-X}\,\gamma_t \tag{63}$$

is a Lévy information process with message X and gamma noise, with fiducial exponent $\psi_0(\alpha) = -m\ln(1-\alpha)$ for $\alpha \in \{w \in \mathbb{C} \,|\, \mathrm{Re}\,w < 1\}$.

Proof. It is evident that $\{\xi_t\}$ is $\mathcal{F}^X$-conditionally a scaled gamma process. As a consequence of (59) we have

$$\frac{1}{t}\ln\mathbb{E}^{\mathbb{P}}\left[\exp(\alpha\xi_t)|X\right] = \frac{1}{t}\ln\mathbb{E}^{\mathbb{P}}\left[\exp\left(\frac{\alpha\gamma_t}{1-X}\right)\Big|X\right] = \ln\left(1 - \frac{\alpha}{1-X}\right)^{-m} \tag{64}$$

for $\alpha \in \mathbb{C}^{\mathrm{I}}$. Then we note that

$$-m\ln\left(1 - \frac{\alpha}{1-X}\right) = -m\ln\left(1 - (X+\alpha)\right) + m\ln\left(1 - X\right). \tag{65}$$

It follows that the $\mathcal{F}^X$-conditional $\mathbb{P}$ exponent of $\{\xi_t\}$ is $\psi_0(X + \alpha) - \psi_0(X)$. $\qquad\square$

The gamma filter arises as follows. An agent observes a process of accumulation. Typically there are many small increments, but now and then there are large increments. The rate at which the process is growing is the figure that the agent wishes to estimate as accurately as possible. The accumulation can be modelled by gamma information, and the associated filter can be used to estimate the growth rate. It has long been recognised that the gamma process is useful in describing phenomena such as the water level of a dam or the totality of the claims made in a large portfolio of insurance contracts (Gani 1957, Kendall 1957, Gani & Pyke 1960). Use of the gamma information process and related bridge processes, with applications in finance and insurance, is pursued in Brody *et al.* (2008b), Hoyle (2010), and Hoyle *et al.* (2011). We draw the reader's attention to Yor (2007) and references cited therein, where it is shown how certain additive properties of Brownian motion have multiplicative analogues in the case of the gamma process. One notes in particular the remarkable property that γ_t and γ_s/γ_t are independent for $t \geq s \geq 0$. Making use of this relation, it will be instructive to present an alternative derivation of the optimal filter for gamma noise. We begin by establishing that the process defined by (63) has the Markov property. We observe first that for any times $t \geq s \geq s_1 \geq s_2 \geq \cdots \geq s_k$ the variables $\gamma_{s_1}/\gamma_s, \gamma_{s_2}/\gamma_{s_1}$, and so on, are independent of one another and are independent of γ_s and γ_t. It follows that

$$\begin{aligned}
\mathbb{P}\left(\xi_t < a | \xi_s, \xi_{s_1}, \ldots, \xi_{s_k}\right) &= \mathbb{P}\left(\xi_t < a | (1-X)^{-1}\gamma_s, \ldots, (1-X)^{-1}\gamma_{s_k}\right)\\
&= \mathbb{P}\left(\xi_t < a \,\Big|\, (1-X)^{-1}\gamma_s, \frac{\gamma_{s_1}}{\gamma_s}, \frac{\gamma_{s_2}}{\gamma_{s_1}}, \ldots, \frac{\gamma_{s_k}}{\gamma_{s_{k-1}}}\right)\\
&= \mathbb{P}\left(\xi_t < a \,|\, \xi_s\right),
\end{aligned} \tag{66}$$

since $\{\gamma_t\}$ and X are independent, and that gives us (29). In working out the distribution of X given $\mathcal{F}_t$ it suffices therefore to work out the distribution of X given ξ_t. We note that the Bayes formula implies that

$$\pi_t(\mathrm{d}x) = \frac{\rho(\xi_t|X = x)}{\int \rho(\xi_t|X = x)\,\pi(\mathrm{d}x)}\,\pi(\mathrm{d}x), \tag{67}$$

where $\pi(\mathrm{d}x)$ is the unconditional distribution of X, and $\rho(\xi|X = x)$ is the conditional density for the random variable ξ_t, which can be calculated as follows:

$$\rho(\xi|X = x) = \frac{\mathrm{d}}{\mathrm{d}\xi} \mathbb{P}(\xi_t \leq \xi|X = x) = \frac{\mathrm{d}}{\mathrm{d}\xi} \mathbb{P}((1 - X)^{-1}\gamma_t \leq \xi|X = x)$$

$$= \frac{\mathrm{d}}{\mathrm{d}\xi} \mathbb{P}\left(\gamma_t \leq (1 - X)\xi\middle| X = x\right) = \frac{\xi^{mt-1}(1 - x)^{mt}\mathrm{e}^{-(1-x)\xi}}{\Gamma[mt]}. \tag{68}$$

It follows that the optimal filter in the case of gamma noise is given by

$$\pi_t(\mathrm{d}x) = \frac{(1 - x)^{mt} \exp(x\xi_t)}{\int_{-\infty}^{1}(1 - x)^{mt} \exp(x\xi_t)\pi(\mathrm{d}x)} \pi(\mathrm{d}x), \tag{69}$$

We conclude with the following observation. In the case of Brownian information, it is well known (and implicit in the example of Wiener 1948) that if the signal is Gaussian, then the optimal filter is a linear function of the observation ξ_t. One might therefore ask in the case of a gamma information process if some special choice of the signal distribution gives rise to a linear filter. The answer is affirmative. Let U be a gamma-distributed random variable with the distribution

$$\mathbb{P}(U \in \mathrm{d}u) = \mathbb{1}\{u > 0\} \frac{\theta^r u^{r-1} \exp(-\theta u)}{\Gamma[r]} \mathrm{d}u, \tag{70}$$

where $r > 1$ and $\theta > 0$ are parameters, and set $X = 1 - U$. Let $\{\xi_t\}$ be a gamma information process carrying message X, let $Y = \psi_0'(X) = m/(1 - X)$, and set $\tau = (r - 1)/m$. Then the optimal filter for Y is given by

$$\hat{Y}_t := \mathbb{E}^{\mathbb{P}}[Y|\mathcal{F}_t] = \frac{\xi_t + \theta}{t + \tau}. \tag{71}$$

Example 4: Variance-gamma information. The so-called variance-gamma or VG process (Madan & Seneta 1990, Madan & Milne 1991, Madan *et al.* 1998) was introduced in the theory of finance. The relevant definitions and conventions are as follows. By a VG process with drift $\mu \in \mathbb{R}$, volatility $\sigma \geq 0$, and rate $m > 0$, we mean a Lévy process with exponent

$$\psi(\alpha) = -m \ln\left(1 - \frac{\mu}{m}\alpha - \frac{\sigma^2}{2m}\alpha^2\right). \tag{72}$$

The VG process admits representations in terms of simpler Lévy processes. Let $\{\gamma_t\}$ be a standard gamma process on $(\Omega, \mathcal{F}, \mathbb{P})$, with rate m, as defined in the previous example, and let $\{B_t\}$ be a standard Brownian motion, independent of $\{\gamma_t\}$. We call the scaled process $\{\Gamma_t\}$ defined by $\Gamma_t = m^{-1}\gamma_t$ a gamma subordinator with rate m. Note that Γ_t has dimensions of time and that $\mathbb{E}^{\mathbb{P}}[\Gamma_t] = t$. A calculation shows that the Lévy process $\{V_t\}$ defined by

$$V_t = \mu\Gamma_t + \sigma B_{\Gamma_t} \tag{73}$$

has the exponent (72). The VG process thus takes the form of a Brownian motion with drift, time-changed by a gamma subordinator. If $\mu = 0$ and $\sigma = 1$, we say that $\{V_t\}$ is a "standard" VG process, with rate parameter m. If $\mu \neq 0$, we say that $\{V_t\}$ is a "drifted"

VG process. One can always choose units of time such that $m = 1$, but for applications it is better to choose conventional units of time (seconds for physics, years for economics), and treat m as a model parameter. In the limit $\sigma \to 0$ we obtain a gamma process with rate m and scale μ/m. In the limit $m \to \infty$ we obtain a Brownian motion with drift μ and volatility σ.

An alternative representation of the VG process results if we let $\{\gamma_t^1\}$ and $\{\gamma_t^2\}$ be independent standard gamma processes on $(\Omega, \mathcal{F}, \mathbb{P})$, with rate m, and set

$$V_t = \kappa_1 \gamma_t^1 - \kappa_2 \gamma_t^2, \tag{74}$$

where κ_1 and κ_2 are nonnegative constants. A calculation shows that the exponent is of the form (72). In particular, we have

$$\psi(\alpha) = -m \ln \left(1 - (\kappa_1 - \kappa_2)\,\alpha - \kappa_1 \kappa_2\,\alpha^2\right), \tag{75}$$

where $\mu = m(\kappa_1 - \kappa_2)$ and $\sigma^2 = 2m\kappa_1\kappa_2$, or equivalently

$$\kappa_1 = \frac{1}{2m}\left(\mu + \sqrt{\mu^2 + 2m\sigma^2}\right) \quad \text{and} \quad \kappa_2 = \frac{1}{2m}\left(-\mu + \sqrt{\mu^2 + 2m\sigma^2}\right), \tag{76}$$

where $\alpha \in \{w \in \mathbb{C} : -1/\kappa_2 < \operatorname{Re} w < 1/\kappa_1\}$. Now let $\{\xi_t\}$ be a standard VG process on $(\Omega, \mathcal{F}, \mathbb{P}_0)$, with exponent $\psi_0(\alpha) = -m \ln(1 - (2m)^{-1}\alpha^2)$ for $\alpha \in \{w \in \mathbb{C} : |\operatorname{Re} w| < \sqrt{2m}\}$. Under the transformed measure $\mathbb{P}_\lambda$ defined by the change-of-measure martingale (6), one finds that $\{\xi_t\}$ is a drifted VG process, with

$$\mu = \lambda \left(1 - \frac{1}{2m}\lambda^2\right)^{-1} \quad \text{and} \quad \sigma = \left(1 - \frac{1}{2m}\lambda^2\right)^{-\frac{1}{2}} \tag{77}$$

for $|\lambda| < \sqrt{2m}$. Thus in the case of the VG process an Esscher transformation affects both the drift and the volatility. Note that for large m the effect on the volatility is insignificant, whereas the effect on the drift reduces to that of an ordinary Girsanov transformation.

With these facts in hand, we are now in a position to construct the VG information process. We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a number $m > 0$.

Proposition 9 *Let $\{\Gamma_t\}$ be a standard gamma subordinator with rate m, let $\{B_t\}$ be an independent Brownian motion, and let the independent random variable X satisfy $|X| < \sqrt{2m}$ almost surely. Then the process $\{\xi_t\}$ defined by*

$$\xi_t = X \left(1 - \frac{1}{2m}X^2\right)^{-1} \Gamma_t + \left(1 - \frac{1}{2m}X^2\right)^{-\frac{1}{2}} B(\Gamma_t) \tag{78}$$

is a Lévy information process with message X and VG noise, with fiducial exponent

$$\psi_0(\alpha) = -m \ln \left(1 - \frac{1}{2m}\alpha^2\right) \tag{79}$$

for $\alpha \in \{w \in \mathbb{C} : \operatorname{Re} w < \sqrt{2m}\}$.

Proof. Observe that $\{\xi_t\}$ is $\mathcal{F}^X$-conditionally a drifted VG process of the form

$$\xi_t = \mu_X \Gamma_t + \sigma_X B(\Gamma_t), \tag{80}$$

where the drift and volatility coefficients are

$$\mu_X = X \left(1 - \frac{1}{2m} X^2\right)^{-1} \quad \text{and} \quad \sigma_X = \left(1 - \frac{1}{2m} X^2\right)^{-\frac{1}{2}}. \tag{81}$$

The $\mathcal{F}^X$-conditional $\mathbb{P}$-exponent of $\{\xi_t\}$ is by (72) thus given for $\alpha \in \mathbb{C}^I$ by

$$
\begin{aligned}
\psi_X(\alpha) &= -m \ln\left(1 - \frac{1}{m}\mu_X\,\alpha - \frac{1}{2m}\sigma_X^2\,\alpha^2\right) \\
&= -m \ln\left(1 - \frac{1}{m}X\left(1 - \frac{1}{2m}X^2\right)^{-1}\alpha - \frac{1}{2m}\left(1 - \frac{1}{2m}X^2\right)^{-1}\alpha^2\right) \\
&= -m \ln\left(1 - \frac{1}{2m}(X+\alpha)^2\right) + m \ln\left(1 - \frac{1}{2m}X^2\right),
\end{aligned}
\tag{82}
$$

which is evidently by (79) of the form $\psi_0(X+\alpha) - \psi_0(X)$, as required. $\qquad\square$

An alternative representation for the VG information process can be established by the same method if one randomly rescales the gamma subordinator appearing in the time-changed Brownian motion. The result is as follows.

Proposition 10 *Let $\{\Gamma_t\}$ be a gamma subordinator with rate m, let $\{B_t\}$ be an independent standard Brownian motion, and let the independent random variable X satisfy $|X| < \sqrt{2m}$ almost surely. Write $\{\Gamma_t^X\}$ for the subordinator:*

$$\Gamma_t^X = \left(1 - \frac{1}{2m}X^2\right)^{-1}\Gamma_t. \tag{83}$$

Then the process $\{\xi_t\}$ defined by $\xi_t = X\Gamma_t^X + B(\Gamma_t^X)$ is a VG information process with message X.

A further representation of the VG information process arises as a consequence of the representation of the VG process as the asymmetric difference between two independent standard gamma processes. In particular, we have:

Proposition 11 *Let $\{\gamma_t^1\}$ and $\{\gamma_t^2\}$ be independent standard gamma processes, each with rate m, and let the independent random variable X satisfy $|X| < \sqrt{2m}$ almost surely. Then the process $\{\xi_t\}$ defined by*

$$\xi_t = \frac{1}{\sqrt{2m} - X}\gamma_t^1 - \frac{1}{\sqrt{2m} + X}\gamma_t^2 \tag{84}$$

is a VG information process with message X.

Example 5: Negative-binomial information. By a negative binomial process with rate parameter m and probability parameter q, where $m > 0$ and $0 < q < 1$, we mean a Lévy process with exponent

$$\psi_0(\alpha) = m \ln \left(\frac{1-q}{1-qe^\alpha} \right) \tag{85}$$

for $\alpha \in \{w \in \mathbb{C} \,|\, \mathrm{Re}\, w < -\ln q\}$. There are two representations for the negative binomial process (Kozubowski & Podgórski 2009; Brody *at al.* 2012). The first of these is a compound Poisson process for which the jump size $J \in \mathbb{N}$ has a logarithmic distribution

$$\mathbb{P}_0(J = n) = -\frac{1}{\ln(1-q)} \, \frac{1}{n} \, q^n \,, \tag{86}$$

and the intensity of the Poisson process determining the timing of the jumps is given by $\lambda = -m \ln(1-q)$. One finds that the characteristic function of J is

$$\phi_0(\alpha) := \mathbb{E}^{\mathbb{P}_0}[\exp(\alpha J)] = \frac{\ln(1 - qe^\alpha)}{\ln(1-q)} \tag{87}$$

for $\alpha \in \{w \in \mathbb{C} \,|\, \mathrm{Re}\, w < -\ln q\}$. Then if we set

$$n_t = \sum_{k=1}^\infty \mathbb{1}\{k \leq N_t\} \, J_k, \tag{88}$$

where $\{N_t\}$ is a Poisson process with rate λ, and $\{J_k\}_{k\in\mathbb{N}}$ denotes a collection of independent identical copies of J, representing the jumps, one deduces that

$$\mathbb{P}_0(n_t = k) = \frac{\Gamma(k + mt)}{\Gamma(mt)\Gamma(k+1)} q^k (1-q)^{mt}, \tag{89}$$

and that the resulting exponent is given by (85). The second representation of the negative binomial process makes use of the method of subordination. We take a Poisson process with rate $\Lambda = mq/(1-q)$, and time-change it using a gamma subordinator $\{\Gamma_t\}$ with rate parameter m. The moment generating function thus obtained, in agreement with (85), is

$$\mathbb{E}^{\mathbb{P}_0}\left[\exp\left(\alpha N(\Gamma_t)\right)\right] = \mathbb{E}^{\mathbb{P}_0}\left[\exp\left(\Lambda(e^\alpha - 1)\Gamma_t\right)\right] = \left(\frac{1-q}{1-qe^\alpha}\right)^{mt}. \tag{90}$$

With these results in mind, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and find:

Proposition 12 *Let $\{\Gamma_t\}$ be a gamma subordinator with rate m, let $\{N_t\}$ be an independent Poisson process with rate m, let the independent random variable X satisfy $X < -\ln q$ almost surely, and set*

$$\Gamma_t^X = \left(\frac{qe^X}{1 - qe^X}\right)\Gamma_t. \tag{91}$$

Then the process $\{\xi_t\}$ defined by

$$\xi_t = N(\Gamma_t^X) \tag{92}$$

is a Lévy information process with message X and negative binomial noise, with fiducial exponent (85).

Proof. This can be verified by direct calculation. For $\alpha \in \mathbb{C}^{\mathrm{I}}$ we have:

$$\mathbb{E}^{\mathbb{P}}\left[e^{\alpha \xi_t}|X\right] = \mathbb{E}^{\mathbb{P}}\left[\exp(\alpha N(\Gamma_t^X))|\,X\right] = \mathbb{E}^{\mathbb{P}}\left[\exp\left(m\frac{qe^X}{1-qe^X}(e^\alpha-1)\Gamma_t\right)\Big|\,X\right]$$

$$= \left(1 - \frac{qe^X\,(e^\alpha-1)}{1-qe^X}\right)^{-mt} = \left(\frac{1-qe^X}{1-qe^{X+\alpha}}\right)^{mt}, \tag{93}$$

which by (85) shows that the conditional exponent is $\psi_0(X+\alpha) - \psi_0(X)$. $\qquad\square$

There is also a representation for negative binomial information based on the compound Poisson process. This can be obtained by an application of Proposition 6, which shows how the Lévy measure transforms under a random Esscher transformation. In the case of a negative binomial process with parameters m and q, the Lévy measure is given by

$$\nu(\mathrm{d}z) = m\sum_{n=1}^\infty \frac{1}{n}\,q^n\,\delta_n(\mathrm{d}z), \tag{94}$$

where $\delta_n(\mathrm{d}z)$ denotes the Dirac measure with unit mass at the point $z = n$. The Lévy measure is finite in this case, and we have $\nu(\mathbb{R}) = -m\ln(1-q)$, which is the overall rate at which the compound Poisson process jumps. If one normalises the Lévy measure with the overall jump rate, one obtains the probability measure (86) for the jump size. With these facts in mind, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and specify the constants m and q, where $m > 1$ and $0 < q < 1$. Then as a consequence of Proposition 6 we have the following:

Proposition 13 *Let the random variable X satisfy $X < -\ln q$ almost surely, let the random variable J^X have the conditional distribution*

$$\mathbb{P}(J^X = n\,|\,X) = -\frac{1}{\ln(1-qe^X)}\,\frac{1}{n}\,(qe^X)^n, \tag{95}$$

let $\{J_k^X\}_{k\in\mathbb{N}}$ be a collection of conditionally independent identical copies of J^X, and let $\{N_t\}$ be an independent Poisson process with rate m. Then the process $\{\xi_t\}$ defined by

$$\xi_t = \sum_{k=1}^\infty \mathbb{1}\{k \le N(-\ln(1-qe^X)t)\}\,J_k^X \tag{96}$$

is a Lévy information process with message X and negative binomial noise, with fiducial exponent (85).

Example 6: Inverse Gaussian information. The inverse Gaussian (IG) distribution appears in the study of the first exit time of Brownian motion with drift (Schrödingier 1915). The name "inverse Gaussian" was introduced by Tweedie (1945), and a Lévy process whose increments have the IG distribution was introduced in Wasan (1968). By an IG process with parameters $a > 0$ and $b > 0$, we mean a Lévy process with exponent

$$\psi_0(\alpha) = a\left(b - \sqrt{b^2 - 2\alpha}\right) \tag{97}$$

for $\alpha \in \{w \in \mathbb{C} \,|\, 0 \le \operatorname{Re} w < \frac{1}{2}b^2\}$. Let us write $\{G_t\}$ for the IG process. The probability density function for G_t is

$$\mathbb{P}_0(G_t \in \mathrm{d}x) = \mathbb{1}\{x > 0\} \frac{at}{\sqrt{2\pi x^3}} \exp\left(-\frac{(bx - at)^2}{2x}\right) \mathrm{d}x, \tag{98}$$

and we find that $\mathbb{E}^{\mathbb{P}_0}[G_t] = at/b$ and that $\operatorname{Var}^{\mathbb{P}_0}[G_t] = at/b^3$. It is straightforward to check that under the Esscher transformation $\mathbb{P}_0 \to \mathbb{P}_\lambda$ induced by (6), where $0 < \lambda < \frac{1}{2}b^2$, the parameter a is left unchanged, whereas $b \to (b^2 - 2\lambda)^{1/2}$. With these facts in mind we are in a position to introduce the associated information process. We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and find the following:

Proposition 14 *Let $G(t)$ be an inverse Gaussian process with parameters a and b, let X be an independent random variable satisfying $0 < X < \frac{1}{2}b^2$ almost surely, and set $Z = b^{-1}(b^2 - 2X)^{1/2}$. Then the process $\{\xi_t\}$ defined by*

$$\xi_t = Z^{-2}G(Zt) \tag{99}$$

is a Lévy information process with message X and inverse Gaussian noise, with fiducial exponent (97).

Proof. It should be evident by inspection that $\{\xi_t\}$ is $\mathcal{F}^X$-conditionally Lévy. Let us therefore work out the conditional exponent. For $\alpha \in \mathbb{C}^{\mathrm{I}}$ we have:

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}} &\left[\exp(\alpha\,\xi_t)|X\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\left.\exp\left(\alpha\,\frac{b^2}{b^2 - 2X}\,G\left(b^{-1}\sqrt{b^2 - 2X}\,t\right)\right)\right| X\right] \\
&= \exp\left(at\left(\sqrt{b^2 - 2X} - \sqrt{b^2 - 2(\alpha + X)}\right)\right) \\
&= \exp\left(at\left(b - \sqrt{b^2 - 2(\alpha + X)}\right) - at\left(b - \sqrt{b^2 - 2X}\right)\right), \tag{100}
\end{aligned}$$

which shows that the conditional exponent is of the form $\psi_0(\alpha + X) - \psi_0(X)$. $\qquad\square$

Example 7: Normal inverse Gaussian information. By a normal inverse Gaussian (NIG) process (Rydberg 1997, Barndorff-Nielsen 1998) with parameters a, b, and m, such that $a > 0$, $|b| < a$, and $m > 0$, we mean a Lévy process with an exponent of the form

$$\psi_0(\alpha) = m\left(\sqrt{a^2 - b^2} - \sqrt{a^2 - (b + \alpha)^2}\right) \tag{101}$$

for $\alpha \in \{w \in \mathbb{C} : -a - b < \operatorname{Re} w < a - b\}$. Let us write $\{I_t\}$ for the NIG process. The probability density for its value at time t is given by

$$\mathbb{P}_0(I_t \in \mathrm{d}x) = \frac{amt K_1\left(a\sqrt{m^2t^2 + x^2}\right)}{\pi\sqrt{m^2t^2 + x^2}} \exp\left(mt\sqrt{a^2 - b^2} + bx\right) \mathrm{d}x, \tag{102}$$

where K_ν is the modified Bessel function of third kind (Erdélyi 1953). The NIG process can be represented as a Brownian motion subordinated by an IG process. In particular, let $\{B_t\}$ be a standard Brownian motion, let $\{G_t\}$ be an independent IG process with parameters a'

and b', and set $a' = 1$ and $b' = m(a^2 - b^2)^{1/2}$. Then the characteristic function of the process $\{I_t\}$ defined by

$$I_t = bm^2 G_t + mB(G_t) \tag{103}$$

is given by (101). The associated information process is constructed as follows. We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the parameters a, b, and m.

Proposition 15 *Let the random variable X satisfy $-a - b < X < a - b$ almost surely, let $\{G_t^X\}$ be $\mathcal{F}^X$-conditionally IG, with parameters $a' = 1$ and $b' = m(a^2 - (b + X)^2)^{1/2}$, and let $F_t = m^2 G_t^X$. Then the process $\{\xi_t\}$ defined by*

$$\xi_t = (b + X)F_t + B(F_t) \tag{104}$$

is a Lévy information process with message X and NIG noise, with fiducial exponent (101).

Proof. We observe that the condition on $\{G_t^X\}$ is that

$$\frac{1}{t} \ln \mathbb{E}^{\mathbb{P}} \left[\exp \left(\alpha G_t^X \right) | X \right] = \delta \sqrt{a^2 - (b + X)^2} - \sqrt{m^2(a^2 - (b + X)^2) - 2\alpha} \tag{105}$$

for $\alpha \in \mathbb{C}^{\mathrm{I}}$. Thus if we set $\psi_X(\alpha) = \mathbb{E}^{\mathbb{P}} \left[\exp(\alpha \xi_t) | X \right]$ for $\alpha \in \mathbb{C}^{\mathrm{I}}$ it follows that

$$\begin{aligned}
\psi_X(\alpha) &= \mathbb{E}^{\mathbb{P}} \left[\exp \left(\alpha(b + X)F_t + \alpha B(F_t) \right) | X \right] \\
&= \mathbb{E}^{\mathbb{P}} \left[\exp \left((\alpha(b + X) + \tfrac{1}{2}\alpha^2)m^2 G_t^X \right) | X \right] \\
&= \mathbb{E}^{\mathbb{P}} \left[\exp \left(mt\sqrt{a^2 - (b + X)^2} - mt\sqrt{a^2 - (b + X)^2 - 2\left(\alpha(b + X) + \tfrac{1}{2}\alpha^2 \right)} \right) \right],
\end{aligned} \tag{106}$$

which shows that the conditional exponent is of the required form. $\qquad\square$

Similar arguments lead to the construction of information processes based on various other Lévy processes related to the IG distribution, including for example the generalised hyperbolic process (Barndorff-Nielsen 1977), for which the information process can be shown to take the form

$$\xi_t = (b + X)\tilde{G}_t + B(\tilde{G}_t). \tag{107}$$

Here the random variable X is taken to be $\mathbb{P}$-independent of the standard Brownian motion $\{B(t)\}$, and $\{\tilde{G}_t\}$ is $\mathcal{F}^X$-conditionally a generalised IG process with parameters $(\delta, (a^2 - (b + X)^2)^{1/2}, \nu)$. It would be of interest to determine whether explicit models can be obtained for information processes based on the Meixner process (Schoutens & Teugels 1998) and the CGMY process (Carr *et al.* 2002, Madan & Yor 2008).

We conclude this study of Lévy information with the following remarks. Recent developments in the phenomenological representation of physical (Brody & Hughston 2006) and economic (Brody *et al.* 2008a) time series have highlighted the idea that signal processing techniques may have far-reaching applications to the identification, characterisation and categorisation of phenomena, both in the natural and in the social sciences, and that beyond the conventional remits of *prediction*, *filtering*, and *smoothing* there is a fourth and important new domain of applicability: the *description* of phenomena in science and in society.

It is our hope therefore that the theory of signal processing with Lévy information herein outlined will find a variety of interesting and exciting applications.

[1] Adler, S. A., Brody, D. C., Brun, T. A. & Hughston, L. P. 2001 Martingale models for quantum state reduction. *J. Phys.* A**34**, 8795–8820. (doi:10.1088/0305-4470/34/42/306)

[2] Ahn, H. & Feldman, R. E. 1999 Optimal filtering of a Gaussian signal in the presence of Lévy noise. *SIAM J. Appl. Math.* **60**, 359–369. (doi:10.1137/S0036139996312004)

[3] Applebaum, D. 2004 *Lévy Processes and Stochastic Calculus* (Cambridge: Cambridge University Press).

[4] Bain, A. & Crisan, D. 2010 *Fundamentals of Stochastic Filtering* (New York: Springer).

[5] Barndorff-Nielsen, O. E. 1977 Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London* A**353**, 401–419. (doi:10.1098/rspa.1977.0041)

[6] Barndorff-Nielsen, O. E. 1998 Processes of normal inverse Gaussian type. *Finance Stochast.* **2**, 41–68. (doi:10.1007/s007800050032)

[7] Bellman, R. E. 1961 *Adaptive Control Processes* (New Jersey: Princeton University Press).

[8] Bertoin, J. 2004 *Lévy Processes* (Cambridge: Cambridge University Press).

[9] Bingham, N. P. 1975 Fluctuation theory in continuous time. *Adv. Appl. Prob.* **7**, 705–766. (www.jstor.org/stable/10.2307/1426397)

[10] Brémaud, P 1981 *Point Processes and Queues: Martingale Dynamics.* (New York: Springer-Verlag).

[11] Brody, D. C. & Hughston, L. P. 2002 Efficient simulation of quantum state reduction. *J. Math. Phys.* **43**, 5254–5261. (doi:10.1063/1.1512975)

[12] Brody, D. C. & Hughston, L. P. 2005 Finite-time stochastic reduction models. *J. Math. Phys.* **46**, 082101. (doi:10.1063/1.1990108)

[13] Brody, D. C. & Hughston, L. P. 2006 Quantum noise and stochastic reduction. *J. Phys.* A**39**, 833–876. (doi:10.1088/0305-4470/39/4/008)

[14] Brody, D. C., Hughston, L. P. & Macrina, A. 2007 Beyond hazard rates: a new framework for credit-risk modelling. In *Advances in Mathematical Finance: Festschrift Volume in Honour of Dilip Madan* (Basel: Birkhäuser). (doi:10.1007/978-0-8176-4545-8_13)

[15] Brody, D. C., Hughston, L. P. & Macrina, A. 2008a Information-based asset pricing. *Int. J. Theoretical Appl. Finance* **11**, 107-142. (doi:10.1142/S0219024908004749) Reprinted as Chapter 5 in *Finance at Fields.* Grasselli, M. R. & Hughston, L. P., eds. (Singapore: World Scientific Publishing, 2012).

[16] Brody, D. C., Hughston, L. P. & Macrina, A. 2008b Dam rain and cumulative gain. *Proc. Roy. Soc. London* A**464**, 1801–1822. (doi:10.1098/rspa.2007.0273)

[17] Brody, D. C., Davis, M. H. A., Friedman, R. L. & Hughston, L. P. 2009 Informed traders. *Proc. Roy. Soc. London* A**465**, 1103–1122. (doi:10.1098/rspa.2008.0465)

[18] Brody, D. C., Hughston, L. P. & Mackie, E. 2012 General theory of geometric Lévy models for dynamic asset pricing. *Proc. Roy. Soc. London* A**468**, 1778–1798. (doi:10.1098/rspa.2011.0670)

[19] Carr, P., Geman, H., Madan, D. & Yor, M. (2002) The fine structure of asset returns: an empirical investigation. *Journal of Business* **75**, 305–332.

[20] Chan, T. 1999 Pricing contingent claims on stocks driven by Lévy processes. *Ann. App. Prob.* **9**, 504–528. (www.jstor.org/stable/2667343)

[21] Davis, M. H. A. 1977 *Linear Estimation and Stochastic Control* (London: Chapman & Hall).

[22] Di Masi, G. B. & Runggaldier, W. J. 1983 Non-linear filtering with discontinuous observations and applications to life sciences. *Bull. Math. Bio.* **45**, 571–577. (doi:10.1007/BF02459588)

[23] Erdélyi, A. 1953 *Higher Transcendental Functions.* Bateman Manuscript Project Vol. 2 (New York: McGraw-Hill).

[24] Esscher, F. 1932. On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift* **15**, 175–195. (doi:10.1080/03461238.1932.10405883)

[25] Filipović, D., Hughston, L. P. & Macrina, A. 2012 Conditional density models for asset pricing. *Int. J. Theoretical Appl. Finance*, **15**, 1250002. (doi:10.1142/S0219024912500021) Reprinted as Chapter 9 in *Finance at Fields.* Grasselli, M. R. & Hughston, L. P., eds. (Singapore: World Scientific Publishing, 2012).

[26] Gani, J. 1957 Problems in the probability theory of storage systems. *J. Roy. Statist. Soc.* B**19**, 181–206. (www.jstor.org/stable/2983814)

[27] Gani, J. & Pyke, R. 1960 The content of a dam as the supremum of an infinitely divisible process. *J. Math. Mech.* (*Indiana Univ. Math. J.*) **9**, 639–651. (doi:10.1512/iumj.1960.9.09038)

[28] Gerber, H. U. & Shiu, E. S. W. 1994 Option pricing by Esscher transforms (with discussion). *Trans. Soc. Actuaries* **46**, 99–191.

[29] Grigelionis, B & Mikulevicius, R. 2011 Nonlinear filtering equations for stochastic processes with jumps. In *The Oxford Handbook of Nonlinear Filtering*, Crisan, D. & Rozovskii, B., eds. (Oxford: Oxford University Press).

[30] Heunis, A. J. 2011 The innovation problem. In *The Oxford Handbook of Nonlinear Filtering*, Crisan, D. & Rozovskii, B., eds. (Oxford: Oxford University Press).

[31] Hoyle, E. 2010 Information-Based Models for Finance and Insurance. PhD thesis, Department of Mathematics, Imperial College London. (arXiv:1010.0829)

[32] Hoyle, E., Hughston, L. P. & Macrina, A. 2011 Lévy random bridges and the modelling of financial information. *Stochast. Process. App.* **121**, 856–884. (doi:10.1016/j.spa.2010.12.003)

[33] Hubalek, F. & Sgarra, C. 2006 Esscher transform and the minimal entropy martingale measure for exponential Lévy models. *Quant. Finance* **6**, 125–145. (doi:10.1080/14697680600573099)

[34] Kailath, T. 1974 A view of three decades of linear filtering theory. *IEEE Trans. Info. Theo.* **20**, 146–181. (doi:10.1109/TIT.1974.1055174)

[35] Kailath, T. & Poor, H. V. 1998 Detection of stochastic processes. *IEEE Trans. Info. Theo.* **44**, 2230–2259. (doi:10.1109/18.720538)

[36] Kallianpur, G. & Striebel, C. 1968 Estimation of stochastic systems: arbitrary system process with additive white noise observation errors. *Ann. Math. Stat.* **39**, 785–801. (doi:10.1214/aoms/1177698311)

[37] Kalman, R. E. 1994 Randomness reexamined. *Modelling, Ident. Cont.* **15**, 141–151. (doi:10.4173/mic.1994.3.3)

[38] Kallsen, J. & Shiryaev, A. N. 2002 The cumulant process and Esscher's change of measure. *Fin. Stochastics* **6** 97–428. (doi:10.1007/s007800200069)

[39] Kendall, D. G. 1957 Some problems in the theory of dams. *J. Roy. Statist. Soc.* B**19**, 207–212. (www.jstor.org/stable/2983815)

[40] Kozubowski, T. J. & Podgórski, K. 2009 Distributional properties of the negative binomial Lévy process. *Probability and Mathematical Statistics* **29**, 43–71. (www.math.uni.wroc.pl/∼pms/files/29.1/Article/29.1.3.pdf)

[41] Kyprianou, A. E. 2006 *Introductory Lectures on Fluctuations of Lévy Processes with Applications* (Berlin: Springer).

[42] Liptser, R. S. & Shiryaev, A. N. 2000 *Statistics of Random Processes* Vols. I and II, 2nd ed. (Berlin: Springer).

[43] Macrina, A. & Parbhoo, P. A. 2011 Randomised mixture models for pricing kernels. (arXiv:1112.2059).

[44] Madan, D. B., Carr, P., & Chang, E. C. (1998) The variance gamma process and option pricing. *European Fin. Rev.* **2**, 79–105. (doi:10.1023/A:1009703431535)

[45] Madan, D. B. & Milne, F. 1991 Option pricing with V.G. martingale components. *Mathematical Finance* **1**, 39–55. (doi:10.1111/j.1467-9965.1991.tb00018.x)

[46] Madan, D. B. & Seneta, E. 1990 The variance gamma (V.G.) model for share market returns. *Journal of Business* **63**, 511–524. (doi:10.1086/296519)

[47] Madan, D. B. & Yor, M. 2008 Representing the CGMY and Meixner Lévy processes as time changed Brownian motion. *Journal of Computational Finance* **12**, 27–47.

[48] Meyer-Brandis, T. & Proske, F. 2004 Explicit solution of a nonlinear filtering problem for Lévy processes with application to finance. *Appl. Math. Optim.* **50**, 119–134. (doi:10.1007/s00245-004-0798-6)

[49] Poklukar, D. R. 2006 Nonlinear filtering for jump-diffusions. *J. Comput. Appl. Math.* **197**, 558–567. (doi:10.1016/j.cam.2005.11.014)

[50] Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., & Mishchenko, E. F. 1962 *The Mathematical Theory of Optimal Processes* (New York: John Wiley & Sons).

[51] Popa, S. & Sritharan, S. S. 2009 Nonlinear filtering of Ito-Lévy stochastic differential equations with continuous observations. *Commun. Stoch. Anal.* **3**, 313–330. (www.math.lsu.edu/cosa/3-3-01[195].pdf)

[52] Protter, P. E. 2005 *Stochastic Integration and Differential Equations*, 2nd ed. (New York: Springer).

[53] Rutkowski, M. 1994 Optimal linear filtering and smoothing for a discrete time stable linear model. *J. Multivariate Anal.* **50**, 68–92. (doi:10.1006/jmva.1994.1035)

[54] Rutkowski, M. & Yu, N. 2007 On the Brody-Hughston-Macrina approach to modelling of defaultable term structure. *Int. J. Theoretical Appl. Finance* **10**, 557–589. (doi:10.1142/S0219024907004263)

[55] Rydberg, T. H. 1997 The normal inverse Gaussian Lévy process: simulation and approximation. *Commun. Statist. Stochastic Models* **13**, 887-910. (doi:10.1080/15326349708807456)

[56] Sato, K. 1999 *Lévy Processes and Infinitely Divisible Distributions* (Cambridge: Cambridge University Press).

[57] Schoutens, W. 2003 *Lévy Processes in Finance: Pricing Financial Derivatives* (New York: Wiley).

[58] Schoutens, W. & Teugels, J. L. 1998 Lévy processes, polynomials and martingales. *Commun. Statist. Stochastic Models* **14**, 335-349. (doi:10.1080/15326349808807475)

[59] Schrödinger, E. 1915 Zur Theorie der Fall- und Steigversuche an Teilchen mit Brownscher Bewegung. *Physikalische Zeitschrift* **16**, 289–295.

[60] Segall, A. & Kailath, T. 1975 The modelling of randomly modulated jump processes. *IEEE Trans. Information Theory* **21**, 135–143. (doi:10.1109/TIT.1975.1055359)

[61] Segall, A., Davis, M. H. A. & Kailath, T. 1975 Nonlinear filtering with counting observations. *IEEE Trans. Information Theory* **21**, 143–149. (doi:10.1109/TIT.1975.1055360)

[62] Shefrin, H. 2008 *A Behavioral Approach to Asset Pricing*, 2nd ed. (Burlington, MA: Academic Press).

[63] Tweedie, M. C. K. 1945 Inverse statistical variates. *Nature* **155**, 453–453. (doi:10.1038/155453a0)

[64] Wasan, M. T. 1968 On an inverse Gaussian process. *Scand. Actuarial J.* **1968**, 69–96. (doi:10.1080/03461238.1968.10413264)

[65] Wiener, N. 1948 *Cybernetics: or the Control and Communication in the Animal and the Machine.* (New York: John Wiley & Sons).

[66] Wiener, N. 1949 *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* (New York: John Wiley & Sons).

[67] Wiener, N. 1954 *The Human Use of Human Beings.* (Boston: Houghton Mifflin Co.).

[68] Wonham, W. M. 1965 Some applications of stochastic differential equations to optimal nonlinear filtering. *J. Soc. Indus. and Appl. Math.* A**2**, 347–369. (doi:10.1137/0302028)

[69] Xiong, J. 2008 *An Introduction to Stochastic Filtering Theory.* (Oxford: Oxford University Press).

[70] Yor, M. 2007 Some remarkable properties of gamma processes. In: *Advances in Mathematical Finance, Festschrift Volume in Honour of Dilip Madan.* R. Elliott, M. Fu, R. Jarrow & Ju-Yi Yen, eds. (Basel: Birkhäuser). (doi:10.1007/978-0-8176-4545-8_3)