Information-Based Commodity Pricing
and the
Theory of Signal Processing with Lévy Information

by

Xun Yang

Department of Mathematics
Imperial College London
London SW7 2AZ, United Kingdom
and
Shell International Limited
Shell Centre, London SE1 7NA, United Kingdom

Submitted to Imperial College London
for the degree of
Doctor of Philosophy

2013
Abstract

In mathematical finance, increasing attention is being paid to (a) the construction of explicit models for the flow of market information, and (b) the use of such models as a basis for asset pricing. One notable approach in this spirit is the information-based asset pricing theory of Brody, Hughston and Macrina (BHM), in which so-called information processes are introduced and ingeniously integrated into the general theory of asset pricing. Building on the BHM theory, this thesis presents a number of new developments in this area.

I begin with a brief review of the BHM framework, leading to a discussion of the simplest asset pricing models. Then the first main topic of the thesis, which is based in part on Brody, Hughston & Yang (2013b), is developed, which concerns asset pricing with continuous cash flows in the presence of noisy information. In particular, an information-based model for the pricing of storable commodities and associated derivatives thereof is introduced. The model employs the concept of market information about future supply and demand as a basis for valuation. Physical ownership of a commodity is regarded as providing the beneficiary with a continuous “convenience dividend”, equivalent to a continuous cash flow. The market filtration is assumed to be generated jointly by: (i) an information process concerning the future convenience-dividend flow; and (ii) a convenience-dividend process that provides information about current and past dividend levels. The price of a commodity is given by the risk-neutral expectation of the cumulative future convenience dividends, suitably discounted, conditional on the information provided by the market filtration. In the situation where the convenience dividend is modelled by an Ornstein-Uhlenbeck process, the prices of options on commodities, both when the underlying is a spot price and when the underlying is a futures price, can be derived in closed form. The dynamical equation of the price process is worked out, leading to an identification of the associated innovations process. The resulting model is sufficiently tractable to allow for simulation studies of the resulting commodity price trajectories.

The second main topic of the thesis, which is based in part on Brody, Hughston & Yang (2013a), concerns a generalisation of concept of information process to the situation where the noise is modelled by a general Lévy process. There are many practical circumstances in which signal or noise, or both, exhibit discontinuities. This part of the thesis develops a rather general theory of signal processing involving Lévy noise, with the view to the introduction of a broad and tractable family of information processes suitable for modelling situations involving discontinuous signals, discontinuous noise, and discontinuous information. In this context, each information process is associated with a certain “noise type”, and an information process of a given noise type is distinguished by the message that it carries. More specifically, each information process
is associated in a precise way to a Lévy process, which I call the fiducial process. The fiducial process is the information process that results in the case of a null message, and can be regarded as a “pure noise” process of the given noise type. Information processes can be classified by the characteristics of the associated fiducial processes. To keep the discussion simple, I mainly consider the case where the message is represented by a single random variable. I construct the optimal filter in the form of a map that takes the \textit{a priori} distribution of the message to an \textit{a posteriori} distribution that depends on the information made available. A number of examples are presented and worked out in detail. The results vary remarkably in detail and character for the different types of information processes considered, and yet there is an overall unity in the scheme that allows for the construction of numerous explicit and interesting examples. The results presented in the second part of the thesis therefore have the potential to pave the way toward a variety of new applications, including applications to problems in finance.
Declaration of Authorship

I, Xun Yang, declare that this thesis, entitled “Information-Based Commodity Pricing and the Theory of Signal Processing with Lévy Information”, and the work presented in it, are my own.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
Acknowledgements

In carrying out the work presented in this thesis I have had the opportunity to be supervised both by Professor D. C. Brody and Professor L. P. Hughston, and I would like to express my gratitude to each of them for their help and support. I would also like to thank my examiners, Professor D. Brigo and Professor G. Peskir, for valuable suggestions. This work was carried out under the terms of a sponsored PhD studentship at Imperial College supported in full by Shell International Limited, London. I would like to thank Paul Downie, Nick Wakefield and Andy Longden at Shell for making this opportunity available to me, which has given me an invaluable experience for which I am truly grateful. During the course of my PhD studies I have had the opportunity to meet many people with whom I have had beneficial discussions, and to make many friends, both among staff and students at Imperial College—too many to name—and among colleagues at Shell—again too many to name—but all of whom I would like to thank. I have also benefited from participation in the activities of the wider London mathematical finance community, and from useful feedback concerning presentations of my work by participants at the AMaMeF (Advanced Mathematical Methods in Finance) conferences in Alesund, Norway (2009), and Bled, Slovenia (2010), the Workshop on Derivatives Pricing and Risk Management at the Fields Institute, Toronto (2010), the Young Researchers in Mathematics Conference at Warwick (2011), and the London Graduate School PhD Day at the London School of Economics (2012). And finally I would especially like to express my thanks to my mother Zhilan Zhang, my father Wanzhang Yang, and my wife Xing Xiong, for their constant support, encouragement, and unreserved love.
Communication is the cement of society. And since sociology and anthropology are primarily sciences of communication, they therefore fall under the general head of cybernetics. That particular branch of sociology with is known as economic ... is a branch of cybernetics.

—Norbert Wiener
# Contents

Abstract

Declarations

Acknowledgements

List of Figures

I Theory of Information-Based Commodity Pricing

1 Introduction

2 Commodity Prices

## I Theory of Information-Based Commodity Pricing

1 Introduction

1.1 Models for commodity pricing

1.2 Information-based approach

1.3 Review of information-based asset pricing framework

1.3.1 Overview

1.3.2 Modelling framework

1.3.3 Modelling the cash flows and the market information

1.3.4 Asset price processes

1.3.5 Asset price dynamics

1.3.6 European call option pricing formula

1.3.7 Applications to commodity prices

2 Commodity Prices

2.1 Chapter overview

2.2 Model setup

2.3 Modelling the convenience dividend

2.4 Properties of the Ornstein-Uhlenbeck process

2.4.1 Reinitialization and orthogonal decomposition

2.4.2 The Ornstein-Uhlenbeck bridge

2.5 Commodity pricing formula

2.6 Special case of constant interest rates

2.7 Alternative derivation of commodity price process

2.8 Price dynamics and innovation representation

2.9 Monte Carlo simulation of crude oil price process
3 Commodity Derivatives 43
  3.1 Pricing options on commodity spot instruments 43
  3.2 Option price analysis 47
  3.3 Futures contracts and derivatives on futures 48
    3.3.1 Futures prices 48
    3.3.2 European options on futures prices 49
  3.4 Time-inhomogeneous extensions 53
    3.4.1 Modelling the convenience dividend 53
  3.5 Properties and applications of time-inhomogeneous OU process 54
    3.5.1 Reinitialisation property and orthogonal decomposition 55
    3.5.2 Time-inhomogeneous OU bridge 56
  3.6 Commodity prices in a time-dependent setting 57
  3.7 Commodity derivatives in a time-dependent setting 58

II Theory of Signal Processing with Lévy Noise 61

4 Introduction 63
  4.1 Motivation 63
  4.2 Synopsis of the theory of Lévy information 65

5 The Theory of Signal Processing with Lévy Information 69
  5.1 Overview of Lévy processes 69
    5.1.1 Infinitely divisible random variables 69
    5.1.2 Lévy processes: definitions and main properties 70
  5.2 Lévy information: definition 72
  5.3 Asymptotic behaviour of Lévy information 77
  5.4 Existence of Lévy information 78
  5.5 Conditional expectations 80
  5.6 General characterisation of Lévy information 83
  5.7 Martingales associated with Lévy information 84
  5.8 On the role of Legendre transforms 86
  5.9 Time-dependent Lévy information 88
    5.9.1 Time-dependent information flow rate 89
    5.9.2 Application to change-point detection problem 90
  5.10 Entropy and mutual information 91
    5.10.1 Entropy and uncertainty 91
    5.10.2 Mutual information 93

6 Examples of Lévy Information Processes 97
  6.1 Chapter overview 97
  6.2 Brownian information process 97
    6.2.1 Definition and properties 97
    6.2.2 Brownian information 98
    6.2.3 Mutual Brownian information 101
  6.3 Poisson information process 103
    6.3.1 Definition and properties 103
    6.3.2 Poisson information 104
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3.3 Mutual Poisson information</td>
<td>106</td>
</tr>
<tr>
<td>6.4 Gamma information process</td>
<td>108</td>
</tr>
<tr>
<td>6.4.1 Definition and properties</td>
<td>108</td>
</tr>
<tr>
<td>6.4.2 Gamma information</td>
<td>109</td>
</tr>
<tr>
<td>6.4.3 Mutual gamma information</td>
<td>111</td>
</tr>
<tr>
<td>6.5 Variance-gamma information process</td>
<td>114</td>
</tr>
<tr>
<td>6.5.1 Definition and properties</td>
<td>114</td>
</tr>
<tr>
<td>6.5.2 Variance-gamma information</td>
<td>117</td>
</tr>
<tr>
<td>6.6 Negative-binomial information process</td>
<td>118</td>
</tr>
<tr>
<td>6.6.1 Definition and properties</td>
<td>118</td>
</tr>
<tr>
<td>6.6.2 Negative-binomial information</td>
<td>119</td>
</tr>
<tr>
<td>6.7 Inverse Gaussian information process</td>
<td>121</td>
</tr>
<tr>
<td>6.7.1 Definition and properties</td>
<td>121</td>
</tr>
<tr>
<td>6.7.2 Inverse Gaussian information</td>
<td>122</td>
</tr>
<tr>
<td>6.8 Normal inverse Gaussian information process</td>
<td>123</td>
</tr>
<tr>
<td>6.8.1 Definition and properties</td>
<td>123</td>
</tr>
<tr>
<td>6.8.2 Normal inverse Gaussian information</td>
<td>124</td>
</tr>
<tr>
<td>6.9 Generalised hyperbolic information process</td>
<td>124</td>
</tr>
<tr>
<td>6.9.1 Definition and properties</td>
<td>124</td>
</tr>
<tr>
<td>6.9.2 Generalised hyperbolic information</td>
<td>127</td>
</tr>
<tr>
<td>6.10 Concluding remarks</td>
<td>128</td>
</tr>
</tbody>
</table>

A Proof of joint Markov property satisfied by generators of commodity information filtration 129

B Derivation of European commodity option pricing formula 133

C Constant parameter OU process vs time-inhomogeneous OU process 135

D Conditional variance of convenience dividend flow 139

Bibliography 143
List of Figures

2.1 Two sample paths of OU Bridge with the following parameters: $X_0 = 0.5$, $\theta = 1.2$, $\psi = 0.4$, $\kappa = 0.2$, $T = 1$. The number of steps is 365. The mean (red) and variance (blue) of the OU bridge are also plotted. .................. 21

2.2 Nine sets of ten sample paths of the commodity price process, with variable parameters $\psi$, $\kappa$ and the following parameters: $S_0 = 75$, $S_\infty = 60$, $\sigma = 0$, $r = 0.05$, $T = 1$. The number of steps is 365. This simulation shows the daily price movement for a year, when $\sigma = 0$. ................. 40

2.3 Nine sets of ten sample paths of the commodity price process, with variable parameters $\sigma$ (rate at which information is revealed to market participants) ranging from 0.01 to 1 (value increases from top to bottom and left to right), and the following parameters: $S_0 = 75$, $S_\infty = 60$, $\psi = 0.3$, $\kappa = 0.03$, $r = 0.05$, $T = 1$. The number of steps is 365. This simulation shows the daily price movement for a year, for a range of values for $\sigma$. . . . 40

2.4 Brent crude daily spot price from 4 November 2008 to 20 April 2010 (black path) plotted against five simulated paths from the model, with the following parameters: $S_0 = 62.78$, $S_\infty = 60$, $\psi = 0.4$, $\kappa = 0.05$, $r = 0.025$, $T = 1$. The number of steps is 365. ......................... 41

3.1 Commodity call option price surface as a function of the initial asset price in the OU model and the time to maturity of the option. The parameters are set as follows: $\kappa = 0.15$, $\theta$ ranges from 0.3 to 0.8 with increments of 0.01, $\sigma = 0.25$, $X_0 = 0.6$, $\psi = 0.15$, $r = 0.05$, and $K = 10$. The range of the maturities is from $T = 0$ to $T = 3.0$. ................................. 47

3.2 The commodity call option delta as a function of the initial asset price in the OU model. The parameters are set as follows: $\kappa = 0.15$, $\theta$ ranges from 0.3 to 0.8 with increments of 0.01, $\sigma = 0.25$, $X_0 = 0.6$, $\psi = 0.15$, $r = 0.05$, and $K = 10$. The three maturities are $T = 0.5$ (blue), $T = 1.0$ (green), and $T = 3.0$ (red). ................................. 48
6.1 Mutual information in the case of Brownian noise. The mutual information $I$ measures the amount of information contained in the observation about the value of the unknown signal $X$. At time zero, no data is available so that the accumulated information content is zero. However, as time progresses, data is forthcoming that enhances the knowledge of the observer. Eventually, sufficient information, equivalent to the amount of the initial uncertainty $-\sum_i p_i \ln p_i$, is gathered, at which point the value of $X$ is revealed. Strictly speaking this happens asymptotically as $t \to \infty$, although for all practical purposes the value of $X$ will be revealed with high confidence level after a passage of finite amount of time. In this example, the parameters are chosen to be $x_1 = 1$, $x_2 = 2$, $p_1 = 0.4$, $p_2 = 0.6$, and $T = 50$. The initial entropy (asymptotic value of $I$) in this example is approximately 0.67.

6.2 Mutual information in the case of Poisson noise. The mutual information $I$ measures the amount of information contained in the observation about the value of the unknown signal $X$. At time zero, no data is available so that the accumulated information content is zero. As time progresses, data is forthcoming that enhances the knowledge of the observer. Eventually, a sufficient amount of information, equivalent to the amount of the initial uncertainty $-\sum_i p_i \ln p_i$, is gathered, at which point the value of $X$ is revealed. Strictly speaking this happens asymptotically as $t \to \infty$, although for all practical purposes the value of $X$ will be revealed with high confidence level after a passage of finite amount of time. In this example, the parameters are chosen to be $x_1 = 1$, $x_2 = 2$, $p_1 = 0.4$, $p_2 = 0.6$, and $T = 50$. The three plots correspond to $m = 1$, $m = 1.3$, and $m = 1.5$. The initial entropy (asymptotic value of $I$) in this example is approximately 0.67.

6.3 Mutual information in the case of the gamma noise. The mutual information $I$ measures the amount of information contained in the observation about the value of the unknown signal $X$. At time zero, no data is available so that the accumulated information content is zero. However, as time progresses, data is forthcoming that enhances the knowledge of the observer. Eventually, a sufficient amount of information, equivalent to the amount of the initial uncertainty $-\sum_i p_i \ln p_i$, is gathered, at which point the value of $X$ is revealed. Strictly speaking this happens asymptotically as $t \to \infty$, although for all practical purposes the value of $X$ will be revealed with high confidence level after a passage of finite amount of time. In this example, the parameters are chosen to be $x_1 = 1$, $x_2 = 2$, $p_1 = 0.4$, $p_2 = 0.6$, and $T = 50$. The initial entropy (asymptotic value of $I$) in this example is approximately 0.67.

6.4 Mutual information comparison. The time dependence of the mutual information $I$ is compared for three types of information processes: Brownian, Poisson, and gamma. The common parameters are chosen to be $x_1 = 1$, $x_2 = 2$, $p_1 = 0.4$, $p_2 = 0.6$, and $T = 50$. Since other parameters embody different meanings, a direct comparison as shown here need not reveal quantitative information. Nevertheless, the intuition one gains from the figure is that the revelation of the modulation signal in the Poisson information is somewhat slower than that in the drift for the Brownian information and the scale for the gamma information.
Part I

Theory of Information-Based Commodity Pricing
Chapter 1

Introduction

1.1 Models for commodity pricing

The starting point of most derivative pricing models—including those for commodity derivatives—is the specification of the price process of the underlying asset—for example, a geometric Brownian motion. More generally, the outcome of chance in the economy is often modelled by a probability space equipped with the filtration generated by a multi-dimensional Brownian motion, and it is assumed that asset prices are Ito processes that are adapted to this filtration. This particular example is, of course, the “standard” model within which a great deal of modern financial engineering has been carried out.

A basic methodological problem with the standard model (and the same applies to various generalisations thereof involving jump processes) is that the market filtration is fixed once and for all. In other words, the filtration, which represents the unfolding of information available to market participants, is modelled first, in an essentially \textit{ad hoc} manner, and then it is assumed that the various asset price processes are adapted to it. But no indication is given about the nature of the “information” implicit in this setup, and it is not obvious at the beginning why Brownian motion, for example, should be regarded as providing information, rather than mere noise, in models constructed in this way.

In the real world—and this is certainly true in the world of commodity trading—the information available to market participants concerning the likely future cash flows associated with an asset is essential to market participants when they make their decisions to buy or sell that asset. A change in such information or the arrival of such information will typically have an effect on the price at which market participants are willing to buy, or to sell, even if the agent’s preferences remain fundamentally unchanged.
Suppose, for example, that a trader working for a firm that holds crude oil in its inventory is thinking of selling a quantity of it at a price that seems attractive. Then along comes a news article drawing attention to the possibility of a shortage of crude oil three months hence. After some reflection, the trader decides it is no longer attractive to sell at that price; the alternative of holding on to the inventory for another three months, even taking into account the cost of carrying the inventory forward, is better. As a result, the trader declines the transaction. This trader is not alone in reaching such a decision, and the price for immediate delivery will increase whereas the price for delivery in three months will drop since the shortages have been eased as a consequence of stock piling.

We take the view that the movement of the price of an asset should, therefore, be regarded as an *emergent phenomenon*, stimulated by information flows. That is, the price process of an asset should be viewed as the output of the various decisions made relating to possible transactions in the asset: these decisions in turn should be understood as being induced primarily by the flow of information to market participants.

### 1.2 Information-based approach


In the theory of commodity pricing presented in Part I of the thesis, which is based on the approach set out in Brody, Hughston & Yang (2013b), we use the concept of market information about future supply and demand as a valuation basis for commodities and commodity derivatives. We shall be concerned here primarily with those commodities that can be physically stored, so we can assume the existence of the so-called convenience yield. Brennan (1991) defines the net convenience yield as “the flows of services accruing to the holder of the physical commodity, but not to the owner of a futures contract”. Indeed, physical inventory provides certain basic elements of service such as offering the owner the possibility (a) to avoid shortage of the spot commodity and thus to maintain
any production dependent upon it, or (b) to benefit from an anticipated future price increase. Generally speaking, the convenience yield represents, in percentage terms, the residual “benefit” conferred upon or derived by the owner of the commodity by virtue of its possession, after storage costs (explicit or implicit), insurance, and the like, have been duly taken into account.

The notion of a convenience yield, treated as a net “dividend” yield paid in effect to the owner of the physical commodity, drives the relation between the spot prices and the futures prices for many commodities (Brennan 1986, Fama & French 1987, 1988). It is reasonable to assume that an inverse relation holds between the overall inventory levels of a commodity and the associated convenience yield. Given the fact that inventories fluctuate, this indicates that the assumption of a constant convenience is in general likely to be an invalid one, and hence that a stochastic approach to convenience yield is needed in the modelling of oil prices (Brennan 1986, Gibson & Schwartz 1990).

In our approach to commodity spot prices we shall assume that the possession of one unit of commodity provides a “convenience dividend” equivalent to a continuous cash flow given by a random process \( \{X_t\}_{t \geq 0} \). Note that we are starting from the spot price because physical ownership of the commodity is required in order to accumulate the “convenience dividend”. Note also that we shall be working directly with the actual flow of convenience from the storage or “possession” of the commodity, rather than the convenience yield. The point is that the convenience yield is in some respects a secondary notion since it depends on the price, which is what we are trying to determine. This is why we refer to \( \{X_t\} \) as a convenience dividend. Generally speaking, we take the view that in some ways it is more natural and more fundamental to model the convenience dividend than the convenience yield.

When a storable commodity is consumed, one can think of it as being exchanged for a consumption good of identical value—think of the difference between a bottle of wine (of known type and quality) which has not yet been opened, and an opened bottle of the same wine—the former is a storable good and the latter, once opened, becomes a “consumption” good. Following this line of argument, Jarrow (2010) takes the view that the value of a commodity derives in part from the element of optionality implicit in the timing of the conversion of the commodity from (in our language) a storable good to a consumption good. From this point of view, one can argue that the convenience yield in effect monetizes the value of the “option to consume” in the form of a cash-flow stream, in some respects similar to the way that in credit theory a randomly timed credit event (the option of the bond issuer to default) is effectively monetized in the form of a (negative) continuous cash flow over the life of the credit risky product—thus leading to valuation formulae where the discounting is carried out not with respect to
the short rate, but rather with respect to the short rate augmented by the hazard rate for the credit event. In the case of commodities, the “consumption” event is a “good” outcome rather than a “bad” one, so when it is monetized over the life of the storable commodity the effect is to diminish the effective short rate used for discounting, rather than to enhance it, and that can be understood as the origin of the so-called convenience yield as a tool for modelling the dynamics of commodities.

In what follows, the convenience dividend (the existence of which we assume) will be modelled explicitly as a stochastic process, as we have indicated. In addition, to address the importance of forward looking information, we introduce a so-called market information process concerning future supply and demand, inventory, and overall market sentiment. The market filtration is assumed to be generated jointly by the convenience dividend process and the market information process. One can interpret the knowledge of the convenience dividend as providing information about current and past dividend levels, whereas the information process gives partial or speculative information about the future dividend flow.

The spot price of the commodity is given in the information-based approach by the discounted expected value of the cumulative future convenience dividends, conditional on the information provided by the market filtration. In the information-based framework, the filtration is modelled in a non-trivial way, so as a consequence the resulting dynamics for the prices is of a novel character. Nevertheless, the resulting scheme has a good deal of tractability, and in the case where convenience dividend is modelled by an Ornstein-Uhlenbeck process, we are able to derive closed-form expressions both for the underlying spot price and for the associated European option pricing formula.

1.3 Review of information-based asset pricing framework

1.3.1 Overview

Certain aspects of the BHM framework were originally developed a number of years ago as part of an attempt to gain a better understanding of the evolution and dynamics of quantum states in the presence of quantum noise (Brody & Hughston 2002; see also Adler, Brody, Hughston & Brun 2001, Brody & Hughston 2005, 2006). The ideas arising in these early investigations were redeveloped and extended by use of a novel Brownian-bridge technique in such a way as to provide a new approach to asset pricing. The first applications of the new approach were to the pricing and hedging of credit risky assets (Macrina 2006, Brody, Hughston & Macrina 2007). Since then the framework has evolved much further, and has found applications to equity pricing and stochastic

The BHM framework can in some respects be seen as part of a larger program with an extensive literature being pursued by a number of authors exploring the role of incomplete information in finance, and attempting under various assumptions to model how prices are determined in such a setting, and how investment strategies should be best formulated. We mention, for example, the work of Gennotte (1986), Detemple (1986, 1991), Veronese (2000), Duffie & Lando (2001), Giesecke & Goldberg (2004), Jarrow & Protter (2004), Gombani, Jaschke & Runggaldier (2005), Coculescu, Geman & Jeanblanc (2008), Frey & Schmidt (2009), Björk, Davis & Landén (2010), and Duffie (2012).

As we have remarked, the BHM framework regards the movements of the price of an asset as an emergent phenomenon. The price process of an asset is to be thought of as the output of the decisions made relating to possible transactions in the asset, and these decisions in turn is understood as being induced primarily by the flow of information to market participants. In other words, the BHM framework for asset pricing is based on modelling of the flow of market information. The information is that concerning the values of the future cash flows associated with the given assets, based upon which market participants determine estimates for the value of the right to the impending cash flows. These estimates in turn lead to the decisions concerning transactions that eventually trigger movements in the market quoted price.

Perhaps the most important difference between general pricing methods and the BHM framework is the following: in the BHM framework the stochastic process that governs the dynamics of an asset is deduced rather than imposed. In particular, rather than being imposed at the beginning of the modelling process in an arbitrary way, the dynamics of the asset are deduced as a consequence of the modelling of (a) the actual cash flows delivered by the asset, and (b) the associated flows of information relating to these random payments. The BHM framework is fully consistent with the general principles of arbitrage-free pricing theory. We present a brief synopsis of the BHM framework below, following the treatment of Brody, Hughston & Macrina (2007, 2008a).
1.3.2 Modelling framework

The BHM asset pricing framework requires three ingredients: (i) the cash flows; (ii) the investor risk preferences; and (iii) the flow of information available to market participants. To incorporate these ingredients into the modelling process, one needs to translate them into mathematical terms, namely: (a) cash flows are modelled as random variables; (b) investors risk preference are modelled with the determination of a pricing kernel; and (c) the market information flow is modelled with the specification of a market filtration.

The modelling process proceeds by the specification of a probability space \((\Omega, \mathcal{F}, Q)\), on which a filtration \(\{\mathcal{F}_t\}_{0 \leq t < \infty}\) is constructed such that \(\{\mathcal{F}_t\}\) can be identified as the market filtration. All asset price processes and other information-providing processes accessible to market participants will be adapted to \(\{\mathcal{F}_t\}\). The probability measure \(Q\) denotes the risk-neutral measure. The framework assumes the absence of arbitrage and the existence of a pricing kernel. With these conditions the existence of a unique preferred risk-neutral measure is ensured.

For simplicity, it is assumed that the default-free system of interest rates is deterministic (this condition is relaxed in Rutkowski & Yu 2007). The absence of arbitrage then implies that the default free discount bond system, denoted by \(\{P_{tT}\}_{0 \leq t \leq T < \infty}\), can be written in the form \(P_{tT} = P_{0T}/P_{0t}\). The discount function \(\{P_{0t}\}_{0 \leq t < \infty}\) is assumed to be differentiable and strictly decreasing, and to satisfy \(0 < P_{0t} < 1\) and \(\lim_{t \to \infty} P_{0t} = 0\).

1.3.3 Modelling the cash flows and the market information

We consider a single isolated cash flow occurring at time \(T\), represented by a random variable \(X_T\). The value \(S_t\) of the cash flow at any earlier time \(t\) in the interval \(0 \leq t \leq T\) is given by the discounted conditional expectation of \(X_T\):

\[
S_t = P_{tT}E^Q[X_T|\mathcal{F}_t].
\]  

(1.1)

The value of the cash flow will be revealed at time \(T\). It is reasonable to assume that some partial information regarding the value of the cash flow \(X_T\) is available at earlier time, and this information will in general be imperfect. We therefore wish the market filtration to fulfil the properties that \(\mathcal{F}_t\) for \(t < T\) embodies partial information concerning the value \(X_T\) of the impending cash flow, and that \(X_T\) is \(\mathcal{F}_T\)-measurable.

With these objective in mind, we proceed as follows. The flow of information available to market participants about the cash flow is assumed to be contained in a process
\{\xi_t\}_{0 \leq t \leq T} defined by

\[ \xi_t = \sigma t X_T + \beta_{tT}. \]  

(1.2)

The process \( \{\xi_t\} \) is referred to as the market information process, or simply the information process. Observe that information process is constructed by two components. The first term \( \sigma t X_T \) contains the “true information” about the value of the cash flow \( X_T \), and is assumed here for simplicity to grow linearly in time at a rate \( \sigma \). The parameter \( \sigma \) can therefore be thought of as representing the signal-to-noise ratio; that is, the rate at which information about the true value of \( X_T \) is revealed as time passes. If \( \sigma \) is low, then the value of \( X_T \) is effectively hidden until very near the time \( T \) of the occurrence of the cash flow. If \( \sigma \) is high, then the value of the cash flow is for all practical purposes revealed before \( T \). The second term is given by the process \( \{\beta_{tT}\}_{0 \leq t \leq T} \), which is assumed to be a standard Brownian bridge over the time interval \([0, T]\). It is well-known that Brownian bridge process is a Gaussian process with mean zero, and the covariance of \( \beta_{sT} \) and \( \beta_{tT} \) is given by \( s(T - t)/T \) for \( s \leq t \). Also we have \( \beta_{0T} = 0 \) and \( \beta_{TT} = 0 \).

In the information-based framework it is assumed that the cash flow \( X_T \) and the process \( \{\beta_{tT}\} \) are independent. Therefore, the information contained in the bridge process represents “pure noise”. The market filtration \( \{\mathcal{F}_t\} \) is thus assumed to be generated by the market information process:

\[ \mathcal{F}_t = \sigma (\{\xi_s\}_{0 \leq s \leq t}). \]  

(1.3)

An immediate consequence of this setup is that \( X_T \) is \( \mathcal{F}_T \)-measurable, but not \( \mathcal{F}_t \)-measurable for \( t < T \). Thus, as required, the value of \( X_T \) becomes “know” at time \( T \), but not earlier. The bridge process cannot be accessed directly by the market participants because it is not \( \mathcal{F}_t \)-measurable for \( t < T \). This is consistent with the fact that until the cash is paid at time \( T \), the market participants cannot fully extract the “true information” about \( X_T \) from the noisy information in the market.

### 1.3.4 Asset price processes

Having constructed the market information structure in the case of a single cash flow, we proceed, following BHM, to calculate the associated price process introduced in (1.1). It will be assumed that the a priori probability distribution of the cash flow \( X_T \) is known. Let us further assume that \( X_T \) has a density function \( p(x) \). Note, for convenience of demonstration and continuity with the materials to come later in this thesis, that
although we consider here the example of a continuous random variable $X_T$, the result are equally applicable to random variables with more general distributions.

The determination of the asset price simplifies if we make note of the fact that the information process $\{\xi_t\}$ is Markov. This can be seen by noting that

$$Q(\xi_t \leq x|\xi_s, \xi_{s_1}, \xi_{s_2}, \ldots, \xi_{s_k})$$

$$= Q\left(\xi_t \leq x|\xi_s, \frac{\xi_{s_1}}{s_1}, \frac{\xi_{s_2}}{s_2}, \ldots, \frac{\xi_{s_{k-1}}}{s_{k-1}}, \frac{\xi_{s_k}}{s_k}\right)$$

$$= Q\left(\xi_t \leq x|\xi_s, \frac{\beta_{sT}}{s} - \frac{\beta_{s_1}}{s_1}, \frac{\beta_{s_2}}{s_2}, \ldots, \frac{\beta_{s_{k-1}}}{s_{k-1}} - \frac{\beta_{s_k}}{s_k}\right)$$

for $s_1 \leq s_2 \leq \cdots \leq s_k \leq s \leq t$. Now it is a property of a Brownian bridge that the random variables $\beta_{sT}/s - \beta_{s_1}/s_1$ and $\beta_{s_1}/s_1 - \beta_{s_2}/s_2$ are independent, and more generally that $\beta_{sT}/s - \beta_{s_1}/s_1$ and $\beta_{s_1}/s_1 - \beta_{s_2}/s_2$ are independent. It follows that

$$Q(\xi_t \leq x|\xi_s, \xi_{s_1}, \xi_{s_2}, \ldots, \xi_{s_k}) = Q(\xi_t \leq x|\xi_s) \quad (1.4)$$

for arbitrary $s_1 \leq s_2 \leq \cdots \leq s_k \leq s \leq t$. In addition, from the fact that $X_T$ is $\mathcal{F}_T$ measurable, we deduce that for $t \leq T$, we have

$$S_t = P_T E^Q[X_T|\xi_t] = P_T \int_0^\infty x \pi_t(x) dx, \quad (1.5)$$

where $\pi_t(x)$ is the conditional probability density function for the random variable $X_T$:

$$\pi_t(x) = \frac{d}{dx} Q(X_T \leq x|\xi_t). \quad (1.6)$$

Note that by using a form of the Bayes formula, we can work out the conditional probability density process for the cash flow:

$$\pi_t(x) = \frac{p(x) \rho(\xi_t|X_T = x)}{\int_0^\infty p(x) \rho(\xi_t|X_T = x) dx}. \quad (1.7)$$

Since $\beta_{sT}$ is a Gaussian random variable with mean zero and variance $t(T-t)/T$, and since that conditional on $X_T = x$ we have

$$\rho(\xi_t|X_T = x) = Q(\beta_{sT} < \xi_t - \sigma X_T t|X_T = x), \quad (1.8)$$

we deduce that the conditional probability density for $\xi_t$ is:

$$\rho(\xi_t|X_T = x) = \sqrt{\frac{T}{2\pi t(T-t)}} \exp \left(-\frac{\left(\xi_t - \sigma x\right)^2}{2t(T-t)}\right). \quad (1.9)$$
Chapter 1. *Introduction*

It follows by inserting this into the Bayes formula (1.7) that

\[
\pi_t(x) = \frac{p(x) \exp \left[ \frac{T}{T-t} (\sigma x \xi_t - \frac{1}{2} \sigma^2 x^2 t) \right]}{\int_0^\infty p(x) \exp \left[ \frac{T}{T-t} (\sigma x \xi_t - \frac{1}{2} \sigma^2 x^2 t) \right] dx}.
\] (1.10)

Hence, the price process \( \{S_t\}_{0 \leq t \leq T} \) can be expressed in the form:

\[
S_t = P_{tT} \frac{\int_0^\infty x p(x) \exp \left[ \frac{T}{T-t} (\sigma x \xi_t - \frac{1}{2} \sigma^2 x^2 t) \right] dx}{\int_0^\infty p(x) \exp \left[ \frac{T}{T-t} (\sigma x \xi_t - \frac{1}{2} \sigma^2 x^2 t) \right] dx}.
\] (1.11)

### 1.3.5 Asset price dynamics

In standard approaches to asset pricing, the starting point is the specification of the price process in the form of a stochastic differential equation. In the BHM approach, the price process is directly deduced from the specification of the cash flow as well as the market filtration. Having obtained the price process (1.11), however, it will be of interest to identify to which stochastic differential equation (1.11) is the solution. To investigate this, let us define

\[
X_{tT} = E^Q [X_T | \xi_t].
\] (1.12)

From the foregoing calculation it should be evident that we can express \( X_{tT} \) in the form \( X_t = X(\xi_t, t) \), where \( X(\xi_t, t) \) is defined by

\[
X(\xi_t, t) = \frac{\int_0^\infty x p(x) \exp \left[ \frac{T}{T-t} (\sigma x \xi_t - \frac{1}{2} \sigma^2 x^2 t) \right] dx}{\int_0^\infty p(x) \exp \left[ \frac{T}{T-t} (\sigma x \xi_t - \frac{1}{2} \sigma^2 x^2 t) \right] dx}.
\] (1.13)

An application of Ito’s lemma shows that the dynamical equation for \( X_{tT} \) is given by

\[
dX_{tT} = \frac{\sigma T}{T-t} V_t \left[ \frac{1}{T-t} (\xi_t - \sigma T X_{tT}) dt + d\xi_t \right].
\] (1.14)

Here \( V_t \) is the conditional variance of the cash flow \( X_T \):

\[
V_t = E_t \left[ (X_T - E_t [X_T])^2 \right] = \int_0^\infty x^2 \pi_t(x) dx - \left( \int_0^\infty x \pi_t(x) dx \right)^2,
\] (1.15)

where \( E_t[-] \) denotes the conditional expectation \( E^Q [-| \xi_t] \). We proceed further by defining the new process \( \{W_t\}_{0 \leq t \leq T} \) according to

\[
W_t = \xi_t - \int_0^t \frac{1}{T-s} (\sigma T X_{sT} - \xi_s) ds.
\] (1.16)
and substitute this back into equation (1.14). Then we have

$$dX_{tT} = \frac{\sigma T}{T-t} V_t dW_t. \quad (1.17)$$

It follows from the expression for the price process and the relation that

$$dP_{tT} = r_t P_{tT} dt, \quad (1.18)$$

where $r_t$ denotes the short rate, the dynamics of the price process is

$$dS_t = r_t S_t dt + \Sigma_{tT} dW_t, \quad (1.19)$$

where

$$\Sigma_{tT} = P_{tT} \frac{\sigma T}{T-t} V_t \quad (1.20)$$

is the absolute volatility process.

It is a straightforward exercise to check the process $\{W_t\}$ defined in equation (1.16) is a $(\mathbb{Q}, \mathcal{F}_t)$-Brownian motion. One can check this by use of the Lévy characterisation of the Brownian motion, that is, to check that $\{W_t\}$ is a $(\mathbb{Q}, \mathcal{F}_t)$-martingale and that the quadratic variation of $\{W_t\}$ is $t$.

It is interesting to observe that in the BHM framework we have been able to derive the Brownian motion that drives the price process. In particular, it is a nonlinear functional of the history of the information process, which in turn depends on the future cash flow of the asset, as well as on ambient noise. This observation shows that the “standard” interpretation, that random movements of prices are generated by noise, is perhaps not satisfactory. Indeed, in Macrina (2006) and Brody et al. (2008a) it is shown that the geometric Brownian motion model used in the Black-Scholes theory can be derived from a market information process consisting of a normal random variable for the log-return of the asset determining the “signal” component, and an independent Brownian bridge for market noise.

### 1.3.6 European call option pricing formula

The information-based framework can be used to price various financial derivatives, in spite of the apparent sophisticated form (1.11) of the price process. For the purpose of illustration, we consider the valuation of a European call option written on a asset for which the dynamics of the price process is given by equation (1.19). The option has strike price $K$ and matures at time $t$; the underlying asset pays a single cash flow at
time $T > t$. Given this setup, we can write down the initial value of this option as

$$C_0 = P_{0t}E^Q \left[(S_t - K)^+\right].$$

(1.21)

Substitute the expression for $S_t$ into the above formula we obtain

$$C_0 = P_{0t}E^Q \left[P_T \int_0^\infty x\pi_t(x) dx - K\right]^+. $$

(1.22)

For simplicity, let us denote the conditional probability $\pi_t(x)$ of (1.10) in the form

$$\pi_t(x) = \frac{p_t(x)}{\int_0^\infty p_t(x) dx},$$

(1.23)

where the unnormalised density process $p_t(x)$ is defined by

$$p_t(x) = p(x)\exp \left[\frac{T}{T-t} \left(\sigma x \xi_t - \frac{1}{2} \sigma^2 x^2 t\right)\right].$$

(1.24)

Given these ingredient, we can rewrite the expression for $C_0$ as

$$C_0 = P_{0t}E^Q \left[\frac{1}{\Psi_t} \left(\int_0^\infty p_t(x) (xP_T - K) dx\right)^+\right],$$

(1.25)

where

$$\Psi_t = \int_0^\infty p_t(x) dx.$$  

(1.26)

It can be shown that the factor $1/\Psi_t$ appearing in (1.25) is a $(Q, F_t)$-martingale, and can be used as a change of measure density martingale. This can be seen as follows.

First, from (1.24) we have

$$dp_t(x) = \sigma T \frac{T}{T-t} x p_t(x) \left(d\xi_t + \frac{1}{T-t} dt\right),$$

(1.27)

and hence

$$d\Psi_t \Psi_t^{-1} = \sigma T \frac{T}{T-t} X_{tT} \left(d\xi_t + \frac{1}{T-t} dt\right),$$

(1.28)

It follows, on account of the Ito lemma, that

$$d\Psi_t^{-1} \Psi_t = -\sigma T \frac{T}{T-t} X_{tT} dW_t,$$

(1.29)

where we have substituted (1.16). Since $\{W_t\}$ is a $Q$-Brownian motion, the desired conclusion follows.
By using \( \{ \Psi^{-1}_t \} \) a new measure \( \mathbb{B} \) on \((\Omega, \mathcal{F}_t)\), which will be called a “bridge measure”. The option price can then be written in the new measure as

\[
C_0 = P_0 \mathbb{E}^\mathbb{B} \left[ \left( \int_0^\infty p_t(x) (xP_t - K) \, dx \right)^+ \right].
\]

(1.30)

What is special about the bridge measure \( \mathbb{B} \) is that under this measure, the process \( \{ \xi_t \} \) is Gaussian with mean zero and variance \( t(T - t)/T \). That is to say, under \( \mathbb{B} \), the information process \( \{ \xi_t \} \) is a standard Brownian bridge. Furthermore, since \( p_t(x) \) can be written as a function of \( \xi_t \), which is \( \mathbb{B} \)-Gaussian, we are able to carry out the expectation and obtain a tractable formula for value of the option. In order to determine the option price, define for each \( t, T \), and \( K \) a constant \( \xi^* \), which is a unique critical value of \( \xi_t \) such that \( S_t = K \) when \( \xi_t = \xi^* \). This implies the following condition:

\[
\int_0^\infty p(x) \exp \left[ \frac{T}{T - t} \left( \sigma x \xi^* - \frac{1}{2} \sigma^2 x^2 t \right) \right] (xP_t - K) \, dx = 0.
\]

(1.31)

Together with the fact that \( \xi_t \) is \( \mathbb{B} \)-Gaussian, we can perform the Gaussian integration explicitly and find the option price:

\[
C_0 = P_0 \int_0^\infty x p(x) N \left( -z^* + \sigma x \sqrt{\tau} \right) \, dx - P_0 K \int_0^\infty p(x) N \left( -z^* + \sigma x \sqrt{\tau} \right) \, dx.
\]

(1.32)

Here \( N(z) \) denotes the cumulative distribution function of a standard normal random variable, and

\[
\tau = \frac{tT}{T - t}, \quad z^* = \xi^* \sqrt{\frac{T}{t(T - t)}}.
\]

(1.33)

We see, therefore, that in spite of the elaborate model (1.11) for the price process, the option pricing formula reduced to something that is analogous to that of Black and Scholes, albeit requiring the relatively simple numerical determination of the solution \( \xi^* \) of (1.31) and performing the integration (1.32).

### 1.3.7 Applications to commodity prices

In what follows, the goal is to apply the BHM method to commodities. Some rethinking of the BHM technique is required, since the convenience dividend of a commodity supplies, in effect, constitutes a continuous cash flow, as opposed to a single cash flow occurring at a prefixed future time in the example considered above. Nevertheless, under appropriate assumptions we are able to obtain exact solutions.
Chapter 1. Introduction

The structure of the remainder of the material in this part of the thesis is as follows. In section 2.2-2.3 we introduce the model for the convenience dividend and for the market filtration. Some useful facts about the mean-reverting Ornstein-Uhlenbeck (OU) process are recalled in section 2.4, in particular, various features of the OU bridge are outlined in section 2.4.2. These are used in the derivation of the commodity price process. In section 2.8, we derive the stochastic differential equation satisfied by the price process. In doing so, we are able to obtain an innovations representation for the associated filtering problem in closed form. Simulation studies of the price process are presented for various values of the model parameters. We work out pricing formulae for call options on the underlying spot price in proposition 3.1.1. In section 3.3, the model is applied to obtain the corresponding price processes for futures contracts, which are useful since futures contracts are very commonly traded in commodity markets. The formulae developed thus far have made use of a simplifying assumption to the effect that the parameters of the model are constant in time. We know, however, from various other models for derivative pricing, that it can be highly advantageous to introduce “time dependent” parameters into the model, to open up the possibility of calibration of the model to various market instruments. Therefore, with this point of view in mind, in section 3.4 we proceed to extend the model to a time-inhomogeneous setup. The resulting formulae in the time-dependent situation are of course of rather greater complexity, but the overall modeling scheme remains analytically tractable, as will be demonstrated.
Chapter 2

Commodity Prices

2.1 Chapter overview

In this chapter, an information-based continuous-time model for the prices of storable commodities is introduced. The model employs the concept of market information about future supply and demand as a basis for valuation. The physical ownership of commodities is regarded as providing “convenience dividends” equivalent to cash flows. The market filtration is assumed to be that generated by the aggregate of (i) an information process concerning the future convenience dividend flow; and (ii) a convenience dividend process that provides information about current and past dividend levels. The price of a commodity is given by the risk-neutral expectation of the discounted cumulative future convenience dividends, conditional on the information provided by the market filtration.

2.2 Model setup

As indicated in the introduction, the starting point of the BHM framework is the specification of: (a) the random variables (called “market factors”) determining the cash flows associated with a given asset, and (b) the flow of information to market participants concerning these market factors. The models that have been considered so far under this framework have the property that the cash flows occur at pre-specified times. The specification of the market factors and the associated information processes for such discrete cash flows is useful for the modelling of many different types of financial contracts. The purpose of the present approach to storable commodity pricing is to introduce an extension of the framework to the situation where an asset pays a continuous dividend.
Chapter 2. Commodity Prices

As an illustrative example we consider, in particular, the case for which the cash flow is an Ornstein-Uhlenbeck process \( \{X_t\} \). As usual we begin with the probability space \((\Omega, \mathcal{F}, \mathbb{P})\), where \(\mathbb{P}\) denotes the market measure. We assume the absence of arbitrage, and that a pricing kernel \(\{\pi_t\}\) has been established. These assumptions ensure the existence of a unique preferred pricing measure (the risk neutral measure), which will be denoted by \(\mathbb{Q}\).

Based on these assumptions, the value at time \(t\) of the storable commodity that generates a continuous stream of benefit equivalent to the cash flow \(\{X_t\}\) is given by the pricing formula

\[
S_t = \frac{1}{\pi_t} \mathbb{E}^\mathbb{P}\left[ \int_t^\infty \pi_u X_u du \right].
\]

(2.1)

Equivalently, transforming to the risk-neutral measure, we can write

\[
S_t = \frac{1}{P_t} \mathbb{E}^\mathbb{Q}\left[ \int_t^\infty P_u X_u du \right],
\]

(2.2)

where

\[
P_t = \exp \left( -\int_0^t r_s ds \right)
\]

is the discount factor, with the associated short rate \(\{r_t\}\). In what follows, we shall use \(\mathbb{E}^\mathbb{Q}[\cdot]\) or \(\mathbb{E}[\cdot]\) to represent the expectation under the risk-neutral measure \(\mathbb{Q}\). For simplicity of exposition, let us assume that the interest rate system is deterministic. Once we work things out for deterministic \(\{r_t\}\) then we can consider the more general situation. From this assumption, we have \(P_t = P_0 t\), where \(\{P_0 t\}_{t \geq 0}\) are initial prices of discount bonds. We shall further assume that the market filtration is generated jointly by the following processes:

(a) the convenience dividend process \(\{X_t\}_{t \geq 0}\); and

(b) an “information process” \(\{\xi_t\}_{t \geq 0}\) of the form:

\[
\xi_t = \sigma t \int_t^\infty P_u X_u du + B_t,
\]

(2.4)

where \(\mathbb{Q}\)-Brownian motion \(\{B_t\}\) is independent of \(\{X_t\}\).

In other words, at time \(t\) the market filtration \(\mathcal{F}_t\) is generated by:

\[
\mathcal{F}_t = \sigma (\{X_s\}_{0 \leq s \leq t}, \{\xi_s\}_{0 \leq s \leq t}).
\]

(2.5)

We can interpret the choice of market filtration as arising from the following two components:
Chapter 2. Commodity Prices

- The knowledge of the convenience dividend \{X_t\}, providing information about the \textbf{current} and \textbf{past} dividend levels.
- The information process \{\xi_t\}, providing partial information about the \textbf{future} dividend flow.

2.3 Modelling the convenience dividend

Motivated in part by the pioneering work of Gibson & Schwartz (1990, 1991), in which the convenience yield is assumed to follow a mean-reverting process, we shall introduce a simple model for the commodity convenience dividend. Specifically, we consider the case for which \{X_t\} is an Ornstein-Uhlenbeck (OU) process.

We shall begin by considering the constant parameter case so that we have the following mean-reverting dynamics for the convenience dividend in the risk neutral measure:

$$dX_t = \kappa(\theta - X_t)dt + \psi d\beta_t,$$

(2.6)

with initial condition \(X_0\). Here \{\beta_t\} is a \(Q\)-Brownian motion that is independent of \{\(B_t\)\}, \(\theta\) is the mean reversion level, \(\kappa\) is the mean reversion rate, and \(\psi\) is the dividend volatility. We shall look at the constant parameter (time homogeneous) case first, and then extend the results into time-dependent (time inhomogeneous) situation. The linear stochastic equation (2.6) can easily be solved to yield the solution

$$X_t = e^{-\kappa t}X_0 + \theta(1 - e^{-\kappa t}) + \psi e^{-\kappa t} \int_0^t e^{\kappa s} d\beta_s.$$

(2.7)

An elementary way of establishing this is to apply Ito’s lemma on \(f(X_t, t) = X_t e^{\kappa t}\).

The Gaussian process (2.7) is fully specified by its mean

$$E[X_t] = e^{-\kappa t}X_0 + \theta(1 - e^{-\kappa t})$$

(2.8)

and the covariance

$$\text{Cov}[X_t, X_T] = \frac{\psi^2}{2\kappa} e^{-\kappa T} (e^{\kappa T} - e^{-\kappa T}).$$

(2.9)

In particular, setting \(T = t\) in (2.9) we find that

$$\text{Var}[X_t] = \frac{\psi^2}{2\kappa} (1 - e^{-2\kappa t}).$$

(2.10)
2.4 Properties of the Ornstein-Uhlenbeck process

In what follows we present some elementary but perhaps not entirely obvious properties of the OU process, which will be required for the subsequent commodity pricing analysis. In particular, here we highlight two orthogonal decomposition properties of the OU process.

2.4.1 Reinitialization and orthogonal decomposition

We begin by noting that the OU process possesses the reinitialisation property in the sense that

$$X_T = e^{-\kappa(T-t)}X_t + \theta(1 - e^{-\kappa(T-t)}) + \psi e^{-\kappa T} \int_t^T e^{\kappa u} d\beta_u.$$  \hspace{1cm} (2.11)

This follows from a direct substitution of (2.7). Now since \(\{X_t\}\) is a Gaussian process, by use of the variance-covariance relations, one can easily verify that the random variables \(X_t\) and \(X_T - e^{-\kappa(T-t)}X_t\) are independent. This property implies that an OU process admits an orthogonal decomposition of the form

$$X_T = (X_T - e^{-\kappa(T-t)}X_t) + e^{-\kappa(T-t)}X_t,$$  \hspace{1cm} (2.12)

for \(T > t\). Note that when the mean reversion rate \(\kappa\) is set to zero this relation reduces to the usual independent increments decomposition for Brownian motion.

2.4.2 The Ornstein-Uhlenbeck bridge

Interestingly, there is another, perhaps somewhat less appreciated, orthogonal decomposition, associated with the OU process. This decomposition takes the form

$$X_t = \left( X_t - \frac{e^{\kappa t} - e^{-\kappa t}}{e^{\kappa T} - e^{-\kappa T}} X_T \right) + \frac{e^{\kappa t} - e^{-\kappa t}}{e^{\kappa T} - e^{-\kappa T}} X_T.$$  \hspace{1cm} (2.13)

The process \(\{b_{tT}\}_{0 \leq t \leq T}\) defined for fixed \(T\) by

$$b_{tT} = X_t - \frac{e^{\kappa t} - e^{-\kappa t}}{e^{\kappa T} - e^{-\kappa T}} X_T,$$  \hspace{1cm} (2.14)

appearing in the decomposition (2.13) is the Ornstein-Uhlenbeck bridge (OU bridge). Figure 2.1 shows sample paths of the OU bridge process, as well as its mean and variance. With the help of the hyperbolic functions, we can expressing the OU bridge in the form

$$b_{tT} = X_t - \frac{\sinh(\kappa t)}{\sinh(\kappa T)} X_T.$$  \hspace{1cm} (2.15)
Clearly, we have that
\[ b_0T = X_0 \] and \[ b_T T = 0. \] (2.16)

By use of the covariance relations, one can check that \( b_t \) and \( X_T \) are independent. Note that the OU bridge is a Gaussian process with mean
\[
\mathbb{E}[b_T b] = \frac{\sinh(\kappa(T - t))}{\sinh(\kappa T)} X_0 + \left[ 1 - \frac{\sinh(\kappa t) + \sinh(\kappa(T - t))}{\sinh(\kappa T)} \right] \theta, \quad (2.17)
\]
and variance
\[
\text{Var}[b_T] = \frac{\variance^2}{2\kappa} \left[ \frac{\cosh(\kappa T) - \cosh(\kappa(T - 2t))}{\sinh(\kappa T)} \right]. \quad (2.18)
\]

The main observation one can make about the OU bridge is the independence of \( b_t \) and \( X_T \), which will become very useful in what follows when we prove the Markovian property of the generating processes of the joint filtration.

### 2.5 Commodity pricing formula

Putting together all the properties we have derived from the last few sections, we are now ready to calculate the price of the commodity. To achieve this, we need to focus on the expectation below:
\[
S_t = \frac{1}{\mathcal{F}_t} \mathbb{E}_t \left[ \int_t^\infty P_u X_u du \right]. \quad (2.19)
\]

**Figure 2.1**: Two sample paths of OU Bridge with the following parameters: \( X_0 = 0.5 \), \( \theta = 1.2 \), \( \psi = 0.4 \), \( \kappa = 0.2 \), \( T = 1 \). The number of steps is 365. The mean (red) and variance (blue) of the OU bridge are also plotted.
Note that the conditioning here is with respect to the joint filtration (2.5). In general, performing such a conditional expectation is difficult. However, we have the following result that simplifies the computation considerably.

**Proposition 2.5.1.** The information process \( \{ \xi_t \} \) and the convenience dividend process \( \{ X_t \} \) are jointly Markovian, implying that, the following relation holds:

\[
E \left[ \int_t^\infty P_u X_u du \middle| \xi_s, X_s \right] = E \left[ \int_t^\infty P_u X_u du \middle| \xi_t, X_t \right].
\]  

The details of the proof of this proposition can be found in the Appendix A. From the orthogonal decomposition (2.12), we can isolate the dependence of the commodity price on the current level of the convenience dividend \( X_t \) and, remarkably, this turns out to be linear in our model.

In particular, we have the following decomposition of the discounted cumulative future dividend flow into orthogonal components:

\[
\int_t^\infty P_u X_u du = \int_t^\infty P_u \left( X_u - e^{-\kappa(u-t)} X_t \right) du + \left( \int_t^\infty P_u e^{-\kappa(u-t)} du \right) X_t.
\]  

Note that the independence of the two terms on the right side of (2.21) can easily be checked by considering their covariance, since both terms are Gaussian. It follows from (2.2) and the orthogonal decomposition property that the commodity price can be expressed in the form:

\[
S_t P_t = E \left[ \int_t^\infty P_u \left( X_u - e^{-\kappa(u-t)} X_t \right) du \middle| \xi_t, X_t \right] + E \left[ \left( \int_t^\infty P_u e^{-\kappa(u-t)} du \right) X_t \middle| \xi_t, X_t \right].
\]  

We now observe that, since the term \( X_u - e^{-\kappa(u-t)} X_t \) is independent of \( X_t \), the conditioning with respect to \( X_t \) in the first expectation in the right side of (2.22) drops out, and we obtain

\[
S_t P_t = E [A_t \sigma t + B_t] + \left( \int_t^\infty P_u e^{-\kappa(u-t)} du \right) X_t,
\]  

where

\[
A_t = \int_t^\infty P_u \left( X_u - e^{-\kappa(u-t)} X_t \right) du,
\]  

and \( B_t \) is the value of the Brownian motion at time \( t \). Note that \( \sigma t + B_t = \xi_t \). Now we are working with the expectation of the form \( E [A | A + B] \), where \( A \) and \( B \) are independent Gaussian random variables, each with a known mean and variance. More
specifically, we have:

\[ A = \int_t^\infty P_u \left( X_u - e^{-\kappa(u-t)}X_t \right) du, \]  

(2.25)

and

\[ B = \frac{B_t}{\sigma t}. \]  

(2.26)

In order to compute the expectation in (2.23), the following lemma will become useful.

**Lemma 2.5.1.** Let \( A, B \) be a pair of independent Gaussian random variables, and write

\[ A = z(A + B) + (1 - z)A - zB. \]  

(2.27)

Then \( A + B \) and \( (1 - z)A - zB \) are orthogonal, and hence independent, if we set

\[ z = \frac{\text{Var}[A]}{\text{Var}[A] + \text{Var}[B]}. \]  

(2.28)

**Proof.** Since \( A + B \) and \( (1 - z)A - zB \) are Gaussian random variables, it suffices to check that

\[ \text{Cov}[(A + B), (1 - z)A - zB] = 0. \]  

(2.29)

After a straightforward calculation, we deduced that the necessary condition for (2.29) to hold is for \( z \) to be equal to (2.28).

\[ \square \]

It follows from the above lemma that is \( A \) and \( B \) are independent and Gaussian, and if \( z \) is given by (2.28), then we have

\[ \mathbb{E}[A|A + B] = z(A + B) + (1 - z)\mathbb{E}[A] - z\mathbb{E}[B]. \]  

(2.30)

To compute the conditional expectation \( \mathbb{E}[A|A + B] \), we are thus required to work out the means and the variances of the random variables \( A \) and \( B \). To begin, let us consider

\[ \mathbb{E}[A] = \mathbb{E} \left[ \int_t^\infty P_u \left( X_u - e^{-\kappa(u-t)}X_t \right) du \right]. \]  

(2.31)

If we recall the reinitialisation property of the OU process

\[ X_u = e^{-\kappa(u-t)}X_t + \theta(1 - e^{-\kappa(u-t)}) + \psi e^{-\kappa u} \int_t^u e^{\kappa v} d\beta_v. \]  

(2.32)
and substitute this in \((2.31)\), then we find

\[
E[A] = E\left[ \theta \int_t^\infty P_u \left( 1 - e^{-\kappa (u-t)} \right) du + \psi \int_t^\infty e^{-\kappa u} P_u \int_t^u e^{-\kappa s} d\beta_s du \right]
\]

\[
= \theta E\left[ \int_t^\infty P_u du \right] - \theta E\left[ \int_t^\infty P_u e^{-\kappa (u-t)} du \right]
+ \psi E\left[ \int_t^\infty e^{-\kappa u} P_u \int_t^u e^{-\kappa s} d\beta_s du \right].
\]

\[(2.33)\]

For the last term of \((2.33)\), interchange order of integration. We have

\[
\psi E\left[ \int_{u=t}^{\infty} e^{-\kappa u} P_u \int_{s=t}^{u} e^{-\kappa s} d\beta_s du \right]
= \psi E\left[ \int_{s=t}^{\infty} e^{-\kappa s} \int_{u=s}^{\infty} P_u e^{-\kappa u} du d\beta_s \right]
= 0.
\]

\[(2.34)\]

Therefore, we are left with the expression

\[
E[A] = \theta E\left[ \int_t^\infty P_u du \right] - \theta E\left[ \int_t^\infty P_u e^{-\kappa (u-t)} du \right]
\]

\[(2.35)\]

Note also that

\[
E[B] = E\left[ \frac{B_t}{\sigma t} \right] = 0,
\]

\[(2.36)\]

where \(B_t\) is a standard Brownian motion. On the other hand, we have

\[
A + B = \int_t^\infty P_u \left( X_u - e^{-\kappa (u-t)} X_t \right) du + \frac{B_t}{\sigma t}
\]

\[
= \int_t^\infty P_u X_u du - \int_t^\infty P_u X_t e^{-\kappa (u-t)} du + \frac{B_t}{\sigma t}
\]

\[
= \frac{1}{\sigma t} \xi_t - X_t \int_t^\infty P_u e^{-\kappa (u-t)} du.
\]

\[(2.37)\]

In order to simplify the calculation, let us introduce \(p_t\) and \(q_t\) according to

\[
p_t = \int_t^\infty P_u du \quad \text{and} \quad q_t = \int_t^\infty P_u e^{-\kappa (u-t)} du.
\]

\[(2.38)\]

We then have:

\[
P_t S_t = (1 - z_t) \left[ \theta p_t + q_t (X_t - \theta) \right] + z_t \frac{\xi_t}{\sigma t}.
\]

\[(2.39)\]
Now from
\[ A - \mathbb{E} [ A ] = \psi \int_t^\infty e^{\kappa s} \left( \int_s^\infty P_u e^{-\kappa u} du \right) d\beta_s \]
\[ = \psi \int_t^\infty \int_s^\infty P_u e^{-\kappa (u-s)} du \ d\beta_s \]
\[ = \psi \int_t^\infty q_s d\beta_s, \quad (2.40) \]
we find that the variance of \( A \) is
\[ \text{Var} [ A ] = \mathbb{E} \left[ (A - \mathbb{E} [ A ])^2 \right] \]
\[ = \psi^2 \mathbb{E} \left[ \int_t^\infty q_s d\beta_s \right]^2 \]
\[ = \psi^2 \int_t^\infty q_s^2 ds, \quad (2.41) \]
on account of the Wiener-Ito isometry. The variance of \( B \) is
\[ \text{Var} [ B ] = \text{Var} \left[ \frac{B_t}{\sigma^2 t^2} \right] = \frac{1}{\sigma^4 t}. \quad (2.42) \]

Putting these together, we obtain the following explicit formula for the commodity price:
\[ S_t = (1 - z_t) P_t^{-1} \left[ \theta p_t + q_t (X_t - \theta) \right] + z_t P_t^{-1} \frac{\xi_t}{\sigma t}, \quad (2.43) \]
where
\[ z_t = \frac{\sigma^2 \psi^2 t \int_t^\infty q_s^2 ds}{1 + \sigma^2 \psi^2 t \int_t^\infty q_s^2 ds}, \quad (2.44) \]
with \( p_t \) and \( q_t \) given in (2.38).

**Remark 2.5.1.** From the expression of the weighting factor \( z_t \) in (2.44), we see that for large \( \psi \) and (or) large \( \sigma \), the value of \( z_t \) tends to unity. On the other hand, for small \( \psi \) and (or) small \( \sigma \), the value of \( z_t \) tends to zero. Hence, if the market information has a low noise content, i.e. high information flow rate \( \sigma \), then the market information is what mainly determines the price of the commodity. On the other hand, if the volatility of the convenience dividend is high, then market participants also rely heavily on “the best information about the future” in their determination of prices, rather than simply assuming that the current value of the dividend is a good guide to the future.

The other term, that is, the term proportional to \((1 - z_t)\) in the expression for \( S_t \) is essentially an annuities valuation of a constant dividend rate set at the mean reversion level, together with a correction term to adjust for the present level of the dividend rate.
This term dominates in situations when the market information is of low quality. It also dominates in situations when the dividend volatility is low. In other words, in the absence of significant information concerning potential future return, our judgements are formed on the basis of a kind of average of the status quo and the market consensus regarding long term average. But we also rely on the status quo in situations where there is little uncertainty, i.e. when the dividend volatility is very low.

## 2.6 Special case of constant interest rates

Under the assumption of a constant interest rate, we can make further simplifications on the commodity valuation formulae (2.43) and (2.44). Whether this is a valid approximation or not depends on the application one has in mind, but in any case one gains insights by considering the constant interest rate situation. The discount bond price in this case becomes

\[ P_t = e^{-rt}. \] (2.45)

Therefore, we have

\[ p_t = \frac{1}{r} e^{-rt} \quad \text{and} \quad q_t = \frac{1}{r + \kappa} e^{-rt}. \] (2.46)

Substituting this in (2.43), we find, after a short calculation, the pricing formula under constant interest rate setup takes the form

\[ S_t = \left(1 - z_t\right) \frac{\kappa \theta + r X_t}{r (r + \kappa)} + z_t e^{rt} \frac{\xi_t}{\sigma t}, \] (2.47)

with the weighting factor

\[ z_t = \frac{\sigma^2 \psi^2 t}{2 r (r + \kappa)^2} e^{2rt} + \sigma^2 \psi^2 t. \] (2.48)

**Remark 2.6.1.** The following observations can be made from a closer inspection of the expression (2.47) for the commodity price:

1. The first term is a kind of annuitised weighted average of the mean reversion level and the current level of the dividend, multiplied by the weighting factor \((1 - z_t)\).
2. If we set \(\sigma = 0\), then the first term alone determines the price of the commodity.
3. The second term then modifies the price by bringing in the market information available about the future dividend stream.
2.7 Alternative derivation of commodity price process

In this section, an alternative derivation of the price process is presented. The purpose of this is to derive a number of other relations that in turn will provide a way of identifying the innovations arising in connection with the associated filtering problem. For this purpose, let us consider the case when interest rate is constant. Therefore, we have the purpose, let us consider the case when interest rate is constant. Therefore, we have

\[ P_tS_t = \mathbb{E}_t \left[ \int_t^\infty P_u X_u du \right] = \mathbb{E}_t \left[ \int_t^\infty e^{-ru} X_u du \right] 
\]

From (2.7) the integral term in the expectation in (2.49) can be simplified according to

\[ \int_0^\infty e^{-ru} X_u du = \int_0^\infty e^{-ru} \left[ e^{-\kappa u} X_0 + \theta (1 - e^{-\kappa u}) + \psi e^{-\kappa u} \int_0^u e^{\kappa s} d\beta_s \right] du 
\]

Solving the first two integrals and rearrange terms, we get

\[ \int_0^\infty e^{-ru} X_u du = \frac{X_0}{r + \kappa} + \theta \left( \frac{1}{r} - \frac{1}{r + \kappa} \right) + \psi \int_0^\infty e^{-(r+\kappa)u} \int_0^u e^{\kappa s} d\beta_s du 
\]

For the double integral, one can use a change of variables to interchange the order of integration to yield:

\[ \int_0^\infty e^{-ru} X_u du = \frac{rX_0 + \kappa \theta}{r (r + \kappa)} + \psi \int_0^\infty e^{-(r+\kappa)u} \int_0^u e^{\kappa s} d\beta_s du 
\]

Then, substituting (2.52) back into (2.49), we have

\[ P_tS_t = \mathbb{E}_t \left[ \frac{rX_0 + \kappa \theta}{r (r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^\infty e^{-rs} d\beta_s \right] - \int_0^t e^{-ru} X_u du 
\]

\[ = \mathbb{E}_t \left[ \frac{rX_0 + \kappa \theta}{r (r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-rs} d\beta_s + \frac{\psi}{r + \kappa} \int_0^t e^{-rs} d\beta_s \right] - \int_0^t e^{-ru} X_u du 
\]

\[ = \frac{rX_0 + \kappa \theta}{r (r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-rs} d\beta_s + \frac{\psi}{r + \kappa} \mathbb{E}_t \left[ \int_t^\infty e^{-rs} d\beta_s \right] - \int_0^t e^{-ru} X_u du. \]
The problem is now reduced to the determination of the conditional expectation
\[ E \left[ \int_t^\infty e^{-rs} \, d\beta_s \ \bigg| \ {\xi_s}_{0\leq s\leq t}, \ {X_s}_{0\leq s\leq t} \right]. \] (2.54)

Before we begin to solve the problem, the following observations can be made. Recall that the model filtration is generated jointly by \( \{\xi_t\} \) and \( \{X_t\} \). Furthermore, \( \xi_t \) can be expressed in the form:
\[
\xi_t = \sigma t \int_t^\infty e^{-ru} X_u \, du + B_t
= \sigma t \left( \int_0^\infty e^{-ru} X_u \, du - \int_0^t e^{-ru} X_u \, du \right) + B_t
= \sigma t \left( \frac{r X_0 + \kappa \theta}{r + \kappa} + \frac{\psi}{r + \kappa} \int_0^\infty e^{-ru} \, d\beta_u - \int_0^t e^{-ru} X_u \, du \right) + B_t
= \omega_t + \sigma t \left( \frac{r X_0 + \kappa \theta}{r + \kappa} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} \, d\beta_u - \int_0^t e^{-ru} X_u \, du \right). \] (2.55)

Here we have defined
\[
\omega_t = \sigma \psi t \int_t^\infty e^{-ru} \, d\beta_u + B_t := \sigma \psi t Y_t + B_t, \] (2.56)
where
\[
Y_t = \int_t^\infty e^{-ru} \, d\beta_u. \] (2.57)

Note that the integrand in \( Y_t \) is a deterministic function of time. Therefore, for each \( t \geq 0 \), \( Y_t \) is a normal random variable. It is easy to see that the filtration generated jointly by \( \{\xi_t\} \) and \( \{X_t\} \) is equivalent to the filtration generated jointly by \( \{\omega_t\} \) and \( \{X_t\} \). Therefore, the conditional expectation appearing in (2.53) can be expressed in the form
\[
E \left[ \int_t^\infty e^{-rs} \, d\beta_s \ \bigg| \ {\xi_s}_{0\leq s\leq t}, \ {X_s}_{0\leq s\leq t} \right] = E \left[ \int_t^\infty e^{-rs} \, d\beta_s \ \bigg| \ \omega_t \right] = \int_{-\infty}^\infty y \pi_t(y) \, dy. \] (2.58)

Here we have made use of the facts that \( Y_t \) is independent of \( \{X_s\}_{0\leq s\leq t} \), and that \( \{\omega_t\} \) is Markov. We write \( \pi_t(y) \) for the conditional density of the random variable \( Y_t \). Specifically, we have:
\[
\pi_t(y) = \frac{d}{dy} Q( Y_t \leq y \mid \omega_t). \] (2.59)
Note that the conditional density function of $Y_t$ can be worked out by using a form of the Bayes formula:

$$\pi_t(y) = \frac{p(y)p(\omega_t|Y = y)}{\int p(y)p(\omega_t|Y = y)dy}. \quad (2.60)$$

Here $p(y)$ denotes the \textit{a priori} density for $Y_t$, which we shall work out shortly, and $\rho(\omega_t|Y = y)$ denotes the conditional density for the random variable $\omega_t$ given that $Y = y$. The fact that $Y_t$ is a Gaussian random variable means we only need the mean and variance to find out the \textit{a priori} density $p(y)$. After a straightforward calculation, we find that $Y_t$ has mean zero and variance $e^{-2rt}/(2r)$. Therefore,

$$p(y) = \sqrt{\frac{re^{2rt}}{\pi}} \exp\left(-\frac{re^{2rt}y^2}{2}\right). \quad (2.61)$$

Also the fact that $B_t$ is a Gaussian random variable with mean zero and variance $t$ implies that conditional on $Y = y$, $\omega_t$ is Gaussian with mean $\sigma\psi ty/(r + \kappa)$ and variance $t$. We thus deduce that the conditional density for $\omega_t$ is

$$\rho(\omega_t|Y = y) = \sqrt{\frac{1}{2\pi t}} \exp\left[-\frac{(\omega_t - \sigma\psi ty/(r + \kappa) y)^2}{2t}\right]. \quad (2.62)$$

Inserting this expression into the Bayes formula we get the \textit{a posteriori} density function:

$$\pi_t(y) = \frac{p(y)\sqrt{\frac{1}{2\pi t}} \exp\left[-\frac{(\omega_t - \sigma\psi ty/(r + \kappa) y)^2}{2t}\right]}{\int p(y)\sqrt{\frac{1}{2\pi t}} \exp\left[-\frac{(\omega_t - \sigma\psi ty/(r + \kappa) y)^2}{2t}\right]dy}$$

$$= \frac{p(y)exp\left[\frac{\sigma\psi wy}{r + \kappa} y - \frac{1}{2} \frac{\sigma^2\psi^2 t}{(r + \kappa)^2} y^2\right]}{\int p(y)exp\left[\frac{\sigma\psi wy}{r + \kappa} y - \frac{1}{2} \frac{\sigma^2\psi^2 t}{(r + \kappa)^2} y^2\right]dy}. \quad (2.63)$$

It follows that for the conditional expectation we can write

$$\mathbb{E}\left[\int_t^\infty e^{-rs} d\beta_s|\omega_t\right] = \int y\pi_t(y)dy$$

$$= \frac{\int yp(y)exp\left[\frac{\sigma\psi wy}{r + \kappa} y - \frac{1}{2} \frac{\sigma^2\psi^2 t}{(r + \kappa)^2} y^2\right]dy}{\int p(y)exp\left[\frac{\sigma\psi wy}{r + \kappa} y - \frac{1}{2} \frac{\sigma^2\psi^2 t}{(r + \kappa)^2} y^2\right]dy}. \quad (2.64)$$
Substituting the *a priori* density $p(y)$ defined in equation (2.61) into equation (2.64), we have

\[
(2.64) = \frac{\int y \exp \left[ \frac{\sigma \psi \omega}{r + \kappa} y - \frac{1}{2} \left( \frac{\sigma^2 \psi^2 t}{(r + \kappa)^2} + 2re^{2rt} \right) y^2 \right] \, dy}{\int \exp \left[ \frac{\sigma \psi \omega}{r + \kappa} y - \frac{1}{2} \left( \frac{\sigma^2 \psi^2 t}{(r + \kappa)^2} + 2re^{2rt} \right) y^2 \right] \, dy},
\]

where

\[
a = \frac{\sigma \psi \omega}{r + \kappa} \quad \text{and} \quad b = \frac{1}{2} \left[ \frac{\sigma^2 \psi^2 t}{(r + \kappa)^2} + 2re^{2rt} \right].
\]

Performing the integration, we thus obtain

\[
E \left[ \int_t^\infty e^{-rs} \, d\beta_s \Big| \omega_t \right] = \frac{a}{2b}.
\]

Substituting $a$ and $b$ of (2.66) in here, we find

\[
E \left[ \int_t^\infty e^{-rs} \, d\beta_s \Big| \omega_t \right] = \frac{\sigma \psi \omega}{r + \kappa} \frac{\sigma^2 \psi^2 \omega}{(r + \kappa)^2} + 2re^{2rt} = \frac{(r + \kappa)\sigma \psi \omega}{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t}
\]

where

\[
z_t = \frac{\sigma^2 \psi^2 t}{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t}
\]

is the weighting factor obtained in the previous method for the calculation presented earlier. Substituting (2.68) into (2.53), together with the fact that expression for the new information process \{\omega_t\} in terms of the previous one \{\xi_t\} is given by

\[
\omega_t = \xi_t - \sigma \left[ \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} \, d\beta_u - \int_0^t e^{-ru} X_u \, du \right],
\]

we find that the pricing formula (2.53) then becomes

\[
P_tS_t = \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-rs} \, d\beta_s + \frac{\psi}{r + \kappa} E_t \left[ \int_t^\infty e^{-rs} \, d\beta_s \right] - \int_0^t e^{-ru} X_u \, du.
\]
Let us consider the expression for $X_t$ expression for $\omega_t$ from equation (2.56) and rearranging terms, we obtain:

$$P_t S_t = \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u + \frac{\psi}{r + \kappa} \frac{r + \kappa}{\sigma \psi t} z_t \omega_t - \int_0^t e^{-ru} X_u du$$

Recalling the expression for $\omega_t$ from equation (2.56) and rearranging terms, we obtain:

$$P_t S_t = \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u - \int_0^t e^{-ru} X_u du$$

The various terms in equation (2.73) then cancel out or group together to give

$$P_t S_t = (1 - z_t) \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\xi_t}{\sigma t}$$

Let us consider the expression $\int_0^t e^{-ru} X_u du$ in equation (2.74). Since we know the expression for $X_t$, and the interest rate is constant, we can solve the integral:

$$\int_0^t e^{-ru} X_u du$$

$$= \frac{X_0 - \theta}{r + \kappa} \left[ 1 - e^{-(r+\kappa)t} \right] + \frac{\theta}{r} \left( 1 - e^{-rt} \right)$$

$$+ \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u - \frac{\psi}{r + \kappa} e^{-(r+\kappa)t} \int_0^t e^{\kappa u} d\beta_u$$

$$= \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u$$

$$- \frac{e^{-rt}}{r(r + \kappa)} \left( rX_0 e^{-\kappa t} + r\theta e^{-\kappa t} + \frac{r\psi}{r} e^{-\kappa t} \int_0^t e^{\kappa u} d\beta_u + \kappa \theta + r \theta \right).$$

After further cancellation and rearrangement of terms, we get

$$\int_0^t e^{-ru} X_u du = \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u - \frac{e^{-rt}}{r(r + \kappa)} (rX_t - r \theta + \kappa \theta + r \theta)$$

$$= \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u - e^{-\kappa t} \frac{r \theta + \kappa \theta + r \theta}{r(r + \kappa)}. $$

(2.76)
Chapter 2. Commodity Prices

Inserting the above expression (2.76) in the pricing formula (2.74), we obtain

\[ P_t S_t = \left(1 - z_t\right) \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \left(1 - z_t\right) \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u + z_t \frac{\xi_t}{\sigma t} - \left(1 - z_t\right) \left[ \frac{rX_0 + \kappa \theta}{r(r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u - e^{-rt} \frac{\kappa \theta + rX_t}{r(r + \kappa)} \right]. \]

Finally, with further rearrangement of the various terms, we have

\[ S_t = \left(1 - z_t\right) \frac{\kappa \theta + rX_t}{r(r + \kappa)} + z_t e^{rt} \frac{\xi_t}{\sigma t}, \tag{2.77} \]

which agrees with the result in equation (2.47) derived from the previous approach.

In terms of the alternative information process \( \{\omega_t\} \), the price process can be expressed in the form

\[ S_t = \frac{\kappa \theta + rX_t}{r(r + \kappa)} + z_t e^{rt} \frac{\xi_t}{\sigma t}, \tag{2.78} \]

where

\[ \omega_t = \frac{\sigma \psi}{r + \kappa} \int_t^\infty e^{-ru} d\beta_u + B_t. \tag{2.79} \]

We see, therefore, that although this alternative derivation involves a somewhat more elaborate set of calculations, the end result is a relatively simple expression that is going to be very useful in the study below of the dynamics of the price process.

2.8 Price dynamics and innovation representation

In the previous sections, we have shown that the commodity price process can be expressed in two different forms. In this section, we are going to use the simpler expression (2.78) to derive the price dynamics, and derive the innovations representation, which in turn provides the “observable” driving Brownian motion for the price.

Applying Ito’s lemma on the price process (2.78), we have

\[ dS_t = \frac{r}{r(r + \kappa)} dX_t + e^{rt} z_t \frac{\omega_t}{\sigma t} dt - e^{rt} z_t \frac{\omega_t}{\sigma t^2} dt + e^{rt} \frac{\omega_t}{\sigma t} dz_t + \frac{e^{rt} z_t}{\sigma t} d\omega_t. \tag{2.80} \]

Recall that from equation (2.6), one gets

\[ dS_t = \frac{r}{r(r + \kappa)} (\kappa \theta dt - \kappa X_t dt + \psi d\beta_t) + \left[ S_t - \frac{\kappa \theta + rX_t}{r(r + \kappa)} \right] dt - e^{rt} z_t \frac{\omega_t}{\sigma t^2} dt + e^{rt} \frac{\omega_t}{\sigma t} dz_t + \frac{e^{rt} z_t}{\sigma t} d\omega_t. \tag{2.81} \]
Rearranging terms, we have

\[ dS_t = \frac{r\kappa\theta}{r(r+\kappa)} dt - \kappa \frac{rX_t}{r(r+\kappa)} dt + \frac{r\psi}{r(r+\kappa)} d\beta_t + rS_t dt - \frac{r\kappa\theta}{r(r+\kappa)} dt \]

\[-r \frac{rX_t}{r(r+\kappa)} dt - e^{rt} z_t \frac{\omega_t}{\sigma t^2} dt + e^{rt} \frac{\omega_t}{\sigma t} dz_t + \frac{e^{rt} z_t}{\sigma t} d\omega_t \]

\[ = (rS_t - X_t)_t \]

\[ + e^{rt} \left( \frac{\psi}{r+\kappa} e^{-rt} d\beta_t - \frac{z_t \omega_t}{\sigma t^2} dt + \frac{\omega_t}{\sigma t} dz_t + \frac{z_t}{\sigma t} d\omega_t \right). \tag{2.82} \]

We now need the dynamics of the information process \( \{\omega_t\} \) and for the weighting factor \( \{z_t\} \). These are given by

\[ d\omega_t = \frac{\omega_t - B_t}{t} dt - \frac{\sigma \psi t}{r+\kappa} e^{-rt} d\beta_t + dB_t \tag{2.83} \]

and

\[ dz_t = z_t (1 - z_t) \frac{1 - 2rt}{t} dt. \tag{2.84} \]

Therefore, substituting these expressions into (2.82) and making further rearrangements of terms, we obtain:

\[ dS_t = (rS_t - X_t)_t dt + \frac{\psi}{r+\kappa} d\beta_t - e^{rt} z_t \frac{\omega_t}{\sigma t^2} dt + e^{rt} \frac{\omega_t z_t}{\sigma t^2} dt \]

\[ + \frac{e^{rt} z_t}{\sigma t} \left( \frac{\omega_t - B_t}{t} dt - \frac{\sigma \psi t}{r+\kappa} e^{-rt} d\beta_t + dB_t \right) \]

\[ = (rS_t - X_t)_t dt + \frac{\psi}{r+\kappa} d\beta_t - \frac{2re^{rt} \omega t z_t}{\sigma t^2} dt - \frac{e^{rt} \omega t z_t^2}{\sigma t^2} dt \]

\[ + \frac{e^{rt} \omega t z_t}{\sigma t^2} dt - \frac{e^{rt} B_t z_t}{\sigma t^2} dt - \frac{\psi}{r+\kappa} z_t d\beta_t + \frac{e^{rt} z_t}{\sigma t} dB_t \]

\[ = (rS_t - X_t)_t dt + \frac{\psi}{r+\kappa} (1 - z_t) d\beta_t + \frac{e^{rt} \omega t z_t}{\sigma t^2} dt - \frac{2re^{rt} \omega t z_t}{\sigma t} z_t (1 - z_t) dt \]

\[ - \frac{e^{rt} B_t z_t}{\sigma t^2} dt + \frac{e^{rt} z_t}{\sigma t} dB_t. \tag{2.85} \]

This can be further simplified if we make use of the expression for the conditional variance of the discounted future dividends \( \{V_t\} \) obtained in Appendix D. Specifically,
we have
\[
dS_t = (rS_t - X_t) \, dt + e^{rt} \sigma V_t \left[ \frac{2r (r + \kappa) e^{rt}}{\sigma \psi} \, d\beta_t + \omega_t \left( 1 - z_t \right) dt - 2r \omega_t (1 - z_t) dt - \frac{B_t}{t} \, dt + dB_t \right]
\]
\[
= (rS_t - X_t) \, dt + e^{rt} \sigma V_t \left\{ C_t \left[ \frac{2r (r + \kappa) e^{rt}}{\sigma \psi} \, d\beta_t + \frac{(1 - 2rt)(1 - z_t)}{t} \omega_t dt - \frac{B_t}{t} \, dt + dB_t \right] \right\},
\]
where
\[
C_t = \frac{\sigma \psi}{\sqrt{(2r)^2 (r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2}}
\]
and
\[
V_t = \frac{z_t}{\sigma^2 t}.
\]
Here $V_t$ is the conditional variance of the future dividend cash flow $\int_t^\infty e^{-ru} X_u \, du$, and \{\{C_t\}\} is a common factor that has been added for reasons to be made clear in what follows.

As one can see, the drift term of the dynamics is in the form of that of a standard dividend paying asset. Therefore, it makes sense to postulate the existence of a Brownian motion \{\{W_t\}\} that gives rise to the volatility term in (2.86). Indeed, we have:

**Proposition 2.8.1.** The dynamics of the commodity price process can be written in the following form:
\[
dS_t = (rS_t - X_t) \, dt + \Sigma_t \, dW_t,
\]
where \{\{W_t\}\} is a standard \{\{\mathcal{F}_t\}\}-Brownian motion on the probability space \((\Omega, \mathcal{F}, \mathbb{Q})\), and where the process \{\{\Sigma_t\}\} defined by
\[
\Sigma_t = e^{rt} \sigma V_t \frac{C_t}{C_t}
\]
is the absolute volatility of the commodity price.

**Proof.** We have already shown that
\[
\mathbb{E}^\mathbb{Q} \left[ \int_t^\infty P_u X_u \, du \middle| \{\xi_s\}_{0 \leq s \leq t}, \{X_s\}_{0 \leq s \leq t} \right] = \mathbb{E}^\mathbb{Q} \left[ \int_t^\infty P_u X_u \, du \middle| \xi_t, X_t \right].
\]
We can also rearrange the valuation formula of the commodity price into the form

\[ P_t S_t = \mathbb{E}^Q \left[ \int_t^\infty P_u X_u \, du \middle| \mathcal{F}_t \right]. \tag{2.92} \]

Denoting the conditional expectation under \( Q \) with respect to \( \mathcal{F}_t \) by the notation \( \mathbb{E}_t[-] \), and taking out the measurable part in (2.92), we have

\[ P_t S_t = \mathbb{E}_t \left[ \int_t^\infty P_u X_u \, du \right] = \mathbb{E}_t \left[ \int_0^\infty P_u X_u \, du \right] - \int_0^t P_u X_u \, du. \tag{2.93} \]

Define a process \( \{M_t\} \) according to

\[ M_t = \mathbb{E}_t \left[ \int_0^\infty P_u X_u \, du \right] = P_t S_t + \int_0^t P_u X_u \, du. \tag{2.94} \]

It should be evident that \( \{M_t\} \) is an \( \{\mathcal{F}_t\} \)-martingale. This can be shown as follows. Letting \( 0 \leq s < t \), we have

\[ \mathbb{E}_s [M_t] = \mathbb{E}_s \left[ P_t S_t + \int_0^t P_u X_u \, du \right] = \mathbb{E}_s \left[ \mathbb{E}_t \left[ \int_t^\infty P_u X_u \, du \right] \right] + \mathbb{E}_s \left[ \int_0^t P_u X_u \, du \right]. \tag{2.95} \]

Here we have replaced the term \( P_t S_t \) by using its expression in equation (2.92). It follows that

\[ \mathbb{E}_s [M_t] = \mathbb{E}_s \left[ \int_t^\infty P_u X_u \, du \right] + \mathbb{E}_s \left[ \int_s^t P_u X_u \, du + \int_s^t P_u X_u \, du \right] \\
= \mathbb{E}_s \left[ \int_s^\infty P_u X_u \, du - \int_s^t P_u X_u \, du \right] \\
+ \int_s^t P_u X_u \, du + \mathbb{E}_s \left[ \int_s^t P_u X_u \, du \right], \tag{2.96} \]

where we have taken the conditional expectation on both terms in order to cancel out two of the identical terms. Therefore,

\[ \mathbb{E}_s [M_t] = \mathbb{E}_s \left[ \int_s^\infty P_u X_u \, du \right] - \mathbb{E}_s \left[ \int_s^t P_u X_u \, du \right] + \mathbb{E}_s \left[ \int_0^t P_u X_u \, du \right] + \mathbb{E}_s \left[ \int_s^t P_u X_u \, du \right] \\
= \mathbb{E}_s \left[ \int_s^\infty P_u X_u \, du \right] + \mathbb{E}_s \left[ \int_0^t P_u X_u \, du \right] \\
= P_s S_s + \int_0^s P_u X_u \, du \\
= M_s, \tag{2.97} \]
which proves the claim that \( \{ M_t \} \) is an \( \{ F_t \} \)-martingale. On the other hand, by inspecting the relations

\[
M_t = P_t S_t + \int_0^t P_u X_u du, 
\]

\[(2.98)\]

\[
P_t S_t = e^{-rt} \kappa \theta + r X_t + \frac{z_t \omega_t}{r (r + \kappa)}.
\]

\[(2.99)\]

and

\[
\int_0^t P_u X_u du = \frac{r X_0 + \kappa \theta}{r (r + \kappa)} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u - \frac{e^{-rt} \kappa \theta + r X_t}{r (r + \kappa)}.
\]

\[(2.100)\]

we find that

\[
M_t = P_t S_t + \int_0^t P_u X_u du \\
= \frac{z_t \omega_t}{\sigma t} + \frac{\psi}{r + \kappa} \int_0^t e^{-ru} d\beta_u + \frac{r X_0 + \kappa \theta}{r (r + \kappa)}.
\]

\[(2.101)\]

Applying Ito’s lemma, we obtain

\[
dM_t = \frac{\psi}{r + \kappa} e^{-rt} d\beta_t - \frac{z_t \omega_t}{\sigma t} dt + \frac{\omega_t}{\sigma t} dz_t + \frac{z_t}{\sigma t} d\omega_t.
\]

\[(2.102)\]

It is encouraging to realise that (2.102) is the same as the term in the brackets on the right-hand side of (2.82). Now from the martingale representation theorem we know that there exists an \( \{ F_t \} \)-Brownian motion \( \{ W_t \} \) on \((\Omega, \mathcal{F}, \mathbb{Q})\) and an adapted process \( \{ \Gamma_t \} \) such that we can write

\[
M_t = M_0 + \int_0^t \Gamma_u dW_u,
\]

\[(2.103)\]

or equivalently

\[
dM_t = \Gamma_t dW_t.
\]

\[(2.104)\]

Therefore, the two equivalent expressions for the dynamics of the commodity price in (2.82) and (2.86) can be compared directly:

\[
dS_t = (r S_t - X_t) dt + e^{rt} \left( \frac{\psi}{r + \kappa} e^{-rt} d\beta_t - \frac{z_t \omega_t}{\sigma t^2} dt + \frac{\omega_t}{\sigma t} dz_t + \frac{z_t}{\sigma t} d\omega_t \right)
\]

\[
= (r S_t - X_t) dt + e^{rt} dM_t;
\]

\[(2.105)\]
and

\[ dS_t = (rS_t - X_t) \, dt + e^{rt} \sigma V_t \left\{ C_t \left[ \frac{2r (r + \kappa) e^{rt}}{\sigma \psi} \, d\beta_t + \frac{(1 - 2rt)(1 - z_t)}{t} \omega_t \, dt - \frac{B_t}{t} \, dt + dB_t \right] \right\} = (rS_t - X_t) \, dt + e^{rt} \Gamma_t \, dW_t. \] (2.106)

From the second expression, we observe, on one hand, that

\[ (dS_t)^2 = e^{2rt} \sigma^2 V_t^2 \, dt. \] (2.107)

On the other hand, from (2.86), we have

\[
(dS_t)^2 = e^{2rt} \sigma^2 V_t^2 \left\{ C_t^2 \left[ \frac{2r (r + \kappa) e^{rt}}{\sigma \psi} \, d\beta_t + \frac{(1 - 2rt)(1 - z_t)}{t} \omega_t \, dt - \frac{B_t}{t} \, dt + dB_t \right]^2 \right\} = e^{2rt} \sigma^2 V_t^2 \times \frac{(2r)^2 (r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2}{\sigma^2 \psi^2} \, dt = e^{2rt} \sigma^2 V_t^2 \frac{2r}{C_t^2} \, dt.
\] (2.108)

Since \( \Gamma_t \) is unique, then we must have

\[ \Gamma_t = \frac{\sigma V_t}{C_t} \] (2.109)

Given (2.109), we conclude that the following relation must hold:

\[ dW_t = C_t \left[ \frac{2r (r + \kappa) e^{rt}}{\sigma \psi} \, d\beta_t + \frac{(1 - 2rt)(1 - z_t)}{t} \omega_t \, dt - \frac{B_t}{t} \, dt + dB_t \right]. \] (2.110)

Hence, we have the desired result.

Note that the innovations process \( \{W_t\} \) defined by (2.110) is a standard Brownian motion on the probability space \( (\Omega, \mathcal{F}, \mathbb{Q}) \), and is adapted to the market filtration generated by the convenience dividend and the information process. An interesting observation can be made about the dynamics, namely, the volatility term is expressed in terms of the conditional variance of the future dividend cash flow. This is consistent with the basic structure of the general BHM framework, where one observes similar phenomena in more elementary situations.
2.9 Monte Carlo simulation of crude oil price process

We can run some simulations to examine the model behaviour of the crude oil price process in the time-homogeneous setup. That is to say, by varying different parameters, we can gain intuition about how the model behaves.

Under the constant interest rate setup, the first step is to simulate the convenience dividend process \( \{X_t\} \). From the re-initialization property, we have for \( 0 \leq s < t \) that

\[
X_t = e^{-\kappa(t-s)}X_s + \theta \left(1 - e^{-\kappa(t-s)}\right) + \psi e^{-\kappa t} \int_s^t e^{\kappa u} d\beta_u, \tag{2.111}
\]

and we can easily compute the expectation and variance of \( X_t \):

\[
E[X_t] = e^{-\kappa(t-s)}X_s + \theta \left(1 - e^{-\kappa(t-s)}\right), \tag{2.112}
\]

and

\[
\text{Var}[X_t] = \frac{\psi^2}{2\kappa} \left(1 - e^{-2\kappa(t-s)}\right). \tag{2.113}
\]

To simulate \( \{X_t\} \), for \( 0 = t_0 < t_1 < \cdots < t_n \) we write

\[
X_{t_{i+1}} = e^{-\kappa(t_{i+1}-t_i)}X_{t_i} + \theta \left(1 - e^{-\kappa(t_{i+1}-t_i)}\right) + \sqrt{\frac{\psi^2}{2\kappa} \left(1 - e^{-2\kappa(t_{i+1}-t_i)}\right)} Z, \tag{2.114}
\]

with \( Z_1, \ldots, Z_n \) drawn independently and randomly from \( \mathcal{N}(0, 1) \).

The second step is to simulate the information process \( \{\xi_t\}_{t \geq 0} \). Recall that the expression for the information process is given by

\[
\xi_t = \sigma t \int_t^\infty \left[X_0 - \theta + \frac{\theta}{r + \kappa} e^{-\kappa s} + \frac{1}{r + \kappa} \int_0^t e^{-rs} d\beta_s\right] + B_t. \tag{2.115}
\]

Substituting the expression for \( X_t \) into the above equation, and rearranging terms, we have

\[
\xi_t = \sigma t \left[\frac{X_0 - \theta}{r + \kappa} e^{-\kappa t} + \frac{\theta}{r} e^{-rt}\right] + \sigma t \psi \left[\frac{e^{-\kappa t}}{r + \kappa} \int_0^t e^{\kappa s} d\beta_s + \frac{1}{r + \kappa} \int_t^\infty e^{-rs} d\beta_s\right] + B_t. \tag{2.116}
\]

Observe that since \( \{\xi_t\} \) is expressed as a sum of Gaussian processes, \( \{\xi_t\} \) itself is also a Gaussian process. Furthermore, since the two Brownian motions \( \{\beta_t\} \) and \( \{B_t\} \) are independent, we can easily work out the mean and variance of the information process:

\[
\mathbb{E}[\xi_t] = \mu(t) = \sigma t \left[\frac{X_0 - \theta}{r + \kappa} e^{-\kappa t} + \frac{\theta}{r} e^{-rt}\right], \tag{2.117}
\]
Chapter 2. Commodity Prices

and

\[ \text{Var}[\xi_t] = \gamma^2(t) = \frac{\sigma^2 \psi^2 t e^{-2rt}}{2r \kappa (r + \kappa)^2} \left[ r + \kappa - re^{-2rt} \right] + t. \] (2.118)

Therefore, to simulate \( \{\xi_t\} \), we make use of the following relationship

\[
\xi_{t+1} = \xi_t + \mu(t_i, t_{i+1}) + \gamma(t_i, t_{i+1}) Z \\
= \xi_t + \sigma \Delta t_i \left[ \frac{X_0 - \theta}{r + \kappa} e^{-(r+\kappa)\Delta t_i} + \frac{\theta}{r} e^{-r \Delta t_i} \right] \\
+ \sqrt{\frac{\sigma^2 \psi^2 \Delta t_i e^{-2rt}}{2r \kappa (r + \kappa)^2} \left[ r + \kappa - re^{-2r \Delta t_i} \right]} \Delta t_i Z, \] (2.119)

where \( \Delta t_i = t_{i+1} - t_i, \ 0 = t_0 < t - 1 < \cdots < t_n \).

The third step is to perform certain elementary parameter calibrations. In the case of the crude oil price, it is not too difficult to find from market data the current spot price \( \bar{S}_0 \).

We can also estimate the expected long term future spot price from the historical average of spot prices, under the assumption that there exists a supply-demand equilibrium price level, to which the long run price will normally tend. Notice that as \( t \to 0 \) and \( t \to \infty \), we have the weighting factor \( z \to 0 \). Since the expected value of the long run convenience dividend level is equal to the long run mean reverting level \( \theta \), we then have the following relation

\[ \bar{E}[S_\infty] = \bar{E}[S_\infty] = \frac{1}{r} \left[ \kappa \theta \right] \frac{r}{r + \kappa} + \frac{r X_0}{r + \kappa} \bar{E}[X_\infty] = \frac{\theta}{r}. \] (2.120)

Therefore, if we give the long run price \( \bar{E}[S_\infty] \), we can work out \( \theta \). We also know that

\[ S_0 = \frac{1}{r} \left[ \kappa \theta \right] \frac{r}{r + \kappa} + \frac{r X_0}{r + \kappa}. \] (2.121)

Since we know \( S_0 \) (from market data) and \( \theta \), we can work out the expression for \( X_0 \):

\[ X_0 = (r + \kappa) S_0 - \kappa S_\infty. \] (2.122)

Hence the parameters are calibrated according to the relations

\[ \theta = r \bar{S}_\infty \quad \text{and} \quad X_0 = (r + \kappa) S_0 - \kappa \bar{S}_\infty. \] (2.123)

Based on this calibration strategy, we have simulated the price process with a range of parameter values. In particular, in figure (2.2), simulation shows the daily price movement for a year. When \( \sigma = 0 \), the value of \( \psi \), the dividend volatility, increases row-by-row, whereas value of \( \kappa \), the mean reversion rate, increases column-by-column. Observe that the effect of increasing volatility is to increase the range of the price movement, whereas an increase of \( \kappa \) counters this increase. In figure (2.3), simulation shows the daily price movement for a year, for a range of values for \( \sigma \). Observe that the
larger the value of $\sigma$, the earlier and faster the price volatility affects the price. This is realistic because a sudden announcement of new information generally causes large price reactions in the market. The Brent crude oil daily spot price from 4 November 2008 to 20 April 2010 (black path) plotted against five simulated paths from the model is shown in figure (2.4).

Figure 2.2: Nine sets of ten sample paths of the commodity price process, with variable parameters $\psi, \kappa$ and the following parameters: $S_0 = 75, S_\infty = 60, \sigma = 0, r = 0.05, T = 1$. The number of steps is 365. This simulation shows the daily price movement for a year, when $\sigma = 0$.

Figure 2.3: Nine sets of ten sample paths of the commodity price process, with variable parameters $\sigma$ (rate at which information is revealed to market participants) ranging from 0.01 to 1 (value increases from top to bottom and left to right), and the following parameters: $S_0 = 75, S_\infty = 60, \psi = 0.3, \kappa = 0.03, r = 0.05, T = 1$. The number of steps is 365. This simulation shows the daily price movement for a year, for a range of values for $\sigma$. 
Figure 2.4: Brent crude daily spot price from 4 November 2008 to 20 April 2010 (black path) plotted against five simulated paths from the model, with the following parameters: \( S_0 = 62.78 \), \( S_\infty = 60 \), \( \psi = 0.4 \), \( \kappa = 0.05 \), \( r = 0.025 \), \( T = 1 \). The number of steps is 365.

Remark 2.9.1. It is interesting to observe from figure (2.4) that the behaviour of the simulated sample paths is very similar in character to the actual path in the real-world market. Although these simulations have been carried out under constant parameter and interest rate assumptions, we show in Appendix C that a constant parameter OU process, multiplied by a deterministic increasing function of time (in connection with which we have assumed that the commodity price in real term increases over time in general) forms a special case of the time-dependent OU process. This means that the behaviour of the model in the time homogeneous setup can already provide insights regarding the behaviour in the time inhomogeneous setup.
Chapter 3

Commodity Derivatives

3.1 Pricing options on commodity spot instruments

Following from last chapter, where we have introduced a model for storable commodities price, let us now turn the attention to the interesting problem of pricing a commodity derivative in the constant interest rate setup, in the situation where the convenience dividend is modelled by an Ornstein-Uhlenbeck process. The valuation formula for a standard European-style call option with strike $K$ and maturity $T$ is given by:

$$C_0 = e^{-rT}E[(S_T - K)^+] .$$  \hspace{1cm} (3.1)

Recall from equation (2.47) that the commodity price

$$S_t = (1 - z_t) \left[ \frac{\kappa}{r + \kappa} \theta + \frac{r}{r + \kappa} X_t \right] + z_t e^{\kappa t} \xi_t \sigma_t$$

is a linear function of the convenience dividend $X_t$, which in turn is given by

$$X_t = e^{-\kappa t} X_0 + \theta (1 - e^{-\kappa t}) + \psi e^{-\kappa t} \int_0^t e^{\kappa s} d\beta_s .$$  \hspace{1cm} (3.2)

Recall also that, the information process (2.4), under the constant interest rate setup, is given by

$$\xi_t = \sigma t \int_t^\infty e^{-ru} X_u du + B_t .$$  \hspace{1cm} (3.3)

Evidently, \{X_t\}, \{\int_0^\infty e^{-ru} X_u du\}, and \{B_t\} are all Gaussian. Therefore, the price $S_T$ at time $T$ is also Gaussian. Now, returning to our call option valuation formula (3.1),
we thus have
\[ C_0 = e^{-rT} \frac{1}{\sqrt{2\pi \text{Var}[S_T]}} \int_K^\infty (z - K) \exp \left( - \frac{(z - \mathbb{E}[S_T])^2}{2 \text{Var}[S_T]} \right) dz. \] (3.4)

**Proposition 3.1.1.** The European-style commodity call option price \( C_0 \) at time zero is given by
\[ C_0 = e^{-rT} \left[ \sqrt{\frac{\text{Var}[S_T]}{2\pi}} \exp \left( - \frac{(\mathbb{E}[S_T] - K)^2}{2 \text{Var}[S_T]} \right) \right] + e^{-rT} \left[ (\mathbb{E}[S_T] - K) N \left( \frac{\mathbb{E}[S_T] - K}{\sqrt{\text{Var}[S_T]}} \right) \right], \] (3.5)

where \( N(x) \) is the cumulative normal distribution function:
\[ N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left( - \frac{1}{2} z^2 \right) dz. \] (3.6)

The proof of this proposition can be found in Appendix B. The valuation of the call price reduces to the determination of the mean and the variance of the Gaussian random variable \( S_T \). Note that
\[ S_T = (1 - z_T) \frac{1}{r} \left[ \frac{\kappa}{r+\kappa} \theta + \frac{r}{r+\kappa} X_T \right] + z_T e^{rT} \frac{\xi_T}{\sigma T} \]
\[ = \frac{1}{r} \left[ \frac{\kappa}{r+\kappa} \theta + \frac{r}{r+\kappa} X_T \right] \\
- \frac{z_T}{r} \left[ \frac{\kappa}{r+\kappa} \theta + \frac{r}{r+\kappa} X_T \right] + z_T e^{rT} \frac{\xi_T}{\sigma T}, \] (3.7)
\[ X_T = e^{-\kappa T} X_0 + \theta (1 - e^{-\kappa T}) + \psi e^{-\kappa T} \int_0^T e^{\kappa s} d\beta_s, \] (3.8)
and
\[ \mathbb{E}[X_T] = e^{-\kappa T} X_0 + \theta (1 - e^{-\kappa T}). \] (3.9)

Substituting the expression for \( X_T \) of (3.8) into \( S_T \) in equation (3.7), and taking the expectation, we obtain
\[ \mathbb{E}[S_T] = \frac{1}{r} \left[ \frac{\kappa}{r+\kappa} \theta + \frac{r}{r+\kappa} \mathbb{E}[X_T] \right] \\
- \frac{z_T}{r} \left[ \frac{\kappa}{r+\kappa} \theta + \frac{r}{r+\kappa} \mathbb{E}[X_T] \right] + z_T e^{rT} \mathbb{E} \left[ \frac{\xi_T}{\sigma T} \right]. \] (3.10)
Substituting (3.9) in the above, we find

$$
\mathbb{E}[S_T] = \frac{1}{r} \left[ \frac{\kappa}{r + \kappa} \theta + \frac{r}{r + \kappa} \left[ e^{-\kappa T} X_0 + \theta (1 - e^{-\kappa T}) \right] \right] \\
- \frac{zT}{r} \left[ \frac{\kappa}{r + \kappa} \theta + \frac{r}{r + \kappa} \left[ e^{-\kappa T} X_0 + \theta (1 - e^{-\kappa T}) \right] \right] + \frac{zT}{r} \mathrm{e}^{rT} \mathbb{E} \left[ \frac{\xi_T}{\sigma T} \right].
$$

(3.11)

Now we need to work out the expectation of $\xi_T/\sigma T$. From (3.3) we have:

$$
\xi_T \sigma T = \int_T^\infty e^{-ru} X_u du + B_T.
$$

(3.12)

Since $\{B_T\}$ is a Brownian motion, we have $\mathbb{E}[B_T] = 0$, and therefore

$$
\mathbb{E} \left[ \frac{\xi_T}{\sigma T} \right] = \mathbb{E} \left[ \int_T^\infty e^{-ru} X_u du + B_T \right] = \mathbb{E} \left[ \int_T^\infty e^{-ru} X_u du \right].
$$

(3.13)

On account of Fubini’s theorem, we can interchange the expectation and integration to obtain

$$
\mathbb{E} \left[ \frac{\xi_T}{\sigma T} \right] = \int_T^\infty e^{-ru} \mathbb{E}[X_u] du = \int_T^\infty e^{-ru} \left[ e^{-\kappa u} X_0 + \theta (1 - e^{-\kappa u}) \right] du
\\ = \int_T^\infty e^{-u(r + \kappa)} X_0 du + \int_T^\infty \theta (e^{-ru} - e^{-u(r + \kappa)}) du.
$$

(3.14)

Performing the integration, we deduce that

$$
\mathbb{E} \left[ \frac{\xi_T}{\sigma T} \right] = - \frac{X_0}{r + \kappa} \left[ e^{-u(r + \kappa)} \right]_T^\infty - \frac{\theta}{r + \kappa} \left[ e^{-ru} \right]_T^\infty + \frac{\theta}{r + \kappa} \left[ e^{-u(r + \kappa)} \right]_T^\infty
\\ = \frac{X_0}{r + \kappa} e^{-T(r + \kappa)} + \frac{\theta}{r + \kappa} e^{-T} - \frac{\theta}{r + \kappa} e^{-T(r + \kappa)},
$$

(3.15)

and hence that

$$
re^{rT} \mathbb{E} \left[ \frac{\xi_T}{\sigma T} \right] = \frac{\kappa \theta}{r + \kappa} + \frac{r}{r + \kappa} \left[ e^{-\kappa T} X_0 + (1 - e^{-\kappa T}) \theta \right].
$$

(3.16)

From (3.16) and (3.11), we have:

$$
\mathbb{E}[S_T] = \frac{1}{r} \left[ \frac{\kappa}{r + \kappa} \theta + \frac{r}{r + \kappa} \left[ e^{-\kappa T} X_0 + \theta (1 - e^{-\kappa T}) \right] \right].
$$

(3.17)

To calculate the variance of $S_T$, we introduce the following decomposition:

$$
\int_T^\infty P_u X_u du = A_T + \frac{1}{r + \kappa} e^{-rT} X_T.
$$

(3.18)
Here $A_t$ is defined as in (2.24). Under this representation, we can write

$$S_T = (1 - z_T) \left( \frac{1}{r} - \frac{1}{r + \kappa} \right) \theta + \frac{1}{r + \kappa} X_T + z_T e^{rT} A_T + z_T e^{rT} B_T \sigma_T,$$  \hspace{1cm} (3.19)

with $X_T$ as in (3.8). Since the random variables $X_T$, $A_T$ and $B_T$ are independent, from earlier calculation, we know that

$$\text{Var} [X_T] = \frac{\psi^2}{2\kappa} \left( 1 - e^{-2\kappa T} \right),$$  \hspace{1cm} (3.20)

and

$$\text{Var} [A_T] = \psi^2 \int_{T}^{\infty} \left( e^{ks} \int_{s}^{\infty} e^{-u(r+\kappa)} du \right)^2 ds = \psi^2 \int_{T}^{\infty} \left( e^{ks} \left[ \frac{1}{r + \kappa} e^{-s(r+\kappa)} \right] \right)^2 ds = \frac{\psi^2}{2r (r + \kappa)^2} e^{-2rT}.$$  \hspace{1cm} (3.21)

Therefore, we have

$$\text{Var} [S_T] = \frac{1}{(r + \kappa)^2} \text{Var} [X_T] + z_T^2 e^{2rT} \text{Var} [A_T] + z_T^2 e^{2rT} \frac{\text{Var} [B_T]}{\sigma^2 T^2}.$$  \hspace{1cm} (3.22)

Now, we can substitute equations (3.20) and (3.21) into equation (3.22) to obtain:

$$\text{Var} [S_T] = \frac{1}{(r + \kappa)^2} \frac{\psi^2}{2\kappa} \left( 1 - e^{-2\kappa T} \right) + z_T^2 e^{2rT} \frac{\psi^2}{2r (r + \kappa)^2} e^{-2rT} + z_T^2 e^{2rT} \frac{T}{\sigma^2 T^2}.$$  \hspace{1cm} (3.23)

Therefore, using the expressions for the mean and the variance, we find that the European commodity call option price under the time-homogeneous model setup is given by

$$C_0 = e^{-rT} \left[ \sqrt{\frac{\gamma^2 T}{2\pi}} \exp \left( -\frac{(\mu_T - K)^2}{2\gamma^2 T} \right) + (\mu_T - K) N \left( \frac{\mu_T - K}{\gamma_T} \right) \right],$$  \hspace{1cm} (3.24)

where

$$\mu_T = \frac{1}{r} \left\{ \frac{\kappa}{r + \kappa} \theta + \frac{r}{r + \kappa} \left[ e^{-\kappa T} X_0 + \theta (1 - e^{-\kappa T}) \right] \right\},$$  \hspace{1cm} (3.25)

and

$$\gamma^2_T = \frac{\psi^2}{2\kappa (r + \kappa)^2} \left( 1 - e^{-2\kappa T} \right) + z_T^2 \left[ \frac{\psi^2}{2r (r + \kappa)^2} + \frac{e^{2rT}}{\sigma^2 T^2} \right].$$  \hspace{1cm} (3.26)
In figure (3.1) we plot the call price as a function of the underlying price $S_0$ and the option maturity $T$.

### 3.2 Option price analysis

It is interesting to observe that in the present model, a closed-form expression for the commodity option delta can be obtained. Recall that the option delta $\Delta$ measures the rate of change of the option value with respect to changes in the price of the underlying asset. To work out the option delta at time zero we recall that

$$S_0 = \frac{1}{r} \left[ \frac{\kappa \theta}{r + \kappa} + \frac{rX_0}{r + \kappa} \right]. \tag{3.27}$$

Therefore, we can write the option delta in the form

$$\Delta_0 = \frac{\partial C_0}{\partial S_0} = \frac{\partial C_0}{\partial X_0} \frac{\partial X_0}{\partial S_0} = (r + \kappa) \frac{\partial C_0}{\partial X_0} = (r + \kappa) \frac{\partial C_0}{\partial \mu} \frac{\partial \mu}{\partial X_0}, \tag{3.28}$$

where $\mu = \mu_T$ is defined in (3.25). Evidently, we have

$$\frac{\partial \mu}{\partial X_0} = \frac{e^{-rT}}{r + \kappa}. \tag{3.29}$$
Figure 3.2: The commodity call option delta as a function of the initial asset price in the OU model. The parameters are set as follows: $\kappa = 0.15$, $\theta$ ranges from 0.3 to 0.8 with increments of 0.01, $\sigma = 0.25$, $X_0 = 0.6$, $\psi = 0.15$, $r = 0.05$, and $K = 10$. The three maturities are $T = 0.5$ (blue), $T = 1.0$ (green), and $T = 3.0$ (red).

and we also have

$$
\frac{\partial C_0}{\partial \mu} = e^{-rT} \left\{ -\frac{2(\mu - K)}{2\gamma^2} \frac{\gamma}{\sqrt{2\pi}} \exp \left\{ -\frac{(\mu - K)^2}{2\gamma^2} \right\} + N \left( \frac{\mu - K}{\gamma} \right) + \frac{\mu - K}{\gamma} \phi \left( \frac{\mu - K}{\gamma} \right) \right\}.
$$

(3.30)

Here $\phi(x)$ denotes the standard normal density function. Substituting this back to (3.28), we find that the call option delta formula reads as follows:

$$
\Delta_0 = e^{-2rT} \left\{ -\frac{2(\mu - K)}{2\gamma^2} \frac{\gamma}{\sqrt{2\pi}} \exp \left\{ -\frac{(\mu - K)^2}{2\gamma^2} \right\} + N \left( \frac{\mu - K}{\gamma} \right) + \frac{\mu - K}{\gamma} \phi \left( \frac{\mu - K}{\gamma} \right) \right\}.
$$

(3.31)

In figure (3.2) we plot the option delta as a function of the initial asset price for three different values of option maturity, indicating the characteristic behaviour of the delta seen, for example, in the classical Black-Scholes model.

3.3 Futures contracts and derivatives on futures

3.3.1 Futures prices

In the foregoing material we derived the price process that we used as our model for crude oil, and proceeded to value elementary option contracts based on the model. We found, in particular, that our model is fully tractable, on account of its Gaussianity; that
said, the derivation of the associated innovations representation shows that the model is nevertheless in many respects highly nontrivial. In particular, the simulation study indicates the intricate and rich nature of the model. Now in the real crude oil market, assets that are more commonly traded are futures contracts. Hence in order to obtain a form of the model that is more directly applicable it will be useful to extend the analysis to consider futures contracts, and this is the objective of the present section.

Since we are assuming here a constant interest rate, and since any potential credit risky event is assumed negligible, the futures price \( f^T_t \) at time \( t \) that matures at time \( T \), \( 0 \leq t \leq T \), is the same as the forward price with the same maturity. We can write the futures price as the expectation of the future spot price under the risk neutral measure:

\[
 f^T_t = \mathbb{E}^Q [S_T | \mathcal{F}_t].
\]  
(3.32)

This follows as a consequence of the usual system of margin calls, and from the fact that at maturity \( T \), the futures price should be equal to the spot price. For time \( t = 0 \) (today), we thus have

\[
 f^T_0 = \mathbb{E}^Q [S_T].
\]  
(3.33)

Making use of the expression (3.19) for \( S_T \) we thus find, for the initial futures price, that

\[
 f^T_0 = \frac{(r + \kappa)\theta + re^{-\kappa T}(X_0 - \theta)}{r(r + \kappa)}. 
\]  
(3.34)

### 3.3.2 European options on futures prices

To obtain the futures price process we have to work out

\[
 f^T_t = \mathbb{E}_t [S_T] = \frac{\kappa \theta + rE_t [X_T]}{r(r + \kappa)} + \frac{e^{r T} z_T}{\sigma T} \mathbb{E}_t [\omega_T], 
\]  
(3.35)

where

\[
 S_t = \frac{\kappa \theta + r X_t}{r(r + \kappa)} + e^{r t} z_t \frac{\omega_t}{\sigma t} 
\]  
(3.36)

and

\[
 \omega_t = \frac{\sigma \psi t}{r + \kappa} \int_t^\infty e^{-ru} d\beta_u + B_t. 
\]  
(3.37)
Hence, we have to focus on working out the following conditional expectations:

\[ E_t [X_T] = e^{-\kappa(T-t)} X_t + \theta(1 - e^{-\kappa(T-t)}) + \psi e^{-\kappa T} \mathbb{E}_t \left[ \int_t^T e^{\kappa u} d\beta_u \right], \quad (3.38) \]

and

\[ \mathbb{E}_t [\omega_T] = \frac{\sigma \psi T}{r + \kappa} \mathbb{E}_t \left[ \int_T^\infty e^{-ru} d\beta_u \right] + \mathbb{E}_t [B_T]. \quad (3.39) \]

The question thus reduces to solving the three conditional expectations appearing in equation (3.38) and (3.39). Let us calculate these one by one. Consider first, for \( 0 \leq t \leq T < \infty \), the expectation

\[ \mathbb{E} [B_T \mid \omega_t] = \mathbb{E} \left[ \mathbb{E} [B_T \mid \omega_t, B_t] \mid \omega_t \right] = \mathbb{E} [B_t \mid \omega_t]. \quad (3.40) \]

Here we have made use of the Markovian property and the tower property of conditional expectation. We have also made use of the facts that \{\beta_t\} and \{B_t\} are independent, and hence that \( [B_T \mid \omega_t, B_t] = [B_T \mid B_t] \).

If we now substitute (3.37) back into the above and use the linearity of expectation, we obtain

\[ \mathbb{E} [B_T \mid \omega_t] = \mathbb{E} \left[ \omega_t - \frac{\sigma \psi t}{r + \kappa} \int_t^\infty e^{-ru} d\beta_u \mid \omega_t \right] = \omega_t - \frac{\sigma \psi t}{r + \kappa} \mathbb{E} \left[ \int_t^\infty e^{-ru} d\beta_u \mid \omega_t \right]. \quad (3.41) \]

The conditional expectation appearing here has in fact been obtained earlier in (2.68), and hence we find

\[ \mathbb{E} [B_T \mid \omega_t] = \omega_t - \frac{\sigma \psi t}{r + \kappa} z_t \omega_t = \omega_t (1 - z_t). \quad (3.42) \]

Next we consider

\[ \mathbb{E} \left[ \int_T^\infty e^{-ru} d\beta_u \mid \omega_t \right] = \mathbb{E} \left[ \int_T^\infty e^{-ru} d\beta_u - \int_t^T e^{-ru} d\beta_u \mid \omega_t \right]. \quad (3.43) \]

We can utilise the expression (2.68) to obtain

\[ \mathbb{E} \left[ \int_T^\infty e^{-ru} d\beta_u \mid \omega_t \right] = \frac{r + \kappa}{\sigma \psi t} z_t \omega_t - \mathbb{E} \left[ \int_t^T e^{-ru} d\beta_u \mid \omega_t \right]. \quad (3.44) \]
Now the problem reduces to finding the conditional expectation above. Consider, for $0 \leq t \leq T < \infty$, the expectation

$$
\mathbb{E}\left[ \int_t^T e^{-ru}d\beta_u \middle| \omega_t \right] = \mathbb{E}\left[ \mathbb{E}\left[ \int_t^T e^{-ru}d\beta_u \middle| \omega_t, \{\beta_u\}_0 \leq u \leq t \right] \middle| \omega_t \right],
$$

where we have made use of the tower property again. Since $\{\beta_t\}$ is a Brownian motion, one sees that the stochastic integral is a martingale under its own filtration. Hence,

$$
\mathbb{E}\left[ \int_t^T e^{-ru}d\beta_u \middle| \omega_t \right] = \mathbb{E}\left[ \int_0^T e^{-ru}d\beta_u - \int_0^t e^{-ru}d\beta_u \middle| \omega_t \right] = 0.
$$

Therefore, we have

$$
\mathbb{E}\left[ \int_T^\infty e^{-ru}d\beta_u \middle| \omega_t \right] = \frac{r + \kappa}{\sigma \psi t} z_t \omega_t,
$$

and

$$
\mathbb{E}\left[ \int_t^T e^{\kappa u}d\beta_u \middle| \omega_t \right] = 0.
$$

By substituting these results in (3.38) and (3.39), we obtain

$$
\mathbb{E}_t [X_T] = e^{-\kappa(T-t)}X_t + \theta \left[ 1 - e^{-\kappa(T-t)} \right],
$$

and

$$
\mathbb{E}_t [\omega_T] = \frac{\sigma \psi T}{r + \kappa} \frac{r + \kappa}{\sigma \psi t} z_t \omega_t + \omega_t (1 - z_t)
\quad = z_t \omega_t \frac{T - t}{t} + \omega_t.
$$

Putting these together, we have our formula for the futures price at time $t$ for the contract that matures at $T$:

$$
f_t^T = \frac{(r + \kappa)\theta + r(X_t - \theta)e^{-\kappa(T-t)}}{r(r + \kappa)} + \frac{e^{rT}z_T [z_t T + (1 - z_t)T]}{\sigma^2 T} \omega_t.
$$
We can check the consistency of this result in two ways. Given that the futures price is a martingale under $Q$, if we take unconditional expectation of $f^T_t$, we have

$$E \left[ f^T_t \right] = \frac{(r + \kappa)\theta + r(\mathbb{E}[X_t] - \theta)e^{-\kappa(T-t)}}{r(r + \kappa)} + \frac{e^{rT}z_T[z_tT + (1 - z_t)t]}{\sigma^2T^2}E[\omega_t]$$

$$= \frac{(r + \kappa)\theta + r(e^{-\kappa T}X_0 + \theta - \theta e^{-\kappa t} - \theta)e^{-\kappa(T-t)}}{r(r + \kappa)}$$

$$= \frac{(r + \kappa)\theta + re^{-\kappa T}(X_0 - \theta)}{r(r + \kappa)}$$

$$= f^T_0.$$  \hspace{1cm} (3.51)

On the other hand, we must have $f^T_{t|t=0} = f^T_0$, which can be checked and it is true.

Let us now calculate the variance of $f^T_t$. Observe that futures price is linear in $X_t$ and $\omega_t$, and that these two random variables are independent. Therefore, the variance becomes:

$$\text{Var} \left[ f^T_t \right] = \frac{e^{-2\kappa(T-t)}}{(r + \kappa)^2} \text{Var}[X_t] + \frac{e^{2rT}Z^2_T[z_tT + (1 - z_t)t]^2}{\sigma^2t^2T^2} \text{Var}[\omega_t].$$  \hspace{1cm} (3.52)

Recall that we already worked out the variances of these two variables:

$$\text{Var}[X_t] = \frac{\psi^2}{2\kappa}(1 - e^{-2\kappa t})$$  \hspace{1cm} (3.53)

and

$$\text{Var}[\omega_t] = \frac{\sigma^2\psi^2t^2}{2r(r + \kappa)^2e^{2rt}} + \frac{t}{1 - z_t}.$$  \hspace{1cm} (3.54)

It follows that

$$\text{Var} \left[ f^T_t \right] = \frac{\psi^2 \left[ e^{-2\kappa(T-t)} - e^{-2\kappa T} \right]}{2\kappa(r + \kappa)^2} + \frac{e^{2rT}[z_tz_T + (1 - z_t)t]^2}{\sigma^2T^2(1 - z_t)t}.$$  \hspace{1cm} (3.55)

We shall now make use of this expression to price options on futures, since options written on futures contracts are common commodity derivatives traded in the commodity markets. The valuation for this derivative today (at time 0) under $Q$ is:

$$C^T_{0,t} = e^{-rt}E \left[ (f^T_t - K)^+ \right]$$  \hspace{1cm} (3.56)

for time $0 \leq t < T$, where $t$ is the maturity of the option and $T$ is the maturity of the futures contract. Note that the option must mature before its underlying matures. The option strike is $K$. The fact that futures price $f^T_t$ is a linear combination of two independent normally distributed random variables means that futures price itself is a normally distributed random variable. Therefore, we have the following result:
Proposition 3.3.1. The price at time 0 of a t-maturity European-style call option on a T-maturity futures contract can be expressed as follows:

\[
C_{0t}^T = e^{-rt} \left[ \sqrt{\frac{\text{Var} \left[ f_t^T \right]}{2\pi}} \exp \left( -\frac{\left( \text{E} \left[ f_t^T \right] - K \right)^2}{2\text{Var} \left[ f_t^T \right]} \right) \right] + e^{-rt} \left[ \left( \text{E} \left[ f_t^T \right] - K \right) N \left( \frac{\text{E} \left[ f_t^T \right] - K}{\sqrt{\text{Var} \left[ f_t^T \right]}} \right) \right],
\]

(3.57)

where

\[
\text{E} \left[ f_t^T \right] = \frac{(r + \kappa)\theta + re^{-\kappa T}(X_0 - \theta)}{r(r + \kappa)}
\]

(3.58)

and

\[
\text{Var} \left[ f_t^T \right] = \frac{\psi^2 \left[ e^{-2\kappa(T-t)} - e^{-2\kappa T} \right]}{2\kappa(r + \kappa)^2} + \frac{e^{2rT} \left[ z_t z_T T + (1 - z_t) z_T t \right]^2}{\sigma^2 T^2 (1 - z_t) t}.
\]

(3.59)

This result can be checked by essentially the same method as we have used in the case of options in spot markets.

3.4 Time-inhomogeneous extensions

3.4.1 Modelling the convenience dividend

We have seen that in the information-based approach we have been able to derive a reasonable model for the price processes of storable commodities, starting from a model for the convenience dividend based on a time-homogeneous Ornstein-Uhlenbeck process. In particular, we have been able to price commodity derivatives in our modelling framework. The time-homogeneous case, however, is nevertheless somewhat restrictive as a candidate model for the convenience dividend in many situations. Therefore, in what follows we shall extend the model introduced above to more realistic time inhomogeneous case. This is similar in spirit to extending the time-independent Vasicek model of interest-rate dynamics to the time inhomogeneous case knows as the Hull-White model. We shall indicate below that indeed such an extension is feasible in the present context, and we are able to retain analytical tractability of the model (at the expense, of course, of some added complexity).

In the time-inhomogeneous case, our model for the convenience dividend continues to fulfill the mean-reverting dynamical equation of the form (2.6), except that the parameters
of the model are replaced with time-dependent functions. Then we write

\[ dX_t = \kappa_t (\theta_t - X_t) \, dt + \psi_t d\beta_t, \tag{3.60} \]

where \( \{ \beta_t \} \) is again a standard \( \mathcal{Q} \)-Brownian motion, and where \( \{ \kappa_t \}, \{ \theta_t \}, \) and \( \{ \psi_t \} \) are deterministic function of time. The solution to this stochastic equation with initial condition \( X_0 \) is

\[ X_t = e^{-f_t} \left[ X_0 + \int_0^t e^{f_s} \kappa_s \theta_s \, ds + \int_0^t \psi_s e^{f_s} \, d\beta_s \right], \tag{3.61} \]

where we have defined the function \( \{ f_t \} \) by the relation

\[ f_t = \int_0^t \kappa_s \, ds. \tag{3.62} \]

It will be useful to have at hand appropriate expressions for the mean, variance, and covariance of the convenience dividend. These are given by:

\[ \mathbb{E}[X_t] = e^{-f_t} \left( X_0 + \int_0^t e^{f_s} \kappa_s \theta_s \, ds \right) \tag{3.63} \]

for the mean,

\[ \text{Var}[X_t] = e^{-2f_t} \int_0^t e^{2f_s} \psi_s^2 \, ds \tag{3.64} \]

for the variance, and

\[ \text{Cov}[X_t, X_T] = e^{-f_t - f_T} \int_0^t e^{2f_s} \psi_s^2 \, ds \tag{3.65} \]

for the covariance.

\section{3.5 Properties and applications of time-inhomogeneous OU process}

For the purpose of deriving the asset-price process it will be useful to work out some further properties of the time-inhomogeneous Ornstein-Uhlenbeck process. In particular, the construction of the time-inhomogeneous OU bridge will be indispensable. We proceed as follows.
3.5.1 Reinitialisation property and orthogonal decomposition

We recall the reinitialisation property detailed in section 2.4.1. We can use a very similar substitution method to derive a reinitialisation property in the time-inhomogeneous case:

\[
X_T = e^{-\kappa T} \left[ X_t + e^{-f_1} \int_t^T e^{f_s \kappa_s \theta_s} d\beta_s + e^{-f_1} \int_t^T e^{f_s \psi_s} d\beta_s \right].
\]  
(3.66)

Here, for convenience of notation we have defined \( \{ \kappa_T \} \) by the relation

\[
e^{-\kappa_T} = e^{-\int_t^T \kappa_s d\beta_s}.
\]  
(3.67)

On account of the relation (3.66) we have:

**Proposition 3.5.1.** The random variables \( X_t \) and \( X_T - e^{-\kappa_T} X_t \) are independent.

**Proof.** We can verify this by the use of variance-covariance relations. In particular, from (3.61) and (3.66) we have

\[
\text{Cov} \left[ X_t, X_T - e^{-\kappa_T} X_t \right] = \mathbb{E} \left[ (X_t - \mathbb{E}[X_t]) \left( X_T - e^{-\kappa_T} X_t - \mathbb{E} \left[ X_T - e^{-\kappa_T} X_t \right] \right) \right]
\]

\[
= \mathbb{E} \left[ e^{-2f_1 - \kappa_T} \int_t^T e^{f_s \psi_s} d\beta_s \int_0^t e^{f_u \psi_u} d\beta_u \right]
\]

\[
= e^{-2f_1 - \kappa_T} \mathbb{E} \left[ \int_t^T e^{f_s \psi_s} d\beta_s \int_0^t e^{f_u \psi_u} d\beta_u \right]
\]

\[
= 0.
\]

The last term follows since the two stochastic integrals inside the expectation are independent. Since \( X_t \) and \( X_T - e^{-\kappa_T} X_t \) are Gaussian, this establishes the claim. \( \square \)

It follows that we can express the random variable \( X_T \) by means of an orthogonal decomposition of the form

\[
X_T = \left( X_T - e^{-\kappa_T} X_t \right) + e^{-\kappa_T} X_t,
\]  
(3.68)

for \( T > t \).
3.5.2 Time-inhomogeneous OU bridge

As in the time-homogeneous case, there is another interesting but rather less obvious orthogonal decomposition admitted by $X_t$, which plays a crucial role in the time-inhomogeneous setup. This decomposition is given by:

$$ X_t = \left( X_t - \frac{e^{-f t} \int_0^t e^{2f_s \psi_s^2} ds}{e^{-f T} \int_0^T e^{2f_s \psi_s^2} ds} X_T \right) + \frac{e^{-f t} \int_0^t e^{2f_s \psi_s^2} ds}{e^{-f T} \int_0^T e^{2f_s \psi_s^2} ds} X_T. \quad (3.69) $$

The process $\{b_{tT}\}_{0 \leq t \leq T}$ defined for fixed $T$ by

$$ b_{tT} = X_t - \frac{e^{-f t} \int_0^t e^{2f_s \psi_s^2} ds}{e^{-f T} \int_0^T e^{2f_s \psi_s^2} ds} X_T \quad (3.70) $$

defines a time inhomogeneous Ornstein-Uhlenbeck bridge. Clearly, we have

$$ b_{0T} = X_0 \quad \text{and} \quad b_{TT} = 0. \quad (3.71) $$

We can work out the mean and variance of the time inhomogeneous OU bridge. Recall the expectation of $X_t$ from (3.63):

$$ \mathbb{E}[X_T] = e^{-fT} \left( X_0 + \int_0^T e^{\int_s^T \kappa_s \theta_s ds} \right). \quad (3.72) $$

Hence, the mean of $b_{tT}$ is given by:

$$ \mathbb{E}[b_{tT}] = \frac{e^{-f t}}{\int_0^T e^{2f_s \psi_s^2} ds} \left[ \left( X_0 + \int_0^t e^{\int_s^t \kappa_s \theta_s ds} \right) \left( \int_t^T e^{2f_s \psi_s^2} ds \right) \right] - \frac{e^{-f t}}{\int_0^T e^{2f_s \psi_s^2} ds} \left[ \left( \int_0^t e^{2f_s \psi_s^2} ds \right) \left( \int_t^T e^{\int_s^T \kappa_s \theta_s ds} \right) \right]. \quad (3.73) $$

On the other hand, recall the variance of $X_t$ from (3.64):

$$ \text{Var}[X_T] = e^{-2fT} \int_0^T e^{2f_s \psi_s^2} ds. \quad (3.74) $$

The variance of $b_{tT}$ is then given by:

$$ \text{Var}[b_{tT}] = \frac{e^{-2f t} \int_0^t e^{2f_s \psi_s^2} ds \int_t^T e^{2f_s \psi_s^2} ds}{\int_0^T e^{2f_s \psi_s^2} ds}. \quad (3.75) $$
3.6 Commodity prices in a time-dependent setting

The arguments presented in the material of the previous chapter carry through in the case of an extended time-inhomogeneous OU model for the convenience dividend. As in the previous setup, we shall assume that the market filtration is generated jointly by the following processes:

(a) the convenience dividend process \( \{X_t\}_{t \geq 0} \); and

(b) an “information process” \( \{\xi_t\}_{t \geq 0} \) of the form:

\[
\xi_t = \sigma t \int_t^\infty P_u X_u \, du + B_t.
\] (3.76)

Here the \( \mathbb{Q} \)-Brownian motion \( \{B_t\} \) is independent of \( \{X_t\} \). In this case, the time inhomogeneity of the information process comes solely from the time inhomogeneous convenience dividend process. Thus the market filtration is generated as follows:

\[
\mathcal{F}_t = \sigma \left( \{X_s\}_{0 \leq s \leq t}, \{\xi_s\}_{0 \leq s \leq t} \right).
\] (3.77)

One can interpret the generators of the filtration in the following spirit:

- The knowledge of the convenience dividend provides information about the current and past dividend levels.
- The information process gives partial information about the future dividend flows.

Thus, we can write the valuation formula for storable commodity in the form:

\[
S_t = \frac{1}{P_t} \mathbb{E} \left[ \int_t^\infty P_u X_u \, du \bigg| \{\xi_s\}_{0 \leq s \leq t}, \{X_s\}_{0 \leq s \leq t} \right].
\] (3.78)

It is remarkable that the methodology employed in the analysis of the previous chapter for deriving the asset price and the prices of associated derivatives carries through in an essentially identical manner to the time inhomogeneous case. Since a rather detailed derivation has already been provided in the homogeneous case, here we shall not repeat all the steps. We merely state the main results.

The first step that is useful for the calculation is the joint Markovian property, which allows us to write

\[
\mathbb{E} \left[ \int_t^\infty P_u X_u \, du \bigg| \{\xi_s\}_{0 \leq s \leq t}, \{X_s\}_{0 \leq s \leq t} \right] = \mathbb{E} \left[ \int_t^\infty P_u X_u \, du \bigg| \xi_t, X_t \right].
\] (3.79)
We note, incidentally, that the Markov property breaks down if the information flow rate parameter $\sigma$ in (3.76) is replaced with a time-dependent function. However, since we are only making the parameters appearing in $\{X_t\}$ time dependent here, the Markov condition (3.79) allows us to use the orthogonal decomposition (3.68) to isolate the dependence of the commodity price on the current level $X_t$ of the convenience dividend. Following the same steps as in the time homogeneous case, we are able to derive the time-inhomogeneous commodity price formula:

$$S_t = \frac{1 - z_t}{P_t} \left[ \int_t^\infty e^{f_s \kappa_s \theta_s \delta_s ds} + e^{f_t \delta_t X_t} \right] + \frac{z_t \xi_t}{\sigma \tau P_t},$$

where

$$z_t = \frac{\sigma^2 t \int_t^\infty e^{2f_s \psi_s^2 \delta_s^2 ds}}{1 + \sigma^2 t \int_t^\infty e^{2f_s \psi_s^2 \delta_s^2 ds}},$$

is the corresponding weighting factor.

### 3.7 Commodity derivatives in a time-dependent setting

Let us consider the problem of pricing a commodity derivative in the time-inhomogeneous case. As an example, consider the pricing of a European-style call option. The value of a call option with strike $K$ and maturity $T$ is:

$$C_0 = P_T \mathbb{E} \left[ (S_T - K)^+ \right],$$

where $P_T$ is the discount factor:

$$P_t = \exp \left( - \int_0^t r_u du \right),$$

and $\{r_t\}$ is the short rate. We recall Proposition 3.1.1 where a closed-form expression for the initial price $C_0$ of a call option was obtained in the time-homogeneous situation. An analogous result holds in the time-inhomogeneous setup, on account of the Gaussianity of the problem at hand. In particular, in the present context a calculation shows that the call price is given by the expression

$$C_0 = P_T \left[ \sqrt{\frac{\gamma_T^2}{2\pi}} \exp \left( - \frac{(\mu_T - K)^2}{2\gamma_T^2} \right) + (\mu_T - K) N \left( \frac{\mu_T - K}{\gamma_T} \right) \right],$$

where

$$\mu_T = \frac{1}{P_T} \int_T^\infty e^{f_s \kappa_s \theta_s \delta_s ds} + \frac{\delta_T}{P_T} \left( X_0 + \int_0^T e^{f_s \kappa_s \theta_s ds} \right),$$

and

$$\gamma_T = \frac{1}{P_T} \int_T^\infty e^{f_s \kappa_s \theta_s \delta_s ds}.$$
and
\[
\gamma^2_T = \frac{2T}{P_T} \left( \int_T^\infty e^{2fs} \psi^2 \sigma^2 ds + \frac{1}{\sigma^2T} \right) + \frac{\delta_T^2}{P_T} \int_0^T e^{2fs} \psi^2 ds. \tag{3.85}
\]

It is satisfying that a fully tractable and explicit expression for the asset price, as well as option prices, can be obtained in the time-inhomogeneous case. This is useful because the range of applicability of the model is much extended as a consequence.

In summary, we have been able to derive models for prices of commodities, based on more primitive concepts of convenience dividends and market information concerning future benefits. The models thus obtained are fully tractable on account of the Gaussian nature of the resulting price process. Nevertheless, the derivation of the price process, and that of the associated innovations representation, are quite intricate, involving novel features of Ornstein-Uhlenbeck bridges.

It is worth remarking that we are only looking at the area of storable commodities, given the fact that the notion of convenience dividend only arises from physical storage of the underlying commodities. Storable commodities include: crude oil, gold, metals, wheat, orange juice, frozen pork bellies, and many other commodities that are actively traded on exchanges. In the case of non-storable commodities, such as electricity, the arguments of convenience dividend and cash and carry break down since it is impossible to carry such commodities over a significant time period. The spot-forward relationship breaks down in the case of non-storable commodities. One can argue that hydro-electricity (electricity generated by releasing water from dam and harvesting its potential energy when it passes through generators) is a form of storable commodity since one can store water in a reservoir. But hydro-electricity only accounts for a very small amount of the global electricity network. Batteries constitute another way in principle for storing electricity, but even with modern technology batteries remain a rather inefficient and expensive way of storing electricity, and in practice should be viewed as a distinct commodity. In any case, the issues associated with the relation between electricity in its conventional out-of-the-socket form, and its various semi-stored forms (hydro-electricity, batteries, solar panels, and so forth) is rather complicated. The discussion of non-storable commodities is thus very interesting in its own rights and it attracts a growing amount of attention, but it is clearly outside the scope of the present work.
Part II

Theory of Signal Processing with Lévy Noise
Chapter 4

Introduction

4.1 Motivation

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 simple 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 given by $\xi = X + \epsilon$, where $\epsilon$ is another random variable, which is independent of $X$. Wiener shows how, given the observed value of the noisy message $\xi$, the original distribution of $X$ can be transformed into an improved *a posteriori* distribution that has a higher information content than the original distribution (in the sense of lower entropy). 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). In connection with World War II, Wiener undertook to analyse the problem of improving the success of anti-aircraft fire. An anti-aircraft gunner must shoot ahead of where his target is at the time of firing. The amount and direction ahead must be estimated quickly and accurately. Where to aim is based on knowledge of how the plane has been travelling and where it is likely to travel in the time the shell takes to reach the plane even if the pilot takes evasive action. Following the earlier work of Wiener, a subsequent breakthrough for the theory of filtering 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*
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 by 1994 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 noisy 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. The filtering theory involving Gaussian noise has been covered by various authors. Liptser & Shiryaev (2000), for example, covers nonlinear filtering (prediction and smoothing) theory and its applications to the problems of optimal estimation and control with incomplete data; Xiong (2008) focuses on the use of probabilistic tools to estimate the unobservable stochastic processes that arise in many applied fields including communication, target-tracking, and mathematical finance; and Bain & Crisan (2010) provides a modern mathematical treatment of the nonlinear stochastic filtering problem, also focusing on various numerical methods for the solutions of filtering problems.

There is no reason a priori, however, why an information process (signal and noise) should be “additive” in the sense indicated above, 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 in situations where the noise is of a discontinuous nature.

There has been a good deal of recent research carried out on the problem of filtering noisy information containing jumps. For example, Rutkowski (1994) discusses linear filtering, smoothing, and prediction problems for a discrete-time linear model using α-stable process. Ahn & Feldman (1999) focus on solving the problems in engineering applications that require extracting a signal from observations corrupted by additive noise, possibly heavy-tailed. They assume that the observation noise is a Lévy process, while the signal is Gaussian, and derive a nonlinear recursive filter that minimises the mean-square error. A suboptimal filter is also proposed for numerical purposes, and simulations show that it outperforms the existing linear filter. Meyer-Brandis & Proske (2004) consider a nonlinear filtering problem with mixed observations, modelled by a Brownian motion and a generalised Cox process, whose jump intensity is given in terms of a Lévy measure. A Zakai equation for the unnormalised conditional density is solved.
to provide an explicit solution to the filtering problem. As an application, they propose a jump diffusion model adapted to their framework for financial assets, which captures the phenomenon of time inhomogeneity of the jump size density. Hanzon & Ober (2006) proposed an approach to calculate rational density functions using state-space representations of the densities, with applications to filtering problems. The case of Cauchy noise is treated as an illustrative example. Poklukar (2006) develops a treatment of the nonlinear filtering problem for jump-diffusion processes. The optimal filter in continuous time is derived for a stochastic system by using measure transformation, where the dynamics of the signal variable is described by a jump-diffusion equation. The optimal filter is then described by stochastic integral equations. Popa & Sritharan (2009) investigate the $n$-dimensional nonlinear filtering problem for jump-diffusion processes. The optimal filter is derived for the case when the observations are continuous. More recently, Mandrekar et al. (2011) derived a Bayes-type formula for the nonlinear filter where the observation contains both general Gaussian noise as well as Cox noise whose jump intensity depends on the signal. It extends the well-known Kallianpur-Striebel formula in the classical nonlinear filter setting. Zakai type equations for both the unnormalised conditional distribution as well as the unnormalised conditional density in the case the signal is a Markovian jump diffusion are obtained.

We see from the remarks above there has been ample research into aspects of filtering theory involving non-Gaussian noise. However, these works have 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.

4.2 Synopsis of the theory of Lévy information

The purpose of this second part of the thesis, which is based in part on Brody, Hughston & Yang (2013a), 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.

More specifically, 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 are set out more fully in the remainder of this thesis, can be presented as follows. We recall the idea of the Esscher transform as a change of probability measure on a probability space \((\Omega, \mathcal{F}, P_0)\) that supports a Lévy process \(\{\xi_t\}_{t \geq 0}\) that possesses \(P_0\) exponential moments. The space of admissible moments is the set

\[
A = \{ w \in \mathbb{R} : \mathbb{E}^{P_0}[\exp(w \xi_t)] < \infty \}. \tag{4.1}
\]

The associated Lévy exponent

\[
\psi(\alpha) = \frac{1}{t} \ln \mathbb{E}^{P_0}[\exp(\alpha \xi_t)] \tag{4.2}
\]

then exists for all \(\alpha \in A_C := \{ w \in \mathbb{C} : \text{Re } w \in A \}, \) and is independent of the value of \(t\). A parametric family of measure changes \(P_0 \rightarrow P_\lambda\) commonly called Esscher transformations can be constructed by use of the exponential martingale family \(\{\rho_\lambda^t\}_{t \geq 0}\) defined for each \(\lambda \in A\) by

\[
\rho_\lambda^t = \exp (\lambda \xi_t - \psi(\lambda)t). \tag{4.3}
\]

If \(\{\xi_t\}\) is a \(P_0\)-Brownian motion, then \(\{\xi_t\}\) is a \(P_\lambda\)-Brownian motion with drift \(\lambda\); if \(\{\xi_t\}\) is a \(P_0\)-Poisson process with intensity \(m\), then \(\{\xi_t\}\) is \(P_\lambda\)-Poisson with intensity \(e^\lambda m\); if \(\{\xi_t\}\) is a \(P_0\)-gamma process with rate parameter \(m\) and scale parameter \(\kappa\), then \(\{\xi_t\}\) is \(P_\lambda\)-gamma with rate parameter \(m\) and scale parameter \(\kappa/(1-\lambda)\); and so on: each case is different in character.

A natural generalisation of the Esscher transformation results when the parameter \(\lambda\) in the measure change is replaced with a random variable \(X\). From the perspective of the new measure \(P_X\) the process \(\{\xi_t\}\) retains the “noisy” character of its \(P_0\)-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 Lévy process under \( \mathbb{P} \), given \( X \), with Lévy exponent \( \psi_0(\alpha + X) - \psi_0(X) \) for \( \alpha \in \mathbb{C}^I := \{ w \in \mathbb{C} : \Re w = 0 \} \). We are thus able to classify Lévy information processes according to 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 that noise type.

We consider a number of different noise types, and construct explicit representations of the associated information processes. We are also able to 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 Chapter 5 and Chapter 6 of the thesis is as follows. Chapter 5 will develop the general theory of Lévy information, while Chapter 6 will be devoted to working out a number of nontrivial examples.

Specifically, in Section 5.2, after recalling some facts about processes with stationary and independent increments, we define Lévy information, and in Proposition 5.3.1 we show that the signal carried by a Lévy information process is effectively “revealed” after the passage of sufficient time. In Section 5.4 we present (in Proposition 5.4.1) an explicit construction using a change of measure technique that ensures the existence of Lévy information processes, and in Proposition 5.4.2 we prove a converse to the effect that any Lévy information process can be obtained in this way. In Proposition 5.5.1 we construct the optimal filter for general Lévy information processes, and in Proposition 5.5.2 we show that such processes have the Markov property. In Proposition 5.6.1 we establish a result that indicates how the information content of the signal is coded into the structure of an information process. Then in Proposition 5.7.1 we present a general construction of the so-called innovations process associated with Lévy information.

Finally in Chapter 6 we proceed to examine a number of specific examples of Lévy information processes, for which explicit representations are constructed in Propositions 6.4.1–6.7.1.
Chapter 5

The Theory of Signal Processing with Lévy Information

5.1 Overview of Lévy processes

Before we proceed with the analysis of the theory of Lévy information, let us begin by a brief overview of some of the properties of Lévy processes. For simplicity of exposition, one-dimensional Lévy processes are treated here, although extensions into higher dimensional processes are straightforward. The results presented in this section are standard in the theory of Lévy processes. For comprehensive treatments of various aspects of the theory and applications of Lévy processes we refer the reader to Bingham (1975), Sato (1999), Appelbaum (2004), Bertoin (2004), Cont & Tankov (2004), Protter (2005), Kyprianou (2006), and Andersen & Lipton (2013). For a concise overview of some of the specific examples of Lévy processes considered later in Chapter 6, we refer the reader to Schoutens (2003).

5.1.1 Infinitely divisible random variables

Lévy processes are closely related to infinitely divisible random variables. We begin with:

**Definition 5.1.1.** A random variable $X \in \mathbb{R}$ is said to be infinitely divisible if its characteristic function $\phi(\lambda) = \mathbb{E}[\exp(i\lambda X)]$, $\lambda \in \mathbb{R}$, can be written for any integer $n$ as the $n^{th}$ power of a characteristic function $\phi_n$, that is, if

$$
\phi(\lambda) = (\phi_n(\lambda))^n.
$$

(5.1)
Equivalently, $X$ is infinitely divisible if for each value of $n$ there exists a set of $n$ independent and identically distributed random variables $\{X_{i}^{(n)}\}_{i=1,...,n}$ such that we have

\[ X = X_{1}^{(n)} + X_{2}^{(n)} + \cdots + X_{n}^{(n)}. \]

**Example 5.1.1.** Random variables with distributions such as Gaussian, Cauchy, Poisson, gamma, variance gamma, inverse Gaussian, normal inverse Gaussian, and generalised hyperbolic are infinitely divisible. Uniformly distributed random variables are not infinitely divisible, nor are any bounded random variables.

### 5.1.2 Lévy processes: definitions and main properties

**Definition 5.1.2.** An $\mathbb{R}$-valued process $\{\xi_{t}\}_{t \geq 0}$, with $\xi_{0} = 0$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a Lévy process if it satisfies the following conditions: for every $s, t \geq 0$, $\xi_{t+s} - \xi_{s}$ is independent of the $\sigma$-algebra $\mathcal{F}_{s}$ generated by $\{\xi_{u}\}_{0 \leq u \leq s}$; for every $s, t \geq 0$, the random variables $\xi_{t+s} - \xi_{s}$ and $\xi_{t}$ have the same law; and the process $\{\xi_{t}\}$ is continuous in probability—that is to say, for fixed $t$, and for all $\epsilon > 0$, it holds that $\mathbb{P}(|\xi_{t} - \xi_{u}| > \epsilon) \to 0$ as $u \to t$.

There are many examples of Lévy processes that have been named and studied in depth. In what follows we shall be looking at Brownian motion, the Poisson processes, the gamma process, the variance gamma process, the negative binomial process, the inverse Gaussian process, the normal inverse Gaussian process, and the generalised hyperbolic process.

Due to the independent and stationary increments properties, it should be evident that for a Lévy process $\{\xi_{t}\}$ we have:

\[ \mathbb{E}\left[e^{i\lambda\xi_{t}}\right] = \left(\mathbb{E}\left[e^{i\lambda\xi_{1}}\right]\right)^{t}. \]

It follows that:

**Proposition 5.1.1.** If $\{\xi_{t}\}$ is a Lévy process, the random variable $\xi_{t}$ is infinitely divisible for all $t \geq 0$.

**Definition 5.1.3.** By a Lévy measure $\nu(dz)$ we mean a positive measure defined on $\mathbb{R}\setminus\{0\}$ satisfying

\[ \int_{\mathbb{R}\setminus\{0\}} (1 \wedge z^{2}) \nu(dz) < \infty. \]


The Lévy measure associated with a Lévy process has the following interpretation: if $B$ is a measurable subset of $\mathbb{R}\setminus\{0\}$, then $\nu(B)$ is the rate at which jumps arrive for which the jump size lies in $B$. Suppose we consider the sets defined for $n \in \mathbb{N}$ by

$$B_n = \{ z \in \mathbb{R} \mid 1/n \leq |z| \leq 1 \}.$$  \hspace{1cm} (5.4)

Let $\nu(dz)$ be the Lévy measure associated with a Lévy process $\{\xi_t\}$. Then if $\nu(B_n) \to \infty$ 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; and if $\nu(\mathbb{R}\setminus\{0\}) < \infty$ we say that $\{\xi_t\}$ has finite activity, since the process has finite number of jumps in any finite time interval.

With these observations in hand, we introduce the so-called Lévy-Khintchine formula for a Lévy process.

**Proposition 5.1.2.** Let $\{\xi_t\}$ be a Lévy process taking values in $\mathbb{R}$. Then for each $t > 0$, the random variable $\xi_t$ is infinitely divisible and its characteristic function is given for $\lambda \in \mathbb{R}$ by

$$\mathbb{E}[\exp(i\lambda \xi_t)] = \exp \left[ t \left( i\lambda p - \frac{1}{2}q\lambda^2 + \int_{\mathbb{R}} \left( e^{i\lambda z} - 1 - i\lambda z1\{|z|<1\}\right) \nu(dz) \right) \right]$$  \hspace{1cm} (5.5)

where $p \in \mathbb{R}$ and $q \geq 0$ are constants, and $\nu$ is a Lévy measure on $\mathbb{R}\setminus\{0\}$.

A Lévy process admits a so-called Lévy-Itô decomposition of the following form:

**Proposition 5.1.3.** An $\mathbb{R}$-valued Lévy process $\{\xi_t\}$ can be decomposed in the form

$$\xi_t = X_t^{(1)} + X_t^{(2)} + X_t^{(3)},$$  \hspace{1cm} (5.6)

where $X_t^{(1)}$ is a Brownian motion with drift, $X_t^{(2)}$ is a compound Poisson process with jump sizes greater than or equal to unity, and $X_t^{(3)}$ is a Lévy process with jump sizes less than unity. The processes $X_t^{(1)}$, $X_t^{(2)}$, and $X_t^{(3)}$ are independent.

By a Brownian motion with drift we mean a process of the form $\{qB_t + pt\}$, where $\{B_t\}$ is a standard Brownian motion.

**Definition 5.1.4.** The function $\phi : \mathbb{R} \to \mathbb{C}$ defined by

$$\mathbb{E}[\exp(i\lambda \xi_t)] = \exp (-t\phi(\lambda))$$  \hspace{1cm} (5.7)

for $\lambda \in \mathbb{R}$ is called the characteristic exponent of the Lévy process $\{\xi_t\}$.  

Definition 5.1.5. If $E[e^{\alpha \xi_t}] < \infty$ for $\alpha$ in some non-trivial connected region of $\mathbb{R}$ containing the origin, the function $\psi(\alpha)$ defined by

$$E[\exp(\alpha \xi_t)] = \exp(t\psi(\alpha))$$

on that region of $\mathbb{R}$ is called the Lévy exponent, or Laplace exponent, or cumulant function, of the Lévy process $\{\xi_t\}$.

It follows that $\psi(\alpha) = -\phi(-i\alpha)$. In particular, as a consequence of Proposition 5.1.2 we have

$$\phi(\lambda) = -ip\lambda + \frac{1}{2}q\lambda^2 - \int \left( e^{i\lambda z} - 1 - i\lambda z 1_{|z| < 1} \right) \nu(dz)$$

and

$$\psi(\alpha) = p\alpha + \frac{1}{2}q\alpha^2 + \int \left( e^{\alpha z} - 1 - \alpha z 1_{|z| < 1} \right) \nu(dz).$$

Remark 5.1.1. Bearing in mind the finite and infinite activity properties mentioned earlier, if we set $q = 0$, we can make the following observations. In the finite activity case, the process $\{\xi_t\}$ is a compound Poisson process with “drift”. In the infinite activity case, if $\int |z| \nu(dz) < \infty$, the paths of the process $\{\xi_t\}$ are of bounded variation on any finite time interval. Again, in the infinite activity case, if $\int |z| \nu(dz) = \infty$, the paths of the process $\{\xi_t\}$ are no longer of bounded variation on any finite time interval.

Remark 5.1.2. The term “Lévy process” (which, according to Çinlar 2011, was coined by Paul A. Meyer) refers to the French mathematician Paul Lévy (1886-1971). We summarise from Schoutens (2003) the main facts of Lévy’s career: Lévy was born in Paris and studied at the École Polytechnique. He obtained a doctoral degree in mathematics from the University of Paris, and became a professor at the École des Mines in Paris at the age of twenty-seven. He made numerous contributions to the theory of stochastic processes, and is generally regarded as one of the most prominent figures of modern probability theory. During the first world war, Lévy served in the artillery, and carried out mathematical analysis concerning defence against air attacks. In 1963 he was elected to an honorary membership of the London Mathematical Society, and in 1964 he was elected to the Académie des Sciences.

5.2 Lévy information: definition

From the discussion of the previous section, we know that a real-valued process $\{\xi_t\}_{t \geq 0}$ on a probability space $(\Omega, \mathcal{F}, P)$ is a Lévy process if $P(\xi_0 = 0) = 1$, $\{\xi_t\}$ has stationary
and independent increments, \(\lim_{t \to s} \mathbb{P}(|\xi_t - \xi_s| > \epsilon) = 0\), and \(\{\xi_t\}\) 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

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

(5.11)

If \(A\) contains points other than \(w = 0\), then we say that \(\{\xi_t\}\) possesses exponential moments. As indicated in (5.8) above, we define a function \(\psi : A \to \mathbb{R}\) called the Lévy exponent (or cumulant function) such that

\[
\mathbb{E}^\mathbb{P}[\exp(\alpha \xi_t)] = \exp(\psi(\alpha) t)
\]  

(5.12)

for \(\alpha \in A\). If a Lévy process possesses exponential moments, then \(\psi(\alpha)\) is convex on \(A\). This can be seen as follows. Observe that

\[
\psi(\alpha) = \frac{1}{t} \ln \mathbb{E}^\mathbb{P}[\exp(\alpha \xi_t)].
\]  

(5.13)

Differentiate with respect to \(\alpha\) we get

\[
\psi'(\alpha) = \frac{\partial \psi(\alpha)}{\partial \alpha} = \frac{1}{t} \frac{\mathbb{E}^\mathbb{P}[\xi_t \exp(\alpha \xi_t)]}{\mathbb{E}^\mathbb{P}[\exp(\alpha \xi_t)]}.
\]  

(5.14)

Take the second derivative, we obtain

\[
\psi''(\alpha) = \frac{\partial^2 \psi(\alpha)}{\partial \alpha^2}
= \frac{1}{t} \left[ \mathbb{E}^\mathbb{P}[\exp(\alpha \xi_t)] \mathbb{E}^\mathbb{P}[\xi_t^2 \exp(\alpha \xi_t)] - \mathbb{E}^\mathbb{P}[\xi_t \exp(\alpha \xi_t)] \mathbb{E}^\mathbb{P}[\xi_t \exp(\alpha \xi_t)] \right]
= \frac{1}{t} \frac{\mathbb{E}^\mathbb{P}[\xi_t - \hat{\xi}_t]^2 \exp(\alpha \xi_t)]}{\mathbb{E}^\mathbb{P}[\exp(\alpha \xi_t)]},
\]  

(5.15)

where

\[
\hat{\xi}_t = \frac{\mathbb{E}^\mathbb{P}[\xi_t \exp(\alpha \xi_t)]}{\mathbb{E}^\mathbb{P}[\exp(\alpha \xi_t)]}.
\]  

(5.16)

From equation (5.15) it is clear that

\[
\frac{\partial^2 \psi(\alpha)}{\partial \alpha^2} > 0,
\]  

(5.17)
and hence that the Lévy exponent is a convex function of $\alpha$. In fact, more generally, for any random variable $\xi$ satisfying

$$q(\alpha) = \mathbb{E}^P[\exp(\alpha \xi)] < \infty$$

(5.18)

for $\alpha$ in some interval containing the origin, $\ln q(\alpha)$ is convex (see e.g., Billingsley 1995).

The mean and variance of $\xi_t$ are given respectively by $\psi'(0)t$ and $\psi''(0)\tau$, and as a consequence of the convexity of $\psi(\alpha)$ the marginal exponent $\psi'(\alpha)$ possesses a unique inverse $I(y)$ such that

$$I(\psi'(\alpha)) = \alpha$$

(5.19)

for $\alpha \in A$. The Lévy exponent extends to a function $\psi : A_C \to \mathbb{C}$ where

$$A_C = \{w \in \mathbb{C} : \text{Re } w \in A\},$$

(5.20)

and it can be shown (Sato 1999, Theorem 25.17) that $\psi(\alpha)$ admits a Lévy-Khintchine representation of the form

$$\psi(\alpha) = p\alpha + \frac{1}{2}q\alpha^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{\alpha z} - 1 - \alpha z I\{|z| < 1\})\nu(dz)$$

(5.21)

with the property that (5.12) holds for all $\alpha \in A_C$. Here $I\{\cdot\}$ denotes the indicator function, $p \in \mathbb{R}$ and $q \geq 0$ are constants, and as indicated above, the Lévy measure $\nu(dz)$ is a positive measure defined on $\mathbb{R}\setminus\{0\}$ satisfying

$$\int_{\mathbb{R}\setminus\{0\}} (1 \wedge z^2)\nu(dz) < \infty.$$  

(5.22)

If the Lévy process possesses exponential moments then for $\alpha \in A$ it also holds that

$$\int_{\mathbb{R}\setminus\{0\}} e^{\alpha z} I\{|z| \geq 1\} \nu(dz) < \infty.$$  

(5.23)

Recall that the data $K = (p, q, \nu)$ is the characteristic triplet of the associated Lévy process. Thus we can classify a Lévy process abstractly by the specification of 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 $\lambda$ the process $\{\xi_t\}$ is $\mathbb{P}_\lambda$-Lévy. The changes of measure arising in this way are called Esscher transformations. For details, we refer the

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 \( \{ \rho^\lambda_t \} \) defined for \( \lambda \in A \) by the expression

\[
\rho^\lambda_t := \frac{dP_\lambda}{dP_0} \bigg|_{\mathcal{F}_t} = \exp (\lambda \xi_t - \psi_0(\lambda)t),
\]

where \( \mathcal{F}_t = \sigma(\{ \xi_s \}_{0 \leq s \leq t}) \). One can check that \( \{ \rho^\lambda_t \} \) is an \( (\{ \mathcal{F}_t \}, P_0) \)-martingale. Indeed, as a consequence of the fact that \( \{ \xi_t \} \) has stationary and independent increments we have

\[
E_{P_0}^s[\rho^\lambda_t] = E_{P_0}^s[e^{\lambda \xi_t}] e^{-t\psi_0(\lambda)} = e^{(t-s)\psi_0(\lambda)} e^{\lambda \xi_s - t\psi_0(\lambda)} = \rho^\lambda_s
\]

for \( s \leq t \). It is straightforward to show that \( \{ \xi_t \} \) has \( P_\lambda \) stationary and independent increments, and that the \( P_\lambda \) exponent of \( \{ \xi_t \} \), defined on the set

\[
A^\lambda_C := \{ w \in C : \text{Re } w + \lambda \in A \},
\]

is given by

\[
\psi^\lambda(\alpha) := \frac{1}{t} \ln E^\lambda[\exp(\alpha \xi_t)] = \psi_0(\alpha + \lambda) - \psi_0(\lambda),
\]

from which by use of the Lévy-Khintchine representation (5.21) one can work out the characteristic triplet \( K^\lambda \) associated with \( \{ \xi_t \} \) under \( P_\lambda \). In what follows we use the terms “signal” and “message” interchangeably. We write

\[
C^I = \{ w \in C : \text{Re } w = 0 \}.
\]

For any random variable \( Z \) on \( (\Omega, \mathcal{F}, P) \) we write \( \mathcal{F}^Z = \sigma[Z] \), and occasionally we write \( E^P[\cdot | Z] \) for \( E^P[\cdot | \mathcal{F}^Z] \). 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 single 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 something of 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 \((Ω, \mathcal{F}, P)\), and an Esscher family of Lévy characteristics \(K_λ, λ ∈ A\), with associated Lévy exponents \(ψ_λ(α), α ∈ A^λ_C\). We refer to \(K_0\) as the fiducial characteristic, and \(ψ_0(α)\) as the fiducial exponent. The intuition here is that the abstract Lévy process of characteristic \(K_0\) and exponent \(ψ_0(α)\), which we call the “fiducial” process, represents the noise type of the associated information process. Thus we can use \(K_0\), or equivalently \(ψ_0(α)\), to represent the noise type.

**Definition 5.2.1.** By a Lévy information process with fiducial characteristic \(K_0\), carrying the message \(X\), we mean a random process \(\{ξ_t\}\), together with a random variable \(X\), such that \(\{ξ_t\}\) is conditionally \(K_X\)-Lévy given \(\mathcal{F}^X\).

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

\[
ψ_X(α) := \frac{1}{t} \ln E_P[\exp(αξ_t) | \mathcal{F}^X] = ψ_0(α + X) - ψ_0(X)
\]  

(5.29)

for \(α ∈ C^l\), where \(ψ_0(α)\) is the fiducial exponent of the specified noise type. It is implicit in the statement of Definition 5.2.1 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

\[
P(X ∈ F) = 1.
\]  

(5.30)

Then we say that \(X\) is compatible with the fiducial exponent \(ψ_0(α)\) if \(S_X \subset A\). Intuitively speaking, this condition ensures that we can use \(X\) to make a random Esscher transformation. Note that we do not require that the Lévy information process should possess exponential moments under \(P\), but a sufficient condition for this to be the case is that \(S_X\) should be a proper subset of \(A\).
5.3 Asymptotic behaviour of Lévy information

We are thus able to state the Lévy noise-filtering problem as follows: given observations of the Lévy information process up to time $t$, what is the best estimate for $X$? 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 5.3.1.** Let $\{\xi_t\}$ be a Lévy information process with fiducial exponent $\psi_0(\alpha)$ and message $X$. Then for every $\epsilon > 0$ we have

$$\lim_{t \to \infty} \mathbb{P} \left[ \left| I_0 \left( \frac{\xi_t}{t} \right) - X \right| \geq \epsilon \right] = 0. \quad (5.31)$$

**Proof.** It follows from (5.29) that

$$\psi'_X(0) = \psi_0(X), \quad (5.32)$$

and hence that at any time $t$ the conditional mean of the random variable $\frac{1}{t} \xi_t$ is given by

$$\mathbb{E}^{\mathcal{F}^X} \left[ \frac{\xi_t}{t} \middle| \mathcal{F}^X \right] = \psi_0'(X). \quad (5.33)$$

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

$$\text{Var}^{\mathcal{F}^X} \left[ \frac{\xi_t}{t} \middle| \mathcal{F}^X \right] := \mathbb{E}^{\mathcal{F}^X} \left[ \left( \frac{\xi_t}{t} - \psi_0'(X) \right)^2 \middle| \mathcal{F}^X \right] = \frac{1}{t} \psi_0''(X), \quad (5.34)$$

which allows us to conclude that

$$\mathbb{E}^{\mathcal{F}^X} \left[ \left( \frac{\xi_t}{t} - \psi_0'(X) \right)^2 \right] = \frac{1}{t} \mathbb{E}^{\mathcal{F}^X} \left[ \psi_0''(X) \right], \quad (5.35)$$

and hence that

$$\lim_{t \to \infty} \mathbb{E}^{\mathcal{F}^X} \left[ \left( \frac{\xi_t}{t} - \psi_0'(X) \right)^2 \right] = 0. \quad (5.36)$$

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

$$\mathbb{P} \left[ \left| \frac{\xi_t}{t} - \psi_0'(X) \right| \geq \epsilon \right] \leq \frac{1}{\epsilon^2} \mathbb{E}^{\mathcal{F}^X} \left[ \left( \frac{\xi_t}{t} - \psi_0'(X) \right)^2 \right]. \quad (5.37)$$
by Chebychev’s inequality, from which we deduce that
\[
\lim_{t \to \infty} \mathbb{P} \left[ \left| \frac{\xi_t}{t} - \psi_0'(X) \right| \geq \epsilon \right] = 0,
\]
and the result follows on account of the invertibility of the marginal Lévy exponent. □

We see that \( I_0(\xi_t/t) \) converges to \( X \) in probability. It follows 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.

## 5.4 Existence 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 \( K_0 \). The process \( \{\rho_t^X\} \) defined by
\[
\rho_t^X = \exp(X\xi_t - \psi_0(X)t)
\]
is a \( (\{\mathcal{G}_t\}, \mathbb{P}_0) \)-martingale.

**Proof.** The derivation goes as follows:

\[
\mathbb{E}_{\mathbb{P}_0} \left[ \rho_t^X | \mathcal{G}_s \right] = \mathbb{E}_{\mathbb{P}_0} \left[ \exp(X\xi_t - \psi_0(X)t) | \mathcal{G}_s \right] = e^{X\xi_s - \psi_0(X)s} \mathbb{E}_{\mathbb{P}_0} \left[ \exp(X(\xi_t - \xi_s) - \psi_0(X)(t-s)) | \mathcal{G}_s \right] = e^{X\xi_s - \psi_0(X)s} \mathbb{E}_{\mathbb{P}_0} \left[ \exp(X(\xi_t - \xi_s) - \psi_0(X)(t-s)) \right] = e^{X\xi_s - \psi_0(X)s} e^{\psi_0(X)(t-s) - \psi_0(X)(t-s)} = \rho_s^X.
\]

Here we have used the fact that \( \xi_t \) is a Lévy process under the measure \( \mathbb{P}_0 \). □
We are thus able to introduce a change of measure $P_0 \rightarrow P_X$ on $(\Omega, \mathcal{F}, P_0)$ by setting

$$\frac{dP_X}{dP_0}|_{\mathcal{G}_t} = \rho_t^X. \tag{5.42}$$

It should be evident that $\{\xi_t\}$ is conditionally $P_X$-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 $\xi_t$ under $P_X$ is given by

$$\frac{1}{t} \ln \mathbb{E}^{P_X}[\exp(\alpha \xi_t) | \mathcal{F}^X] = \psi_0(\alpha + X) - \psi_0(X) \tag{5.43}$$

for $\alpha \in \mathbb{C}$, which shows that the conditions of Definition 5.2.1 are satisfied, allowing us to conclude the following:

**Proposition 5.4.1.** The $P_0$-Lévy process $\{\xi_t\}$ is a Lévy information process under $P_X$ with noise type $\psi_0(\alpha)$ and message $X$.

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 5.4.2.** Let $\{\xi_t\}$ be a Lévy information process on a probability space $(\Omega, \mathcal{F}, P)$ with noise type $\psi_0(\alpha)$ and message $X$. Then there exists a change of measure $P \rightarrow P_0$ such that $\{\xi_t\}$ and $X$ are $P_0$-independent, $\{\xi_t\}$ is $P_0$-Lévy with exponent $\psi_0(\alpha)$, and the probability law of $X$ under $P_0$ is the same as probability law of $X$ under the measure $P$.

**Proof.** First we establish that the process $\{\tilde{\rho}_t^X\}$ defined by

$$\tilde{\rho}_t^X = \exp(-X \xi_t + \psi_0(X)t) \tag{5.44}$$

is a $\{\mathcal{G}_t\}, \mathbb{P}$-martingale. We have

$$\mathbb{E}^P[\tilde{\rho}_s^X | \mathcal{G}_s] = \mathbb{E}^P[\exp(-X \xi_t + \psi_0(X)t) | \mathcal{G}_s]$$

$$= \mathbb{E}^P[\exp(-X(\xi_t - \xi_s)) | \mathcal{G}_s] \exp(-X \xi_s + \psi_0(X)t)$$

$$= \exp(\psi_X(-X)(t - s)) \exp(-X \xi_s + \psi_0(X)t), \tag{5.45}$$

by virtue of the fact that $\{\xi_t\}$ is $\mathcal{F}^X$-conditionally Lévy under $P$. By use of (5.29) we deduce that

$$\psi_X(-X) = -\psi_0(X), \tag{5.46}$$
and hence that

\[ E_P[\tilde{\rho}^X_{|G_s}] = \tilde{\rho}^X, \quad (5.47) \]

as required. Then we use \( \{\tilde{\rho}^X_t\} \) to define a change of measure \( P \rightarrow P_0 \) on \( (\Omega, \mathcal{F}, P) \) by setting

\[ \frac{dP_0}{dP} \bigg|_{G_t} = \tilde{\rho}^X. \quad (5.48) \]

To show that \( \xi_t \) and \( X \) are \( P_0 \)-independent for all \( t \), it suffices to show that their \( P_0 \) joint characteristic function factorises. Letting \( \alpha, \beta \in \mathbb{C} \), we have

\[ E_{P_0}[\exp(\alpha \xi_t + \beta X)] = \exp(\psi_0(\alpha t)) E_{P_0}[\exp(\beta X)], \quad (5.49) \]

where the last step follows from (5.29). This argument can be extended to show that \( \{\xi_t\} \) and \( X \) are \( P_0 \)-independent. Next we observe that

\[ E_{P_0}[\exp(\alpha(\xi_u - \xi_t) + \beta \xi_t)] \]

\[ = E_{P_0}[\exp((X - X)\xi_t + \psi_0(X)u + \alpha(\xi_u - \xi_t) + \beta \xi_t)] \]

\[ = E_{P_0}[E_{P_0}[\exp((X - X)\xi_t + \psi_0(X)u + \alpha(\xi_u - \xi_t) + \beta \xi_t) | \mathcal{F}_X]] \]

\[ = E_{P_0}[\psi_0(X)u + (\alpha - X)(\xi_u - \xi_t) + (\beta - X)\xi_t] | \mathcal{F}_X] \]

\[ = E_{P_0}[\exp(\psi_0(X)u + \psi_0(X)(u - t) + \psi_0(\beta - X)t)] \]

\[ = \exp(\psi_0(\alpha)(u - t)) \exp(\psi_0(\beta)t) \quad (5.50) \]

for \( u \geq t \geq 0 \), and it follows that \( \xi_u - \xi_t \) and \( \xi_t \) are independent. This argument can be extended to show that \( \{\xi_t\} \) has \( P_0 \)-independent increments. Finally, if we set \( \alpha = 0 \) in (5.49) it follows that the probability laws of \( X \) under \( P_0 \) and \( P \) are identical; if we set \( \beta = 0 \) in (5.49) it follows that the \( P_0 \) exponent of \( \{\xi_t\} \) is \( \psi_0(\alpha) \); and if we set \( \beta = 0 \) in (5.50) it follows that \( \{\xi_t\} \) is \( P_0 \)-stationary.

\[ \square \]

5.5 Conditional expectations

Going forward, we adopt the convention that \( P \) always denotes the “physical” measure in relation to which an information process with message \( X \) is defined, and that \( P_0 \) 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 5.4.2 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 can be written in terms of $\mathbb{P}_0$ expectations, and is given by a “generalised Bayes formula” (Kallianpur & Striebel 1968) of the form

$$
\mathbb{E}^\mathbb{P}[Z \mid \mathcal{F}_t] = \frac{\mathbb{E}^{\mathbb{P}_0}[\rho^X Z \mid \mathcal{F}_t]}{\mathbb{E}^{\mathbb{P}_0}[\rho^X \mid \mathcal{F}_t]}. 
$$

(5.51)

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

$$
F^X_t(y) = \mathbb{P}(X \leq y \mid \mathcal{F}_t)
$$

(5.52)

for $y \in \mathbb{R}$. In the Bayes formula we set $Z = \mathbb{I}\{X \leq y\}$, and the result is

$$
F^X_t(y) = \frac{\int \mathbb{I}\{x \leq y\} \exp(x\xi_t - \psi_0(x)t) \, dF^X(x)}{\int \exp(x\xi_t - \psi_0(x)t) \, dF^X(x)},
$$

(5.53)

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 $\pi_t(dx)$ induced on $\mathbb{R}$ defined by

$$
dF^X_t(x) = \pi_t(dx).
$$

(5.54)

In particular, when $X$ is a continuous random variable with a density function $p(x)$ one can write

$$
\pi_t(dx) = p_t(x)dx,
$$

(5.55)

where $p_t(x)$ is the conditional density function. With these results, the following should be evident.

**Proposition 5.5.1.** 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 $\pi(dx)$. Then the $\mathcal{F}_t$-conditional a posteriori distribution of $X$ is

$$
\pi_t(dx) = \frac{\exp(x\xi_t - \psi_0(x)t)}{\int \exp(x\xi_t - \psi_0(x)t) \pi(dx)} \pi(dx).
$$

(5.56)
It is straightforward to establish by use of a variational argument that the best estimate for the message \( X \) conditional on the information \( \mathcal{F}_t \) is given by

\[
\hat{X}_t := \mathbb{E}^P[X \mid \mathcal{F}_t] = \int x \pi_t(dx).
\] (5.57)

By “best estimate” we mean the \( \mathcal{F}_t \)-measurable random variable \( \hat{X}_t \) that minimises the quadratic error \( \mathbb{E}^P[(X - \hat{X}_t)^2 \mid \mathcal{F}_t] \). It will be observed that at any given time \( t \) the best estimate can be expressed as a function of \( \xi_t \) 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 5.5.2.** The Lévy information process \( \{\xi_t\} \) has the Markov property under the measure \( P \).

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

\[
\mathbb{P}(\xi_t \leq a \mid \mathcal{F}_s) = \mathbb{E}^P(\xi_t \leq a \mid \mathcal{F}_s),
\] (5.58)

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

\[
\Phi_t := \mathbb{E}^{P_0}[\rho_t^X \mid \mathcal{F}_t] = \int \exp(x\xi_t - \psi_0(x)t) \pi(dx),
\] (5.59)

where \( \rho_t^X \) is defined as in equation (5.40). It follows that

\[
\mathbb{P}(\xi_t \leq a \mid \mathcal{F}_s) = \mathbb{E}^P[1\{\xi_t < a\} \mid \mathcal{F}_s]
= \frac{\mathbb{E}^{P_0}[\Phi_t 1\{\xi_t < a\} \mid \mathcal{F}_s]}{\mathbb{E}^{P_0}[\Phi_t \mid \mathcal{F}_s]}
= \frac{\mathbb{E}^{P_0}[\Phi_t 1\{\xi_t < a\} \mid \mathcal{F}_s]}{\mathbb{E}^{P_0}[\Phi_t \mid \mathcal{F}_s]}
= \mathbb{E}^P[1\{\xi_t < a\} \mid \mathcal{F}_s]
= \mathbb{P}(\xi_t \leq a \mid \mathcal{F}_s),
\] (5.60)

since \( \{\xi_t\} \) has the Markov property under the transformed measure \( P_0 \).

\[\square\]

We note that since \( X \) is \( \mathcal{F}_\infty \)-measurable, which follows from Proposition 5.3.1, the Markov property implies that

\[
\mathbb{E}^P[X \mid \mathcal{F}_t] = \mathbb{E}^P[X \mid \mathcal{F}_t].
\] (5.61)
This identity is useful if one wishes 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 (5.56) that, given the conditional distribution \( \pi_s(dx) \) at time \( s \leq t \), we can express \( \pi_t(dx) \) in the form

\[
\pi_t(dx) = \frac{\exp \left( x(\xi_t - \xi_s) - \psi_0(x)(t - s) \right)}{\int \exp \left( x(\xi_t - \xi_s) - \psi_0(x)(t - s) \right) \pi_s(dx)} \pi_s(dx).
\]

(5.62)

Then if for fixed \( s \geq 0 \) we introduce a new time variable \( u := t - s \), and define

\[
\eta_u = \xi_{u+s} - \xi_s,
\]

(5.63)

we find that \( \{\eta_u\}_{u \geq 0} \) is a Lévy information process with fiducial exponent \( \psi_0(\alpha) \) and message \( X \) with a priori distribution \( \pi_s(dx) \). Thus given up-to-date information we can “re-start” the information process at that time so as to produce a new information process of the same type, with an adjusted distribution for the message.

### 5.6 General characterisation of Lévy information

A general characterisation of the nature of Lévy information can be inferred by examination of expression (5.29) for the conditional exponent of an information process. In particular, as a consequence of the Lévy-Khintchine representation (5.21) we deduce that

\[
\psi_0(\alpha + X) - \psi_0(X) = \left( p + qX + \int_{R \setminus \{0\}} z(e^{Xz} - 1)1_{\{|z| < 1\}}\nu(dz) \right) \alpha + \frac{1}{2}q\alpha^2 + \int_{R \setminus \{0\}} (e^{az} - 1 - az1_{\{|z| < 1\}})e^{Xz}\nu(dz)
\]

(5.64)

for \( \alpha \in \mathbb{C}^1 \), which leads to the following:

**Proposition 5.6.1.** The randomisation of the \( \mathbb{P}_0 \)-Lévy process \( \{\xi_t\} \) achieved through the change of measure generated by the density

\[
\rho_t = \exp(X\xi_t - \psi_0(X)t)
\]

(5.65)
induces two effects on the characteristics of the process: (i) a random shift in the drift term, given by

\[ p \to p + qX + \int_{\mathbb{R}\setminus\{0\}} z(e^{Xz} - 1)1\{|z| < 1\})\nu(dz), \tag{5.66} \]

and (ii) a random rescaling of the Lévy measure, given by

\[ \nu(dz) \to e^{Xz}\nu(dz). \tag{5.67} \]

Note that the integral appearing in the definition of the random shift in the drift term is well defined since the term \( z(e^{Xz} - 1) \) vanishes to second order at the origin. It follows from Proposition 5.6.1 that in sampling the values of an information process an agent is in effect trying to detect a random shift in the drift term, as well as an overall 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 determine the value of \( X \).

5.7 Martingales associated with Lévy information

We turn to examine the properties of certain martingales of importance associated with Lévy information. More specifically, 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; Segall et al. 1975). These examples can be seen as arising as special cases in the general theory of Lévy information processes. Throughout the discussion that follows we fix a probability space \((\Omega, \mathcal{F}, \mathbb{P})\).

**Proposition 5.7.1.** 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

\[ \xi_t = \int_0^t \hat{Y}_u \, du + M_t \tag{5.68} \]

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 (5.33) 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

\[
\rho^X_t \epsilon = \exp \left( (X + \epsilon)\xi_t - \psi_0(X + \epsilon)t \right)
\]

for \( \epsilon \in \mathbb{C}^1 \). Expanding this expression to first order in \( \epsilon \), we deduce that the process defined for \( t \geq 0 \) by

\[
\rho^X_t (\xi_t - \psi'_0(X)t) \mid \mathcal{G}_s = \rho^X_s (\xi_s - \psi'_0(X)s).
\]

Then using \( \{\rho^X_t\} \) to make a change of measure from \( \mathbb{P}_0 \) to \( \mathbb{P} \) we obtain

\[
\mathbb{E}^\mathbb{P}[\xi_t - \psi'_0(X)t \mid \mathcal{G}_s] = \xi_s - \psi'_0(X)s,
\]

and the result follows if we set \( Y = \psi'_0(X) \). Next we introduce the “projected” process \( \{\hat{m}_t\} \) defined by \( \hat{m}_t = \mathbb{E}^\mathbb{P}[m_t \mid \mathcal{F}_t] \). We note that since \( \{m_t\} \) is a \( (\{\mathcal{G}_t\}, \mathbb{P}) \)-martingale we have

\[
\mathbb{E}^\mathbb{P}[\hat{m}_t \mid \mathcal{F}_s] = \mathbb{E}^\mathbb{P}[\xi_t - \int_0^t \hat{Y}_u du \mid \mathcal{F}_s]
\]

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

\[
\mathbb{E}^\mathbb{P}[M_t \mid \mathcal{F}_s] = \mathbb{E}^\mathbb{P}[\xi_t - \int_0^t \hat{Y}_u du \mid \mathcal{F}_s]
\]

\[
= \mathbb{E}^\mathbb{P}[\xi_t \mid \mathcal{F}_s] - \mathbb{E}^\mathbb{P}\left[\int_s^t \hat{Y}_u du \mid \mathcal{F}_s\right] - \int_0^s \hat{Y}_u du,
\]
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

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

(5.75)

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

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

$$
X = I_0(Y),
$$

(5.76)

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.

### 5.8 On the role of Legendre transforms

It is worth pointing out that there is an interesting expression that one can deduce for the change-of-measure density martingale that involves the Legendre transformation. This relation leads to a conjecture, the validity of which we have not been able to establish. It seems nevertheless of interest to point this out as one of the open questions in relation to the theory of nonlinear filtering associated with Lévy processes. To this end, we recall from (5.68) that the innovations process {$M_t$} is given by

$$
M_t = \xi_t - \int_0^t \hat{Y}_u \, du,
$$

(5.77)

which evidently is an $(\{\mathcal{F}_t\}, \mathbb{P})$-martingale. In a differential form we can thus write

$$
d\xi_t = dM_t + \hat{Y}_t dt.
$$

(5.78)

On the other hand, let us define a process {$r_t$} according to

$$
r_t = \exp \left[ \int_0^t \lambda_s d\xi_s - \int_0^t \psi_0(\lambda_s) \, ds \right]
$$

(5.79)
for some \( \{ \lambda_t \} \) which is \( \{ \mathcal{F}_t \} \) predictable. As shown earlier in equation (5.40), \( \{ r_t \} \) is a \( \mathbb{P}_0 \)-martingale. For a positive real number \( \epsilon \) let us define a parametric family of processes \( \{ r_t(\epsilon) \} \) by

\[
 r_t(\epsilon) = \exp \left( \int_0^t (\lambda_s + \epsilon)d\xi_s - \int_0^t \psi_0(\lambda_s + \epsilon)ds \right). \tag{5.80}
\]

Now, differentiate with respect to \( \epsilon \) and set \( \epsilon \) equal to zero, we get:

\[
 r'_t(0) = r_t \left( \xi_t - \int_0^t \psi'_0(\lambda_s)ds \right). \tag{5.81}
\]

We can use \( \{ r_t \} \) as a change of measure martingale from \( \mathbb{P}_0 \rightarrow \mathbb{P}_\lambda \), and introduce the process \( \{ N_t \} \), which is a martingale under \( \mathbb{P}_\lambda \), by the expression

\[
 N_t = \xi_t - \int_0^t \psi'_0(\lambda_s)du. \tag{5.82}
\]

Keeping in mind that \( \{ M_t \} \) is a martingale under \( \mathbb{P} \), and \( \{ N_t \} \) is a martingale under \( \mathbb{P}_\lambda \), if we compare (5.77) and (5.82) and set

\[
 \psi'_0(\lambda_t) = \hat{Y}_t, \tag{5.83}
\]

we then have \( M_t = N_t \). This, of course, does not imply that \( \mathbb{P}_{\lambda^*} = \mathbb{P} \), which is subject to further investigation.

Let us recall that the fiducial exponent \( \psi_0(\alpha) \) is convex. Hence, the marginal fiducial exponent is monotonic and thus invertible. It follows that

\[
 \lambda^*_t = I_0(\hat{Y}_t). \tag{5.84}
\]

If we substitute (5.84) back into (5.79), then we get

\[
 r_t = \exp \left[ \int_0^t I_0(\hat{Y}_s)d\xi_s - \int_0^t \psi_0(I_0(\hat{Y}_s))ds \right]. \tag{5.85}
\]

Now, \( \{ r_t \} \) is defined as a change of measure martingale from \( \mathbb{P}_0 \rightarrow \mathbb{P}_{\lambda^*} \). Thus, we can reverse the measure change by use of the martingale \( \{ \tilde{r}_t \} \) defined by

\[
 \tilde{r}_t = \frac{1}{r_t} = \exp \left[ -\int_0^t I_0(\hat{Y}_s)d\xi_s + \int_0^t \psi_0(I_0(\hat{Y}_s))ds \right]. \tag{5.86}
\]

Using (5.78), we can rewrite the above expression as follows:

\[
 \tilde{r}_t = \exp \left[ -\int_0^t I_0(\hat{Y}_s)dM_s + \int_0^t \psi_0(I_0(\hat{Y}_s))ds - \int_0^t I_0(\hat{Y}_s)\hat{Y}_s ds \right]. \tag{5.87}
\]
We now recall that a Legendre-transformation of $\psi$, which is defined by
\[
\tilde{\psi}(y) = \max_x (xy - \psi(x)),
\] (5.88)
gives
\[
\tilde{\psi}(y) = yI_0(y) - \psi(I_0(y)).
\] (5.89)
With this in mind, (5.87) can be rewritten in the form:
\[
\hat{r}_t = \exp \left[ - \int_0^t I_0(\hat{Y}_s) dM_s - \int_0^t \tilde{\psi}_0(\hat{Y}_s) ds \right],
\] (5.90)
which defines a change of measure from $\mathbb{P}_{\lambda^*} \rightarrow \mathbb{P}_0$. On the other hand, recall from (5.59) that we can effect a change of measure from $\mathbb{P} \rightarrow \mathbb{P}_0$ by:
\[
\frac{1}{\Phi_t} = \frac{1}{\int \exp(x\xi_t - \psi_0(x)t)\pi(dx)}.
\] (5.91)
Since $\{M_t\}$ is a martingale in $\mathbb{P}$ as well as in $\mathbb{P}_{\lambda^*}$, this leads to the following conjecture:

**Conjecture 5.8.1.** The probability measures $\mathbb{P}_{\lambda^*}$ and $\mathbb{P}$ are the same, and hence the following relation holds:
\[
\exp \left[ \int_0^t I_0(\hat{Y}_s) dM_s + \int_0^t \tilde{\psi}_0(\hat{Y}_s) ds \right] = \int \exp(x\xi_t - \psi_0(x)t)\pi(dx).
\] (5.92)
In the next chapter, we show, by an application of Ito’s lemma, that this relation does hold in the case of Brownian motion.

### 5.9 Time-dependent Lévy information

Consider the case where time dependency is introduced by a time varying signal strength. In this setup, the fixed random variable $X$ representing the signal becomes a time varying signal $X \int_0^t \sigma_s ds$. Here $\{\sigma_t\}$ is a deterministic function of time $t \geq 0$. In this section, results for such time dependent Lévy information will be presented. We will also touch upon the connection between general Lévy information and the so-called change-point detection problem.
5.9.1 Time-dependent information flow rate

The analysis involving the time-dependent information process, when the time dependency is merely a deterministic adjustment of the information flow rate, is entirely analogous to that of the time-independent case presented earlier. If we recall the change-of-measure martingale (5.40), it should be evident that the relevant density martingale in the time-dependent case should take the form

\[
\rho_t^\sigma = \exp \left( \int_0^t X_{\sigma_s} d\xi_s - \int_0^t \psi_0(X_{\sigma_s}) ds \right),
\]

where the superscript \(\sigma\) in \(\rho_t\) above represents the fact that we are working in the time-dependent information flow rate setup. It is straightforward to check that this is indeed a \(({\mathcal{G}_t}, P_0)\)-martingale. Recall the Kallianpur-Striebel formula (5.51), which in the present context generalises to the following:

\[
E^P[Z | \mathcal{F}_t] = \frac{E^{P_0}[\rho_t^\sigma Z | \mathcal{F}_t]}{E^{P_0}[\rho_t^\sigma | \mathcal{F}_t]},
\]

(5.94)

Now let the random variable \(Z\) be given by \(Z = \mathbb{1}\{X \leq y\}\), and we thus deduce that

\[
F_t^X(y) = \frac{\int \mathbb{1}\{x \leq y\} \exp \left( x \int_0^t \sigma_s d\xi_s - \int_0^t \psi_0(x_{\sigma_s}) ds \right) dF^X(x)}{\int \exp \left( \int_0^t x_{\sigma_s} d\xi_s - \int_0^t \psi_0(x_{\sigma_s}) ds \right) dF^X(x)},
\]

(5.95)

where \(F^X(y) = \mathbb{P}(X < y)\) is the a priori distribution function. Equivalently, we can work with the measure \(\pi(dx)\) induced on \(\mathbb{R}\) defined by \(dF^X(x) = \pi(dx)\). We thus have:

**Proposition 5.9.1.** Let \(\{\xi_t\}\) be a Lévy information process under \(P\) with noise type \(\psi_0(\alpha)\), and let the a priori distribution of the associated message \(X\) be \(\pi(dx)\). Then in the case of a time-dependent information flow rate, the \(\mathcal{F}_t\)-conditional a posteriori distribution of \(X\) is

\[
\pi_t(dx) = \frac{\exp \left( x \int_0^t \sigma_s d\xi_s - \int_0^t \psi_0(x_{\sigma_s}) ds \right) \pi(dx)}{\int \exp \left( x \int_0^t \sigma_s d\xi_s - \int_0^t \psi_0(x_{\sigma_s}) ds \right) \pi(dx)}.
\]

(5.96)

As in the time-independent setup, the best estimate for the message \(X\) given the information \(\mathcal{F}_t\) is

\[
\hat{X}_t = \int x \pi_t(dx).
\]

(5.97)
Hence, the best estimate takes the form
\[
\hat{X}_t = \frac{\int_0^\infty xp(x) \exp \left( x \int_0^t \sigma_s d\xi_s - \int_0^t \psi_0(x\sigma_s) ds \right) dx}{\int_0^\infty p(x) \exp \left( x \int_0^t \sigma_s d\xi_s - \int_0^t \psi_0(x\sigma_s) ds \right) dx}.
\] (5.98)

### 5.9.2 Application to change-point detection problem

Change point detection is the problem of filtering information about points in time at which properties of a time-series data stream change. In this section we consider an application of the theory of time-dependent information processes to a change-point detection problem of the following type. Up to a random time \( \tau \), which is exponentially distributed, the information process \( \{\xi_t\} \) carries no signal, i.e., it is pure Lévy noise, but after \( \tau \) it carries a signal, given by a constant \( \theta \) (the value of this constant is assumed to be known). The objective is to derive the optimal estimate for the time at which the change has occurred or will occur. In the Brownian context, this has been examined by a number of authors, dating back to early work by Shiryaev (1969). We refer the reader to the treatments given, for example, by Karatzas (2003) and by Vellekoop & Clark (2006) for an indication of where modern efforts stand. We show in what follows that there is a natural extension of the basic elements of the problem to the Lévy scenario.

To proceed, we define a time-dependent signal term in (5.93) by replacing the expression \( X\sigma_s \) with the expression \( \theta \mathbb{1}\{s > \tau\} \) where \( \theta \) is a constant. The signal is thus in effect being replaced by the random time \( \tau \), which we assume to have the density \( p(u) \), and we are able as a consequence to make the following substitutions in the previously derived expression for the conditional density. We have that
\[
\int_0^t X\sigma_s ds \rightarrow (t - \tau)\theta \mathbb{1}\{t \geq \tau\},
\] (5.99)
and that
\[
\int_0^t X\sigma_s d\xi_s \rightarrow \theta \int_\tau^t d\xi_s = \theta (\xi_t - \xi_\tau).
\] (5.100)

In addition,
\[
\int_0^t \psi_0(X\sigma_s) ds \rightarrow \int_\tau^t \psi_0(\theta) ds = (t - \tau)\psi_0(\theta).
\] (5.101)

With these changes in place, we find that the least-square optimal estimate for the change point is given by
\[
\hat{\tau}_t = \frac{\int_0^\infty up(u) \exp (\theta(\xi_t - \xi_u) - (t - u)\psi_0(\theta)) du}{\int_0^\infty p(u) \exp (\theta(\xi_t - \xi_u) - (t - u)\psi_0(\theta)) du}.
\] (5.102)
Thus at any time $t > 0$, given knowledge of the current value of the Lévy information process, we can work out the conditional probability density for the change-point time, and hence determine its mean and variance, and other relevant statistics. This provides the basis for a generalisation of the Brownian version of the problem considered by Karatzas (2003) and various other authors to a rather general Lévy setting.

### 5.10 Entropy and mutual information

The concept of information is perhaps too broad to be captured completely by means of a single definition. However, for any probability density function (or a discrete set of probabilities), we can define a quantity called the entropy, which, up to sign convention, has many properties that agree with the intuitive notion of what a measure of information ought to be. This notion is extended to define the mutual information between two random variables, which is a measure of the amount of information one random variable contains about another. Entropy then becomes the “self-information” of a random variable. Mutual information is in fact an example of an entropic measure known as the relative entropy, and provides a distance measure between the joint probability and the product of marginal probabilities. All these quantities share a number of properties, some of which are discussed below.

In this section, we shall briefly introduce the notion and relationship between entropy and mutual information. The purpose in doing so is because in the next chapter, mutual information of various information processes will be calculated and compared. That is to say, given the value at time $t$ of a Lévy information process, we can ask how much information is contained in $\xi_t$ about the unknown quantity $X$.

#### 5.10.1 Entropy and uncertainty

In information theory, entropy is a measure of the uncertainty associated with a random variable, and negative entropy is defined to be the amount of information. The concept was introduced by Claude E. Shannon in his 1948 paper *A Mathematical Theory of Communication*, and by Norbert Wiener in his 1948 book *Cybernetics: or the Control and Communication in the Animal and the Machine*. The Wiener-Shannon entropy represents a measure of the average information content that is missing when one does not know the value of the random variable. In the case of a discrete random variable $X$ that takes values $\{x_i\}_{i=1,\ldots,N}$ with probabilities $\{p_i\}_{i=1,\ldots,N}$, the entropy $H$ is defined
by the formula

\[ H_X = - \sum_{i=1}^{N} p_i \ln p_i. \quad (5.103) \]

From the definition (5.103) we can obtain intuition about the measure of uncertainty associated with the value of \( X \). A traditional measure of uncertainty is the standard deviation of \( X \), which provides an estimate for the degree of spread about the possible values that \( X \) can take. Entropic measure of uncertainty is similar to that of standard deviation. In particular, it is evident from (5.103) that \( H_X \) takes its minimum value \( H_X = 0 \) when \( p_k = 1 \) for some \( k \) and \( p_i = 0 \) for \( i \neq k \). In other words, entropy is minimised when there is no uncertainty. Conversely, when there is maximum amount of uncertainty, that is, when \( p_i = 1/N \) for all \( i \), then \( H_X \) takes its maximum value \( H_X = \ln N \). This example also illustrates the fact that standard deviation does not provide an accurate information about the spread of \( X \) since the uniform distribution does not maximise standard deviation.

In the case of a continuous random variable \( X \) with density \( p(x) \), the analogue of (5.103) reads

\[ H_X = - \int p(x) \ln p(x) \, dx. \quad (5.104) \]

Note, however, that the continuous version of entropy (5.104), as compared to its discrete counterpart (5.103), is defined only up to an additive constant, and it is meaningful only as much as one is comparing the entropies associated with two or more density functions. This follows on account of the fact that a continuous density function \( p(x) \) has the dimension of \( [x^{-1}] \); hence \( \ln p(x) \) is ill defined unless \( p(x) \) is replaced with \( cp(x) \) where \( c \) is an arbitrary nonzero constant of dimension \( [x] \).

In the case where one has a pair of random variables, say, \( X \) and \( Y \), we can speak about joint entropy and conditional entropy. If we let \( p_{ij} \) denote the probability that the random variable pair \((X, Y)\) takes the values \((x_i, y_j)\), and \( p_{i|j} \) the conditional probability that \( X = x_i \) given \( Y = y_j \), then the joint entropy is defined by

\[ H_{X,Y} = - \sum_{i,j} p_{i,j} \ln p_{i,j}, \quad (5.105) \]

whereas for the conditional entropy we have

\[ H_{X|Y} = - \sum_{i,j} p_{i,j} \ln p_{i|j}. \quad (5.106) \]
On account of the relation \( p_{i,j} = p_{i|j}p_j = p_{j|i}p_i \), where

\[
p_i = \sum_j p_{i,j} \quad \text{and} \quad p_j = \sum_i p_{i,j}
\]  \hspace{1cm} (5.107)

are marginal probabilities, one can verify the relation

\[
H_{X,Y} = H_{X|Y} + H_Y = H_{Y|X} + H_X.
\]  \hspace{1cm} (5.108)

We refer the readers to Cover & Thomas (1991) for an in-depth discussion about a wide range of information and uncertainty measures, and their various relations to one another.

### 5.10.2 Mutual information

If a pair of random variables \( X \) and \( \xi \) is given, and if they are not independent, then knowledge of \( \xi \), for instance, will evidently provide partial information about \( X \). To illustrate this idea, let us elaborate further the example considered by Wiener (1948).

Suppose that \( X \) is a random variable with density

\[
p_X(x) = \frac{1}{\sqrt{2\pi a}} \exp \left( -\frac{x^2}{2a} \right),
\]  \hspace{1cm} (5.109)

where \( a \in \mathbb{R}^+ \). Additionally, let the random variable \( \epsilon \), independent of \( X \), be also normally distributed with density

\[
p_\epsilon(y) = \frac{1}{\sqrt{2\pi b}} \exp \left( -\frac{y^2}{2b} \right),
\]  \hspace{1cm} (5.110)

where \( b \in \mathbb{R}^+ \). We can think of \( X \) as the “signal” that requires to be determined, whereas \( \epsilon \) represents background noise that obscures the value of \( X \). The observation is thus given by

\[
\xi = X + \epsilon,
\]  \hspace{1cm} (5.111)

and the question thus arising is as follows: How much information is contained in \( \xi \) about the value of \( X \)?

Now the initial uncertainty about \( X \) is characterised by the entropy associated with the \textit{a priori} density \( p_X(x) \). The knowledge of \( \xi \), however, will reduce the uncertainty since the entropy associated with the \textit{a posteriori} density \( p_X(x|\xi) \) will be smaller than the initial entropy. Thus, roughly speaking, the difference in entropy provides a measure of reduction in the amount of uncertainty. Indeed, on account of the Bayes formula the \textit{a}
The posterior density of $X$ is given by
\[
p_X(x|\xi) = \frac{1}{2\pi\sqrt{ab}} \exp\left(-\frac{(\xi-x)^2}{2b}\right) \exp\left(-\frac{x^2}{2a}\right),
\]
from which we can readily calculate the entropy change
\[
\Delta H_X = -\int p_X(x|\xi) \ln p_X(x|\xi) dx - \left(-\int p_X(x) \ln p_X(x) dx\right)
\]
\[
= -\frac{1}{2} \ln \left(1 + \frac{a}{b}\right).
\]
Clearly $\Delta H_X < 0$, indicating that the amount of uncertainty has reduced. Observe in this example that the amount of entropy change is independent of the observation $\xi$. In the limit where $a \ll b$ holds, that is, either the initial uncertainty about the value of $X$ is already negligible, or the noise is so widely spread (or both), the reduction of uncertainty approaches zero, as should be the case.

The change in entropy thus provides an estimate in the reduction of uncertainty. The amount of information one gains from the data $\xi$ can be quantified more precisely by means of mutual information. The mutual information between the random variables $X$ and $\xi$ is defined by the expression
\[
I(\xi, X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\xi X}(z, x) \ln \frac{p_{\xi X}(z, x)}{p_\xi(z) p_X(x)} dx dz,
\]
where $p_{\xi X}(z, x)$ is the joint density function of the random variables $(\xi, X)$, and $p_\xi(z)$ and $p_X(x)$ are the respective marginal densities. In the present example with $\xi$ and $X$ related according to (5.111) we have
\[
p_{\xi X}(z, x) = \frac{1}{2\pi\sqrt{ab}} \exp\left(-\frac{(z-x)^2}{2b}\right) \exp\left(-\frac{x^2}{2a}\right)
\]
and
\[
p_\xi(z) = \frac{1}{\sqrt{2\pi(a+b)}} \exp\left(-\frac{z^2}{2(a+b)}\right),
\]
from which it follows that
\[
I(\xi, X) = \frac{1}{2} \ln \left(1 + \frac{a}{b}\right),
\]
which is indeed the same as the magnitude of the absolute entropy change. The mutual information $I(\xi, X)$ is symmetric in $\xi$ and $X$. In terms of the entropic measures
introduced above, a short calculation shows that

$$I(\xi, X) = H_\xi - H_{\xi|X} = H_X - H_{X|\xi}. \quad (5.118)$$

In information theory, a commonly used notion of a distance measure between a pair of density functions $p(x)$ and $q(x)$ is that of the Kullback-Leibler information distance (relative entropy) (Kullback & Leibler 1951), which is given by

$$D(p|q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} \, dx. \quad (5.119)$$

By comparing (5.114) and (5.119) we thus find that the mutual information between $\xi$ and $X$ is just the relative entropy between the joint density $p_{\xi X}(z, x)$ of $\xi$ and $X$, and the product $p_\xi(z)p_X(x)$ of the marginals. With this background material in mind, in the next chapter we shall work out, for some of the specific examples of Lévy information processes, the amount of information $I(\xi_t, X)$ contained in $\xi_t$ at time $t$ about the value of the unknown quantity $X$. The time evolution of $I(\xi_t, X)$ then determines the “learning curve” — in particular, since $X$ is $\mathcal{F}_\infty$-measurable, the information one gathers from the observation of the Lévy information will asymptotically reach the amount $H_X$ of the initial uncertainty.
Chapter 6

Examples of Lévy Information Processes

6.1 Chapter overview

In a number of situations it turns out that 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, the normal inverse Gaussian, and the generalised hyperbolic 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.

6.2 Brownian information process

6.2.1 Definition and properties

A general account of the history and theory of Brownian motion can be found in Karatzas & Shreve (1991). We recall a few details of the origin of the subject. In 1828 the botanist Robert Brown observed through a microscope the irregular movement of pollen suspended in water. This became known as Brownian movement. The study of these random movements, generated by collisions between the pollen particles and water molecules, led to many important applications. The first quantitative work on properties of Brownian motion came in the work of Louis Bachelier, in his PhD thesis entitled
“Théorie de la spéculaction”, published in 1900, where he investigated the stochastic modelling of share price movements. In 1905 Einstein published a fundamental work entitled “On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat”, in which he derived the transition density for Brownian motion in the theory of heat.

A rigorous treatment of Brownian motion was carried out by Norbert Wiener in the period 1923-1924, in the works “Differential space” and “Un problème de probabilités dénombrables”, in which he provides the first existence proof. For this reason Brownian motion is also known as the Wiener process. In Part I of this thesis we have made an ample use of Brownian motion, and its re-introduction is in some respects redundant. Nevertheless, since in this Chapter we shall be examining a range of specific examples of Lévy processes, some of which are perhaps less known and thus deserve a brief introduction, for the sake of balance we begin this section with a brief reminder of the definition of Brownian motion. For further details of the theory of Brownian motion, we refer the reader to Hida (1980).

**Definition 6.2.1** (Brownian motion). A continuous stochastic process \( \{B_t\} \) adapted to the filtration \( \{\mathcal{F}_t\}_{0 \leq t < \infty} \) is called a standard Brownian motion (Wiener process) on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) provided: \( B_0 = 0 \) almost surely; for \( 0 \leq s < t \), the increment \( B_t - B_s \) is independent of \( \mathcal{F}_s \); for \( 0 \leq s < t \), the random variable \( B_t - B_s \) is normally distributed with mean zero and variance \( t - s \).

Since the increment \( B_t - B_s \) for \( 0 \leq s < t \) depends on \( s \) and \( t \) only through the difference \( t - s \), we say that \( \{B_t\} \) has stationary, independent increments. Although the definition of a Brownian motion presented above depends on the specific choice of filtration \( \{\mathcal{F}_t\} \), if we are given the process \( B_t \) but no filtration, and if we know that \( \{B_t\} \) has stationary, independent increments and that \( B_t - B_0 \) is normally distributed with mean zero and variance \( t \), then \( \{B_t\} \) is an \( \{\mathcal{F}_t^B\} \)-Brownian motion with respect to its own natural filtration.

### 6.2.2 Brownian information

On a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), let \( \{B_t\} \) be a standard Brownian motion, let \( X \) be an independent random variable, and set

\[
\xi_t = X t + B_t.
\]

The random process \( \{\xi_t\} \) thereby defined, which we call a 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, we see that 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 \( X \), which is consistent with Proposition 5.3.1 if we note that \( \psi_0'(\alpha) = \alpha \) and \( I_0(y) = y \) in the standard Brownian case.

The best estimate for \( X \) conditional on \( \mathcal{F}_t \) is given by (5.57), which can be derived by use of the generalised Bayes formula (5.51). 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 5.5.2 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

\[ \frac{\xi_s}{s} - \frac{\xi_{s_1}}{s_1} = \frac{B_s}{s} - \frac{B_{s_1}}{s_1}. \]  

Extending this line of reasoning, we see that for any \( a \in \mathbb{R} \) we have

\[
\mathbb{P}(\xi_t \leq a \mid \xi_s, \xi_{s_1}, \ldots, \xi_{s_k}) = \mathbb{P}(\xi_t \leq a \mid \xi_s, s^{-1}\xi_s - s_1^{-1}\xi_{s_1}, \ldots, s_{k-1}^{-1}\xi_{s_{k-1}} - s_k^{-1}\xi_{s_k}) \\
= \mathbb{P}(\xi_t \leq a \mid \xi_s, s^{-1}B_s - s_1^{-1}B_{s_1}, \ldots, s_{k-1}^{-1}B_{s_{k-1}} - s_k^{-1}B_{s_k}) \\
= \mathbb{P}(\xi_t \leq a \mid \xi_s),
\]

since \( \xi_t \) and \( \xi_s \) 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 (5.58). Since we have established that \( X \) is \( \mathcal{F}_\infty \)-measurable, it follows that (5.61) 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
\[ \hat{X}_t = \frac{\int x \exp(x\xi_t - \frac{1}{2}x^2t) \pi(dx)}{\int \exp(x\xi_t - \frac{1}{2}x^2t) \pi(dx)}. \] (6.5)

The innovations representation (5.68) 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 (5.59) is a “space-time” function of the form
\[ \Phi_t := \mathbb{E}^{\mathbb{P}_0}[\rho_t | \mathcal{F}_t] = \int \exp \left( x\xi_t - \frac{1}{2}x^2t \right) \pi(dx). \] (6.6)

By use of the Ito calculus together with (6.5), we deduce that
\[ d\Phi_t = \hat{X}_t \Phi_t d\xi_t, \] (6.7)
and thus by integration we obtain
\[ \Phi_t = \exp \left( \int_0^t \hat{X}_s d\xi_s - \frac{1}{2} \int_0^t \hat{X}_s^2 ds \right). \] (6.8)

Since \( \{\xi_t\} \) is an \( (\{\mathcal{F}_t\}, \mathbb{P}_0)\)-Brownian motion, it follows from (6.8) by the Girsanov theorem that the process \( \{M_t\} \) defined by
\[ \xi_t = \int_0^t \hat{X}_s ds + M_t \] (6.9)
is an \( (\{\mathcal{F}_t\}, \mathbb{P})\)-Brownian motion, which we call the innovations process (see, e.g., Heunis 2011). This gives us the innovations representation for the information process in the Brownian case, in which the increments of \( \{M_t\} \) represent the arrival of new information.

**Remark 6.2.1. (Brownian bridge information processes)** 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. One way of modelling such situations 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 5.3.1 shows that the process \( \{\xi_{tT}\} \) defined over the time interval \( 0 \leq t < T \) by
\[ \xi_{tT} = \frac{T - t}{T} \xi \left( \frac{tT}{T-t} \right) \] (6.10)
Chapter 6. Examples of Lévy Information Processes

reveals the value of $X$ in the limit as $t \to T$, and one can check for $0 \leq s \leq t < T$ that

$$\text{Cov}[\xi_tT, \xi_{sT} | \mathcal{F}^X] = \frac{s(T - t)}{T} \psi_0''(X).$$  \hfill (6.11)

In the case where $\{\xi_t\}$ is a Brownian information process represented as above in the form

$$\xi_t = Xt + B_t,$$  \hfill (6.12)

the time-changed process (6.10) takes the form

$$\xi_{tT} = Xt + \beta_{tT},$$  \hfill (6.13)

where $\{\beta_{tT}\}$ is a Brownian bridge over the interval $[0, T]$. Such Brownian bridge information processes have had applications both in physics (Brody & Hughston 2005, 2006) and in finance (Macrina 2006, 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 (6.10) proposed above may be similarly applicable.

6.2.3 Mutual Brownian information

Further insight can be gained if we study the behaviour of the mutual information $I(\xi_t, X)$ between $\xi_t$ and $X$. This follows from the fact that $I(\xi_t, X)$ determines the amount of information contained in $\xi_t$ about the true value of the signal $X$ (see Gel’fand & Yaglom 1957 for a similar analysis involving Gaussian processes). The discussion in the previous chapter shows that the amount of uncertainty at time zero regarding the value of $X$ is given by the entropy associated with the a priori density of $X$. As time progresses, however, the observation of the information process removes the uncertainty. The mutual information as a function of time should thus indicate some form of ‘learning curve’. As an example, let us here consider the case where the random variable $X$ takes discrete values $\{x_i\}_{i=1}^n$ with probabilities $\{p_i\}_{i=1}^n$. Then the mutual information is given by

$$I(\xi_t, X) = \sum_{i=1}^n \int_0^\infty \rho_{\xi X}(\xi, i) \ln \left[ \frac{\rho_{\xi X}(\xi, i)}{\rho_\xi(\xi) \rho_X(i)} \right] d\xi,$$  \hfill (6.14)

where for clarity we have written $\rho_X(i) = p_i$. Here, $\rho_{\xi X}(\xi, i)$ denotes the joint density of the random variable pair $(\xi_t, X)$, and $\rho_\xi(\xi)$ is the marginal density for $\xi_t$. To work
out the joint density function we observe that

\[ P(\xi_t \leq \xi \cap X = x_i) = P(\xi_t \leq \xi | X = x_i) P(X = x_i) \]
\[ = p_i P(x_i + B_t \leq \xi) \]
\[ = p_i P(B_t \leq \xi - x_i t) \]
\[ = p_i F_{B_t}(\xi - x_i t), \quad (6.15) \]

where \( F_{B_t} \) denotes the cumulative density function of Brownian motion, and where \( P(X = x_i) = p_i \). Hence, we deduce that

\[ \rho_{\xi X}(\xi, i) = p_i dF_{B_t}(\xi - x_i t) \]
\[ = \frac{p_i}{\sqrt{2\pi t}} \exp \left[ -\frac{(\xi - x_i t)^2}{2t} \right]. \quad (6.16) \]

It follows that the marginal density function \( \rho_{\xi}(\xi) \) is given by

\[ \rho_{\xi}(\xi) = \frac{1}{\sqrt{2\pi t}} \sum_{i=1}^{n} p_i \exp \left[ -\frac{(\xi - x_i t)^2}{2t} \right]. \quad (6.17) \]

With these expressions at hand, we find

\[ \frac{\rho_{\xi X}(\xi, i)}{\rho_{\xi}(\xi) \rho_X(i)} = \frac{\exp \left[ -\frac{(\xi - x_i t)^2}{2t} \right]}{\sum_{i=1}^{n} p_i \exp \left[ -\frac{(\xi - x_i t)^2}{2t} \right]}. \quad (6.18) \]

Hence, the mutual information can be expressed as follows:

\[ I(\xi_t, X) = \sum_{i=1}^{n} \int_{0}^{\infty} \rho_{\xi X}(\xi, i) \ln \left( \frac{\rho_{\xi X}(\xi, i)}{\rho_{\xi}(\xi) \rho_X(i)} \right) d\xi \]
\[ = \sum_{i=1}^{n} \int_{0}^{\infty} \frac{p_i}{\sqrt{2\pi t}} \exp \left[ -\frac{(\xi - x_i t)^2}{2t} \right] \ln \left( \frac{\exp \left[ -\frac{(\xi - x_i t)^2}{2t} \right]}{\sum_{i=1}^{n} p_i \exp \left[ -\frac{(\xi - x_i t)^2}{2t} \right]} \right) d\xi \]
\[ = \frac{1}{\sqrt{2\pi t}} \int_{0}^{\infty} E \left[ (\xi - X t)^2 \right] \exp \left[ -\frac{(\xi - X t)^2}{2t} \right] \ln \left( \frac{\exp \left[ -\frac{(\xi - X t)^2}{2t} \right]}{\sum_{i=1}^{n} p_i \exp \left[ -\frac{(\xi - x_i t)^2}{2t} \right]} \right) d\xi \]
\[ - \frac{1}{\sqrt{2\pi t}} \int_{0}^{\infty} E \left[ \exp \left[ -\frac{(\xi - X t)^2}{2t} \right] \ln \left( \exp \left[ -\frac{(\xi - X t)^2}{2t} \right] \right) \right] d\xi. \quad (6.19) \]

A plot of the mutual information arising in the case of a Brownian information process and the associated signal \( X \) is shown in figure 6.1, which indicates how accumulated information increases as time goes by.
6.3 Poisson information process

6.3.1 Definition and properties

The Poisson process is the simplest of all Lévy processes entailing jumps. The characteristic exponent of a Poisson process with rate parameter \( m \) is given by

\[
\phi(\alpha, m) = \exp \left( m(e^{i\alpha} - 1) \right).
\]  

(6.20)

A Poisson process \( \{N_t\} \) takes nonnegative integer values \( N = \{0, 1, \ldots\} \), whose jump size obeys the probability law:

\[
P(N_{t+\Delta} - N_t = n) = \exp(-m\Delta)(m\Delta)^n/n!.
\]

(6.21)

Evidently, \( \{N_t\} \) has independent and stationary increments. Like Brownian motion, the Poisson process is a prototype model for characterising empirical observations seen both
in society and in nature, such as the time of arrival of telephone calls at a switchboard, or the time of particle emissions in radioactive decay.

**Remark 6.3.1.** The formula for what is now known as the Poisson distribution was introduced by Simeon Denis Poisson (1781-1840) in 1838. The formula arises more or less as a by-product of Poisson’s discussion of the law of large numbers in his book *Recherches sur la probabilité des jugements en matière criminelle et en matière civile* (“Research on the Probability of Judgements in Criminal and Civil Matters”) published in 1837. In the chapter on the law of large numbers, he points out that the probability of the number of happenings of an event in trials will fall within certain assigned limits. More specifically, the Poisson distribution can be recovered from the binomial distribution in the limit where the number of trial goes to infinity, while the expected number of successes is held fixed. This observation, often referred to as the law of small numbers, or law of rare events, as it refers to events that occur rarely but at the same time have many opportunities to occur, was novel, and was developed further in the book *Das Gesetz der kleinen Zahlen* (“The Law of Small Numbers”) by Ladislaus von Bortkiewicz published in 1898.

### 6.3.2 Poisson information

Let us now turn to the signal detection problem involving Poisson noise. Consider a situation in which an agent observes a series of events taking place at an unknown random rate, and the agent wishes to determine this unknown rate as best as possible since its value conveys an important piece of information. One can model the information flow in this example by a modulated Poisson process for which the jump rate is itself 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 for treating problems of this kind can be seen to emerge rather naturally as an example of our general model for Lévy information.

As in the Brownian case, one can construct the relevant information process directly. The setup is as follows. On a probability space \((\Omega, \mathcal{F}, P)\), 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

\[
\xi_t = N(e^X t).
\]

(6.22)

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 5.2.1. In particular, we have

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

and for fixed \( X \) one obtains a Poisson process with rate \( m e^X \). It follows that (6.22) is an information process. The fiducial characteristic is given by \( K_0 = (0, 0, me^X \delta_1(dz)), \) that of a Poisson process with unit jumps at the rate \( m \), where \( \delta_1(dz) \) is the Dirac measure with unit mass at \( z = 1 \), and the fiducial exponent is

\[
\psi_0(\alpha) = m(e^\alpha - 1).
\]

A calculation using (5.29) shows that \( K_X = (0, 0, me^X \delta_1(dz)), \) and that

\[
\psi_X(\alpha) = me^X (e^\alpha - 1).
\]

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

\[
F^X_t(y) := \mathbb{P}(X \leq y \mid \mathcal{F}_t) = \mathbb{P}(X \leq y \mid \xi_t)
\]

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

\[
\mathbb{P}(X \leq y \mid \xi_t = n) = \frac{\int \mathbb{I}\{x \leq y\} \mathbb{P}(\xi_t = n \mid X = x) dF^X(x)}{\int \mathbb{P}(\xi_t = n \mid X = x) dF^X(x)}.
\]

In the case of Poisson information the relevant conditional distribution is

\[
\mathbb{P}(\xi_t = n \mid X = x) = \exp(-mte^x) \frac{(mte^x)^n}{n!}.
\]

After some cancellation we deduce that

\[
\mathbb{P}(X \leq y \mid \xi_t = n) = \frac{\int \mathbb{I}\{x \leq y\} \exp(xn - m(e^x - 1)t) dF^X(x)}{\int \exp(xn - m(e^x - 1)t) dF^X(x)}.
\]
and hence

\[
F_t^X(y) = \int \mathbb{1}_{\{x \leq y\}} \exp(x\xi_t - m(e^x - 1)t) \, dF^X(x) \quad \left( \frac{\text{integration}}{\int \exp(x\xi_t - m(e^x - 1)t) \, dF^X(x)} \right)
\]

and thus

\[
\pi_t(dx) = \frac{\exp(x\xi_t - m(e^x - 1)t)}{\int \exp(x\xi_t - m(e^x - 1)t) \, \pi(dx)} \pi(dx),
\]

which we can see is consistent with (5.56) if we recall that in the case of noise of the Poisson type the fiducial exponent is given by \( \psi_0(\alpha) = m(e^\alpha - 1) \).

### 6.3.3 Mutual Poisson information

For the purpose of determining the mutual information \( I(\xi_t, X) \) between the observation \( \xi_t \) and the signal \( X \) we need to work out expressions for the joint and marginal densities. As in the previous example, let us assume that \( X \) is a discrete random variables taking the values \( \{x_i\} \) with probabilities \( \{p_i\} \). It turns out that the calculation simplifies slightly if we instead compute the mutual information \( I(\xi_t, Y) \) between \( \xi_t = N_{Y_t} \) and \( Y = e^X \). Now if

\[
\rho_N(k) = \frac{e^{-mt} (mt)^k}{k!}
\]

is the density function for the Poisson process with rate \( m \), then the joint density is given by

\[
\rho_{\xi Y}(\xi, i) = \frac{p_i}{\xi} \rho_N(\xi) e^{-mt y_i} (mt y_i)^\xi.
\]

Here we have written \( y_i = e^{x_i} \) for the values of \( Y \). The marginal density for \( \xi_t \) is thus given by

\[
\rho_\xi(\xi) = \sum_{i=1}^n \frac{p_i}{\xi!} e^{-mt y_i} (mt y_i)^\xi.
\]

It follows that

\[
\frac{\rho_{\xi Y}(\xi, i)}{\rho_\xi(\xi) \rho_Y(i)} = \frac{e^{-mt y_i} (mt y_i)^\xi}{\sum_{i=1}^n p_i e^{-mt y_i} (mt y_i)^\xi}.
\]
Chapter 6. Examples of Lévy Information Processes

Figure 6.2: Mutual information in the case of Poisson noise. The mutual information $I$ measures the amount of information contained in the observation about the value of the unknown signal $X$. At time zero, no data is available so that the accumulated information content is zero. As time progresses, data is forthcoming that enhances the knowledge of the observer. Eventually, a sufficient amount of information, equivalent to the amount of the initial uncertainty $- \sum_i p_i \ln p_i$, is gathered, at which point the value of $X$ is revealed. Strictly speaking this happens asymptotically as $t \to \infty$, although for all practical purposes the value of $X$ will be revealed with high confidence level after a passage of finite amount of time. In this example, the parameters are chosen to be $x_1 = 1$, $x_2 = 2$, $p_1 = 0.4$, $p_2 = 0.6$, and $T = 50$. The three plots corresponds to $m = 1$, $m = 1.3$ and $m = 1.5$. The initial entropy (asymptotic value of $I$) in this example is approximately 0.67.

Putting these together into the formula for the mutual information, we find:

$$I(\xi_t, X) = \sum_{i=1}^{\infty} \sum_{\xi=0}^{\infty} \rho_{\xi Y}(\xi, i) \ln \left[ \frac{\rho_{\xi Y}(\xi, i)}{\rho_{\xi}(\xi)\rho_Y(i)} \right]$$

$$= \sum_{i=1}^{\infty} \sum_{\xi=0}^{\infty} \frac{p_i}{\xi!} e^{-mt_y(i)} \xi \ln \left[ e^{-mt_y(i)\xi} \xi \sum_{i=1}^{\infty} p_i e^{-mt_y(i)\xi} \rho_Y(i) \right]$$

$$= \sum_{\xi=0}^{\infty} \frac{(mt)^\xi}{\xi!} \left[ \sum_{i=1}^{\infty} p_i e^{-mt_y(i)^\xi} \xi \ln \left( e^{-mt_y(i)^\xi} \xi \right) - \ln \mathbb{E} \left( e^{-mt_YY^\xi} \right) \right]$$

$$= \sum_{\xi=0}^{\infty} \frac{(mt)^\xi}{\xi!} \left[ \mathbb{E} \left( e^{-mt_Y^\xi} \ln \left( e^{-mt_Y^\xi} \right) \right) - \mathbb{E} \left( e^{-mt_Y^\xi} \ln \mathbb{E} \left( e^{-mt_Y^\xi} \right) \right) \right].$$

(6.36)

In figure 6.2 we plot the “learning curve” for the information accumulation in the case of the Poisson information process, for a range of parameter values.
6.4 Gamma information process

6.4.1 Definition and properties

It will be convenient first to recall some definitions and conventions (cf. Yor 2007, Brody et al. 2008b, Brody et al. 2012) relating to gamma processes. Let \( m \) and \( \kappa \) be positive numbers. By a gamma process with rate \( m \) and scale \( \kappa \) on a probability space \((\Omega, \mathcal{F}, P)\) we mean a Lévy process \( \{\gamma_t\}_{t \geq 0} \) with exponent

\[
\psi_0(\alpha) = \frac{1}{t} \ln \mathbb{E}^P [\exp(\alpha \gamma_t)] = -m \ln(1 - \kappa \alpha) \tag{6.37}
\]

for \( \alpha \in A_C = \{ w \in \mathbb{C} | \text{Re} \ w < \kappa^{-1}\} \). The probability density for \( \gamma_t \) is

\[
\mathbb{P}(\gamma_t \in dx) = \mathbb{1}\{x > 0\} \kappa^{-mt} x^{mt-1} \exp(-x/\kappa) \frac{\exp}{\Gamma(mt)} \, dx, \tag{6.38}
\]

where \( \Gamma[a] \) is the gamma function. A short calculation making use of the functional equation \( \Gamma[a + 1] = a \Gamma[a] \) shows that

\[
\mathbb{E}^P [\gamma_t] = m \kappa t \quad \text{and} \quad \text{Var}^P [\gamma_t] = m \kappa^2 t. \tag{6.39}
\]

Clearly, the mean and variance determine the rate and scale. The Lévy measure in this example is given by

\[
\nu(dz) = \mathbb{1}\{z > 0\} m z^{-1} \exp(-\kappa z) \, dz. \tag{6.40}
\]

One can check that \( \nu(\mathbb{R}\setminus\{0\}) = \infty \) and thus that the gamma process has infinite activity. If \( \kappa = 1 \) we say that \( \{\gamma_t\} \) is a standard gamma process with rate \( m \), and in that case one finds that \( \{\kappa \gamma_t\} \) is a scaled gamma process with rate \( m \) and scale \( \kappa \).

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 \( \{\rho^\lambda_t\} \) defined by

\[
\rho^\lambda_t = (1 - \lambda)^{mt} e^{\lambda \gamma_t} \tag{6.41}
\]

is an \((\mathcal{F}_t, \mathbb{P}_0)\)-martingale. If we let \( \{\rho^\lambda_t\} \) act as a change of measure density for the transformation \( \mathbb{P}_0 \rightarrow \mathbb{P}_\lambda \), 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, we can proceed and establish results about gamma information processes.
6.4.2 Gamma information

We begin by stating the following result on the characterisation of the gamma information:

**Proposition 6.4.1.** 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

\[
\xi_t = \frac{1}{1 - X} \gamma_t
\]

is a Lévy information process with message \( X \) and gamma noise, with fiducial exponent

\[
\psi_0(\alpha) = -m \ln(1 - \alpha)
\]

for \( \alpha \in \{ w \in \mathbb{C} \mid \text{Re } w < 1 \} \).

**Proof.** It is evident that \( \{\xi_t\} \) is \( \mathcal{F}_X \)-conditionally a scaled gamma process. As a consequence of (6.37) we have

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

for \( \alpha \in \mathbb{C}^I \). Then we note that

\[
-m \ln \left( 1 - \frac{\alpha}{1 - X} \right) = -m \ln (1 - (X + \alpha)) + m \ln (1 - X),
\]

from which it follows that the \( \mathcal{F}_X \)-conditional \( \mathbb{P} \) exponent of \( \{\xi_t\} \) is \( \psi_0(X + \alpha) - \psi_0(X) \).

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 unknown rate at which the process is growing on average is an important figure that the agent wishes to determine as accurately as possible. The accumulation process can be modelled by gamma information, and the associated filter can be utilised to estimate the growth rate. It has long been recognised that the gamma process is useful in characterising 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, Hughston & Macrina (2008b), Hoyle (2010), and Hoyle, Hughston & Macrina (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 the remarkable property that $\gamma_t$ and $\gamma_s/\gamma_t$ 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 in the case of gamma information. We begin by establishing that the process defined by (6.42) 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 random variables $\gamma_{s_1}/\gamma_s, \gamma_{s_2}/\gamma_{s_1}, \ldots$ are independent of one another and are independent of $\gamma_s$ and $\gamma_t$. It follows that

$$P(\xi_t < a | \xi_s, \xi_{s_1}, \ldots, \xi_{s_k}) = P(\xi_t < a | (1 - X)^{-1} \gamma_s, (1 - X)^{-1} \gamma_{s_1}, \ldots, (1 - X)^{-1} \gamma_{s_k})$$

$$= P(\xi_t < a | (1 - X)^{-1} \gamma_s, \gamma_{s_1}/\gamma_s, \gamma_{s_2}/\gamma_{s_1}, \ldots, \gamma_{s_k}/\gamma_{s_{k-1}})$$

$$= P(\xi_t < a | (1 - X)^{-1} \gamma_s)$$

$$= P(\xi_t < a | \xi_s), \quad (6.46)$$

since $\\{\gamma_t\}$ and $X$ are independent, and that gives us the Markov property (5.58). In working out the conditional distribution of $X$ given $\mathcal{F}_t$ it suffices therefore to work out the conditional distribution of $X$ given $\xi_t$, since $X$ is $\mathcal{F}_\infty$-measurable. We note that the Bayes formula implies that

$$\pi_t(dx) = \frac{\rho(\xi_t | X = x)}{\int \rho(\xi_t | X = x) \pi(dx)} \pi(dx), \quad (6.47)$$

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

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

$$= \frac{d}{d\xi} P(\gamma_t \leq (1 - X)\xi | X = x) = \frac{\xi^m(t-1)(1-x)^m \exp(-(1-x)\xi)}{\Gamma[mt]}. \quad (6.48)$$

Therefore, we deduce that

$$\pi_t(dx) = \frac{1}{\int_{-\infty}^{1-x}(1-x)^m \exp(x\xi_t) \pi(dx)} \pi(dx), \quad (6.49)$$

and this gives us the optimal filter for the case of the gamma information process.

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 $\xi_t$. 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

$$P(U \in du) = 1 \{u > 0\} \frac{\theta^r e^{-\theta u} u^{r-1}}{\Gamma[r]} du, \quad (6.50)$$
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
\[
Y = \psi_0'(X) = \frac{m}{(1 - X)},
\]
and set
\[
\tau = \frac{(r - 1)}{m}.
\]
Then the optimal filter for \( Y \) is given by
\[
\hat{Y}_t := E_P[Y | F_t] = \xi_t + \theta t + \tau,
\]
which indeed is a linear function of the observation.

### 6.4.3 Mutual gamma information

As in the previous examples, for working out the mutual information in the gamma information model, we need to determine the joint and marginal densities. Let
\[
\rho_\gamma(y) = \frac{y^{(mt-1)}e^{-y}}{\Gamma[mt]},
\]
be the density function for the gamma process with rate \( m \). We also have
\[
\mathbb{P}(\xi_t \leq \xi \land X = x_i) = \mathbb{P}(\xi_t \leq \xi | X = x_i)\mathbb{P}(X = x_i)
= p_i\mathbb{P}(x_i \gamma_t \leq \xi)
= p_i\mathbb{P}\left(\gamma_t \leq \frac{\xi}{x_i}\right)
= p_i F_\gamma\left(\frac{\xi}{x_i}\right),
\]
where \( F_\gamma \) denotes the cumulative density function for the marginals of the \( \gamma_t \), and \( \mathbb{P}(X = x_i) = p_i \). Hence, we can calculate
\[
\rho_{\xi X}(\xi, i) = \frac{d}{d\xi} F_\gamma\left(\frac{\xi}{x_i}\right)
= p_i \frac{1}{x_i} \rho_\gamma\left(\frac{\xi}{x_i}\right)
= p_i \frac{\xi^{mt-1} x_i^{1-mt} e^{-\xi/x_i}}{\Gamma[mt]} \\
= \frac{p_i^{n-t} x_i^{1-mt} e^{-\xi/x_i}}{\Gamma[mt]}.
\]
For the marginal density, we thus obtain

\[
\rho_{\xi}(\xi) = \sum_{i=1}^{n} \rho_{\xi X}(\xi, i)
\]

\[
= \sum_{i=1}^{n} \frac{p_{i} e^{\xi(m-1)} x_{i}^{m-1} e^{-\xi/x_{i}}}{\Gamma[mt]}
\]

\[
= \frac{\xi^{m-1}}{\Gamma[mt]} \sum_{i=1}^{n} p_{i} x_{i}^{m-1} e^{-\xi/x_{i}}.
\]

Putting these together, we find

\[
I(\xi, X) = \sum_{i=1}^{n} \int_{0}^{\infty} \frac{p_{i} e^{\xi(m-1)} x_{i}^{m-1} e^{-\xi/x_{i}}}{\Gamma[mt]} \left[ -mt \ln x_{i} - \frac{\xi}{x_{i}} - \ln \left( \sum_{i=1}^{n} p_{i} x_{i}^{m-1} e^{-\xi/x_{i}} \right) \right] d\xi
\]

\[
= -\frac{1}{\Gamma[mt]} \sum_{i=1}^{n} \int_{0}^{\infty} mtp_{i} e^{\xi(m-1)} x_{i}^{m-1} e^{-\xi/x_{i}} \ln x_{i} d\xi
\]

\[
= -\frac{1}{\Gamma[mt]} \sum_{i=1}^{n} \int_{0}^{\infty} p_{i} e^{\xi(m-1)} x_{i}^{m-1} e^{-\xi/x_{i}} d\xi
\]

\[
= -\frac{1}{\Gamma[mt]} \sum_{i=1}^{n} \int_{0}^{\infty} p_{i} x_{i}^{m-1} e^{-\xi/x_{i}} \ln \left( \sum_{i=1}^{n} p_{i} x_{i}^{m-1} e^{-\xi/x_{i}} \right) d\xi.
\]

We now observe that

\[
-\frac{1}{\Gamma[mt]} \sum_{i=1}^{n} mtp_{i} x_{i}^{m-1} \ln x_{i} \int_{0}^{\infty} \xi^{m-1} e^{-\xi/x_{i}} d\xi
\]

\[
= -\frac{1}{\Gamma[mt]} \sum_{i=1}^{n} mtp_{i} x_{i}^{m-1} \ln x_{i} \Gamma[mt] x_{i}^{mt}
\]

\[
= -mt \sum_{i=1}^{n} p_{i} \ln x_{i}
\]

\[
= -mt \mathbb{E} \left[ \ln X \right],
\]

and that

\[
-\frac{1}{\Gamma[mt]} \sum_{i=1}^{n} p_{i} x_{i}^{-1-mt} \int_{0}^{\infty} \xi^{mt} e^{-\xi/x_{i}} d\xi
\]

\[
= -\frac{1}{\Gamma[mt]} \sum_{i=1}^{n} p_{i} x_{i}^{-1-mt} \Gamma[mt+1] x_{i}^{mt+1}
\]

\[
= -mt.
\]
Chapter 6. Examples of Lévy Information Processes

Figure 6.3: Mutual information in the case of the gamma noise. The mutual information $I$ measures the amount of information contained in the observation about the value of the unknown signal $X$. At time zero, no data is available so that the accumulated information content is zero. However, as time progresses, data is forthcoming that enhances the knowledge of the observer. Eventually, a sufficient amount of information, equivalent to the amount of the initial uncertainty $-\sum p_i \ln p_i$, is gathered, at which point the value of $X$ is revealed. Strictly speaking this happens asymptotically as $t \to \infty$, although for all practical purposes the value of $X$ will be revealed with high confidence level after a passage of finite amount of time. In this example, the parameters are chosen to be $x_1 = 1$, $x_2 = 2$, $p_1 = 0.4$, $p_2 = 0.6$, and $T = 50$. The three plots corresponds to $m = 1$, $m = 2$ and $m = 3$. The initial entropy (asymptotic value of $I$) in this example is approximately 0.67.

Furthermore, the last term in the right side of (6.58) is

\[
-\frac{1}{\Gamma(mt)} \sum_{i=1}^{n} \int_{0}^{\infty} p_i \xi^{mt-1} x_i^{-mt} e^{-\xi/x_i} \ln \mathbb{E} \left[ X^{-mt} e^{-\xi/X} \right] d\xi
\]

\[
= -\frac{1}{\Gamma(mt)} \sum_{i=1}^{n} p_i \int_{0}^{\infty} \xi^{mt-1} x_i^{-mt} e^{-\xi/x_i} \ln \mathbb{E} \left[ X^{-mt} e^{-\xi/X} \right] d\xi
\]

\[
= -\frac{1}{\Gamma(mt)} \int_{0}^{\infty} \xi^{mt-1} \mathbb{E} \left[ X^{-mt} e^{-\xi/X} \right] \ln \mathbb{E} \left[ X^{-mt} e^{-\xi/X} \right] d\xi. \tag{6.61}
\]

Hence, the expression for mutual information is

\[
I(\xi, X) = -mt \left( 1 + \mathbb{E} \left[ \ln X \right] \right) - \frac{\mathbb{E} \left[ \int_{0}^{\infty} \xi^{mt-1} X^{-mt} e^{-\xi/X} \ln \mathbb{E} \left[ X^{-mt} e^{-\xi/X} \right] d\xi \right]}{\Gamma(mt)}. \tag{6.62}
\]

In figure 6.3 we plot the mutual information $I(\xi, X)$ as a function of $t \in [0, 50]$ for three values of the information flow rate parameter $m$. The information gained by the market
Figure 6.4: Mutual information comparison. The time dependence of the mutual information $I$ is compared for three types of information processes: Brownian, Poisson, and gamma. The common parameters are chosen to be $x_1 = 1$, $x_2 = 2$, $p_1 = 0.4$, $p_2 = 0.6$, and $T = 50$. Since other parameters embody different meanings, a direct comparison as shown here need not reveal quantitative information. Nevertheless, the intuition one gains from the figure is that the revelation of the modulation signal in the Poisson information is somewhat slower than that in the drift for the Brownian information and the scale for the gamma information.

participants increases more rapidly as $m$ increases, as one might expect.

Remark 6.4.1. In figure 6.4, we compare the mutual information in the Brownian, Poisson and gamma examples. In particular, the growth rate parameter $m$ appearing in the Poisson information process and the gamma information process is set to unity in these plots. The results indicate that the learning curve grows significantly more rapidly in the case of a gamma information, which might be due to the existence of infinite activity. The learning curve is even more steep for Brownian information, although there is no direct comparison we can make since the nature of the process is quite different.

6.5 Variance-gamma information process

6.5.1 Definition and properties

The so-called variance-gamma or VG process (Madan & Seneta 1990, Madan & Milne 1991, Madan, Carr & Chang 1998, Carr, Geman, Madan & Yor 2002) was introduced in
the theory of finance. The relevant definitions and conventions are as follows. By a VG process \( V_t \) with drift \( \mu \in \mathbb{R} \), volatility \( \sigma \geq 0 \), and rate \( m > 0 \), we mean a Lévy process with exponent

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

(6.63)

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

\[
\Gamma_t = \frac{\gamma_t}{m}
\]

(6.64)

a gamma subordinator with rate \( m \). Note that \( \Gamma_t \) has dimensions of time and that \( \mathbb{E}^\mathbb{P}[\Gamma_t] = t \). A calculation shows that the Lévy process \( \{V_t\} \) defined by

\[
V_t = \mu \Gamma_t + \sigma B_{\Gamma_t}
\]

(6.65)

has the exponent (6.63). The VG process thus takes the form of a Brownian motion with drift, time-changed by use of 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 physical applications, years for economic applications), and treat \( m \) as a model parameter. In the limiting case \( \sigma \to 0 \) we obtain a gamma process with rate parameter \( m \) and scale parameter \( \mu/m \). In the limiting case \( m \to \infty \) we obtain a Brownian motion with drift \( \mu \) and volatility \( \sigma \).

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

\[
V_t = \kappa_1 \gamma^1_t - \kappa_2 \gamma^2_t,
\]

(6.66)

where \( \kappa_1 \) and \( \kappa_2 \) are nonnegative constants. A calculation shows that the associated characteristic exponent is of the form (6.63). In particular, we have

\[
\psi(\alpha) = -m \ln \left( 1 - (\kappa_1 - \kappa_2) \alpha - \kappa_1 \kappa_2 \alpha^2 \right),
\]

(6.67)
where $\mu = m(\kappa_1 - \kappa_2)$ and $\sigma^2 = 2m\kappa_1\kappa_2$, or equivalently
\[
\kappa_1 = \frac{1}{2m} \left( \mu + \sqrt{\mu^2 + 2m\sigma^2} \right) \quad \text{and} \quad \kappa_2 = \frac{1}{2m} \left( -\mu + \sqrt{\mu^2 + 2m\sigma^2} \right), \tag{6.68}
\]
where $\alpha \in \{ w \in \mathbb{C} : -1/\kappa_2 < \text{Re} \, w < 1/\kappa_1 \}$.

Now let $\{\xi_t\}$ be a standard VG process on a probability space $(\Omega, \mathcal{F}, P_0)$, with exponent
\[
\psi_0(\alpha) = -m \ln(1 - (2m)^{-1} \alpha^2) \tag{6.69}
\]
for $\alpha \in \{ w \in \mathbb{C} : |\text{Re} \, w| < \sqrt{2m} \}$. Under the transformed measure $P_\lambda$ defined by the change-of-measure martingale (5.24), one finds that $\{\xi_t\}$ is a drifted VG process, with
\[
\mu = \lambda \left( 1 - \frac{1}{2m} \lambda^2 \right)^{-1} \quad \text{and} \quad \sigma = \left( 1 - \frac{1}{2m} \lambda^2 \right)^{-\frac{1}{2}}, \tag{6.70}
\]
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.

In the parametrisation used by Carr, Geman, Madan & Yor (2002) the Lévy measure of the VG process takes the form
\[
\nu(dx) = C \exp(Gx) \begin{cases} 
(\frac{-x}{-x^2})^{-1-Y} \, dx & \text{if } x < 0, \\
(\frac{-x}{x^2})^{-1-Y} \, dx & \text{if } x > 0,
\end{cases}
\]
where $C = m > 0$, $G = \kappa_2^{-1} > 0$, $M = \kappa_1^{-1} > 0$, and $Y < 2$.

**Remark 6.5.1.** The Lévy measure has infinite mass. Thus, a VG process has infinitely many jumps in any finite time interval. Since
\[
\int_{-1}^{1} |x| \nu(dx) < \infty, \tag{6.71}
\]
a VG process has paths of finite variation.

The mean and variance of $V_t$ are given by:
\[
\mathbb{E}^{P_0}[V_t] = \mu \quad \text{and} \quad \text{Var}^{P_0}[V_t] = \sigma^2 + \frac{\mu^2}{m}, \tag{6.72}
\]
6.5.2 Variance-gamma information

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 6.5.1.** Let \(\{\Gamma_t\}\) be a standard 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. Then the process \(\{\xi_t\}\) defined by

\[
\xi_t = X \left(1 - \frac{1}{2m} X^2\right)^{-1} \Gamma_t + \left(1 - \frac{1}{2m} X^2\right)^{-\frac{1}{2}} B(\Gamma_t)
\]

(6.73)

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

\[
\psi_0(\alpha) = -m \ln \left(1 - \frac{1}{2m} \alpha^2\right)
\]

(6.74)

for \(\alpha \in \{w \in \mathbb{C} : \text{Re } w < \sqrt{2m}\}\).

*Proof.* We observe that \(\{\xi_t\}\) is \(\mathcal{F}_X\)-conditionally a drifted VG process of the form

\[
\xi_t = \mu_X \Gamma_t + \sigma_X B(\Gamma_t),
\]

(6.75)

where the drift and volatility coefficients are

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

(6.76)

The \(\mathcal{F}_X\)-conditional \(\mathbb{P}\)-exponent of \(\{\xi_t\}\) is by virtue of (6.63) thus given for \(\alpha \in \mathbb{C}\) by

\[
\psi_X(\alpha) = -m \ln \left(1 - \frac{1}{m} \mu_X \alpha - \frac{1}{2m} \sigma_X^2 \alpha^2\right)
\]

\[
= -m \ln \left(1 - \frac{1}{m} X \left(1 - \frac{1}{2m} X^2\right)^{-1} \alpha - \frac{1}{2m} \left(1 - \frac{1}{2m} X^2\right)^{-1} \alpha^2\right)
\]

\[
= -m \ln \left(1 - \frac{1}{2m} (X + \alpha)^2\right) + m \ln \left(1 - \frac{1}{2m} X^2\right),
\]

(6.77)

which, by (6.74), is evidently of the form \(\psi_0(X + \alpha) - \psi_0(X)\), as required.

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 6.5.2.** 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 $\{\Gamma^X_t\}$ for the subordinator defined by

$$\Gamma^X_t = \left(1 - \frac{1}{2m} X^2\right)^{-1} \Gamma_t.$$  \hfill (6.78)

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

$$\xi_t = X \Gamma^X_t + B(\Gamma^X_t) \hfill (6.79)$$

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 6.5.3.** Let $\{\gamma^1_t\}$ and $\{\gamma^2_t\}$ 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 $\{\xi_t\} \hfill (6.80)$

is a VG information process with message $X$.

### 6.6 Negative-binomial information process

#### 6.6.1 Definition and properties

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

$$\psi_0(\alpha) = m \ln \left(\frac{1 - q}{1 - q e^{\alpha}}\right) \hfill (6.81)$$

for $\alpha \in \{w \in \mathbb{C} \mid \text{Re} \ w < -\ln q\}$.

There are two representations for the negative binomial process (Kozubowski & Podgórski 2009, Brody, Hughston & Mackie 2012). The first of these is a compound Poisson process for which the jump size $J \in \mathbb{N}$ has a logarithmic distribution

$$\mathbb{P}_0(J = n) = \frac{1}{\ln(1-q)} \frac{1}{n} q^n, \hfill (6.82)$$
and the intensity of the Poisson process determining the timing of the jumps is given by
\(-m \ln(1 - q)\). One finds that the characteristic function of \(J\) is
\[
\phi_0(\alpha) := \mathbb{E}^{P_0}[\exp(\alpha J)] = \frac{\ln(1 - q e^\alpha)}{\ln(1 - q)}
\] (6.83)
for \(\alpha \in \{w \in \mathbb{C} | \text{Re } w < -\ln q\}\). Then if we set
\[
n_t = \sum_{k=1}^{\infty} 1\{k \leq N_t\} J_k,
\] (6.84)
where \(\{N_t\}\) is a Poisson process with rate \(-m \ln(1 - q)\) and \(\{J_k\}_{k \in \mathbb{N}}\) denotes a collection
of independent identical copies of \(J\), representing the jumps, a calculation shows that
\[
P_0(n_t = k) = \frac{\Gamma(k + mt)}{\Gamma(mt)\Gamma(k + 1)} q^k (1 - q)^{mt},
\] (6.85)
and that the resulting exponent is given by (6.81).

The second representation of the negative binomial process makes use of the method of
subordination. We take a Poisson process with rate \(\Lambda = mq / (1 - q)\) (6.86)
and time-change it using a gamma subordinator \(\{\Gamma_t\}\) with rate parameter \(m\). The
moment generating function thus obtained, in agreement with (6.81), is
\[
\mathbb{E}^{P_0} [\exp (\alpha N(\Gamma_t))] = \mathbb{E}^{P_0} [\exp (\Lambda(e^\alpha - 1)\Gamma_t)] = \left(\frac{1 - q}{1 - q e^\alpha}\right)^{mt}.
\] (6.87)

**Remark 6.6.1.** This distribution is sometimes referred to as the *Pascal distribution*,
named after the French mathematician and philosopher Blaise Pascal (1623-1662).

### 6.6.2 Negative-binomial information

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

**Proposition 6.6.1.** 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
\[
\Gamma_t^X = \left(\frac{q e^X}{1 - q e^X}\right) \Gamma_t.
\] (6.88)
Then the process \( \{ \xi_t \} \) defined by
\[
\xi_t = N(I_t^X)
\] (6.89)
is a Lévy information process with message \( X \) and negative binomial noise, with fiducial exponent (6.81).

**Proof.** This can be verified by direct calculation. For \( \alpha \in \mathbb{C}^I \) we have:
\[
\mathbb{E}^P \left[ e^{\alpha \xi_t} \middle| X \right] = \mathbb{E}^P \left[ \exp(\alpha N(I_t^X)) \middle| X \right] = \mathbb{E}^P \left[ \exp \left( m \frac{q e^{X}}{1 - q e^{X}}(e^{\alpha} - 1)\right) \right] \times
\]
\[
= \left( 1 - \frac{q e^{X}(e^{\alpha} - 1)}{1 - q e^{X}} \right)^{-mt} = \left( \frac{1 - q e^{X}}{1 - q e^{X+\alpha}} \right)^{mt},
\] (6.90)
which by (6.81) shows that the conditional exponent is of the form \( \psi_0(X+\alpha) - \psi_0(X) \). □

There is also a representation for negative binomial information based on the compound Poisson process. This can be obtained by an application of Proposition 5.6.1, 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
\[
\nu(dz) = m \sum_{n=1}^{\infty} \frac{1}{n} q^n \delta_n(dz),
\] (6.91)
where \( \delta_n(dz) \) denotes the Dirac measure with unit mass at the point \( z = n \). The Lévy measure is finite in this case, and we have
\[
\nu(\mathbb{R}) = -m \ln(1 - q),
\] (6.92)
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 (6.82) 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 5.6.1 we have the following:

**Proposition 6.6.2.** Let the random variable \( X \) satisfy \( X < -\ln q \) almost surely, let the random variable \( J^X \) have the conditional distribution
\[
\mathbb{P}(J^X = n \mid X) = \frac{1}{\ln(1 - q e^{X})} \frac{1}{n} (q e^{X})^n,
\] (6.93)
let \(\{J^X_k\}_{k \in \mathbb{N}}\) 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

\[
\xi_t = \sum_{k=1}^{\infty} \mathbb{1}\{k \leq N\left(-\ln(1 - qe^X)t\right)\} J^X_k
\]

is a Lévy information process with message \(X\) and negative binomial noise, with fiducial exponent (6.81).

### 6.7 Inverse Gaussian information process

#### 6.7.1 Definition and properties

We refer to Folks & Chhikara (1978) for a general review of the inverse Gaussian (IG) distribution and its applications. According to Folks & Chhikara, the IG distribution first appears in the work of Schrödinger (1915), in connection with the first passage time of Brownian motion, and the name “inverse Gaussian distribution” was coined by Tweedie (1945). Let us write \(\{G_t\}\) for an IG process, by which we mean a Lévy process that has independent and stationary IG increments. The IG process, which was introduced by Wasan (1968), can be characterised more precisely as follows (see, for example, Schoutens 2003, Kyprianou 2006). We fix a strictly positive “rate” parameter \(a\), and a strictly positive “drift” parameter \(b\). For each value of \(t \geq 0\), let the random variable \(G_t\) denote the first time a standard Brownian motion with positive drift given by \(\{B_u + bu\}_{u \geq 0}\) hits the level \(at\). Then we say that \(\{G_t\}_{t \geq 0}\) is an inverse Gaussian process with parameters \((a, b)\). The corresponding Lévy exponent is

\[
\psi_0(\alpha) = a \left( b - \sqrt{b^2 - 2\alpha} \right)
\]

for \(\alpha \in \{w \in \mathbb{C} | 0 \leq \text{Re } w < \frac{1}{2}b^2\}\). The probability density function for \(G_t\) is

\[
\mathbb{P}_0(G_t \in dx) = \mathbb{1}\{x > 0\} \frac{at}{\sqrt{2\pi x^3}} \exp\left(-\frac{(bx - at)^2}{2x}\right) dx,
\]

and the Lévy measure of the IG process is given by:

\[
\nu(dx) = \mathbb{1}\{x > 0\} \frac{a}{\sqrt{2\pi x^3}} \exp\left(-\frac{b^2x}{2}\right) dx.
\]

All of the positive and negative moments of \(G_t\) exist. These can be worked with the help of the relation

\[
\mathbb{E}_0^\mathbb{P}\left[G_t^{\gamma+1}\right] = \left(\frac{a}{b}t\right)^{2\gamma+1} \mathbb{E}_0^\mathbb{P}\left[G_t^{-\gamma}\right].
\]
which holds for all $\gamma \in \mathbb{R}$. The mean and the variance are given respectively by

$$E_{P_0}[G_t] = \frac{a}{b} t \quad \text{and} \quad \text{Var}_{P_0}[G_t] = \frac{a}{b^3} t. \quad (6.99)$$

The IG process has the following interesting scaling property. If the Lévy exponent of the IG process $\{G_t\}$ has the parameters $(a, b)$ as in (6.95) above, then for $c > 0$ the scaled process $\{cG_t\}$ is also an IG process, with parameters $(a\sqrt{c}, b/\sqrt{c})$.

It is straightforward to check that under the Esscher transformation $P_0 \to P_\lambda$ induced by (5.24), where $0 < \lambda < \frac{1}{2}b^2$, the parameter $a$ is left unchanged, whereas $b \to \sqrt{b^2 - 2\lambda}$.

### 6.7.2 Inverse Gaussian information

With these facts in mind we are in a position to introduce the associated information process. We fix a probability space $(\Omega, \mathcal{F}, P)$ and find the following:

**Proposition 6.7.1.** 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

$$Z = \frac{\sqrt{b^2 - 2X}}{b}. \quad (6.100)$$

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

$$\xi_t = Z^{-2}G(Zt), \quad (6.101)$$

is a Lévy information process with message $X$ and inverse Gaussian noise, with fiducial exponent (6.95).

**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 $\alpha \in C^1$ we have:

$$E^P[\exp(\alpha \xi_t)|X] = E^P\left[\exp\left(\alpha \frac{b^2}{b^2 - 2X} G\left(b^{-1}\sqrt{b^2 - 2X} t\right)\right)|X\right]$$

$$= \exp\left(at\left(\sqrt{b^2 - 2X} - \sqrt{b^2 - 2(\alpha + X)}\right)\right)$$

$$= \exp\left(at\left(b - \sqrt{b^2 - 2(\alpha + X)}\right) - at\left(b - \sqrt{b^2 - 2X}\right)\right), \quad (6.102)$$

which shows that the conditional exponent is of the required form $\psi_0(\alpha + X) - \psi_0(X)$. $\square$
6.8 Normal inverse Gaussian information process

6.8.1 Definition and properties

By a normal inverse Gaussian (NIG) process (Rydberg 1997, Barndorff-Nielsassen 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

$$\psi_0(\alpha) = m\left(\sqrt{a^2 - b^2} - \sqrt{a^2 - (b + \alpha)^2}\right)$$  \hspace{1cm} (6.103)

for $\alpha \in \{w \in \mathbb{C} : -a - b < \Re w < a - b\}$. Let us write $\{I_t\}_{t \geq 0}$ for the NIG process. The probability density for its value at time $t$ is given by

$$P_0(I_t \in dx) = \frac{amtK_1(a \sqrt{m^2t^2 + x^2})}{\pi \sqrt{m^2t^2 + x^2}} \exp\left(mt\sqrt{a^2 - b^2 + bx}\right) dx,$$  \hspace{1cm} (6.104)

where $K_\nu$ is the modified Bessel function of third order. The Lévy measure of the NIG process is given by:

$$\nu(dx) = \frac{am \exp(bx)K_1(a|x|)}{|x|} dx.$$  \hspace{1cm} (6.105)

If $\{I_t\}$ is an NIG process with parameters $(a, b, m)$, then it follows from (6.103) that $\{-I_t\}$ is an NIG process with parameters $(a, -b, m)$. If $b = 0$, then we say that the NIG process is symmetric. A calculation shows that

$$\mathbb{E}^{P_0}[I_t] = \frac{mtb}{\sqrt{a^2 - b^2}} \quad \text{and} \quad \text{Var}^{P_0}[I_t] = \frac{a^2 mt}{(a^2 - b^2)^3}. \hspace{1cm} (6.106)$$

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

$$a' = 1 \quad \text{and} \quad b' = m\sqrt{a^2 - b^2}. \hspace{1cm} (6.107)$$

Then the Lévy exponent of the process $\{I_t\}$ defined by

$$I_t = bm^2G_t + mB_{G_t} \hspace{1cm} (6.108)$$

is given by (6.103).
6.8.2 Normal inverse Gaussian information

The associated information process is constructed as follows. We fix a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and the parameters \(a, b,\) and \(m\). Then we have:

**Proposition 6.8.1.** Let the random variable \(X\) satisfy \(-a - b < X < a - b\) almost surely, let \(\{G_t^X\}\) be \(\mathcal{F}_t^X\)-conditionally IG, with parameters

\[
a' = 1 \quad \text{and} \quad b' = m\sqrt{a^2 - (b + X)^2},
\]

and let

\[
F_t = m^2 G_t^X.
\]

Then the process \(\{\xi_t\}\) defined by

\[
\xi_t = (b + X)F_t + B(F_t)
\]

is a Lévy information process with message \(X\) and normal inverse Gaussian noise, with fiducial exponent \((6.103)\).

**Proof.** We observe that the condition on \(\{G_t^X\}\) is that

\[
t^{-1}\ln \mathbb{E}[\exp(\alpha G_t^X) \mid X] = \delta \sqrt{a^2 - (b + X)^2} - \sqrt{m^2(a^2 - (b + X)^2) - 2\alpha}
\]

for \(\alpha \in \mathbb{C}^I\). Thus setting \(\psi_X(\alpha) = \mathbb{E}[\exp(\alpha \xi_t) \mid X]\) for \(\alpha \in \mathbb{C}^I\) it follows that

\[
\psi_X(\alpha) = \mathbb{E}[\exp(\alpha(b + X)F_t + \alpha B(F_t)) \mid X]
\]

\[
= \mathbb{E}\left[\exp \left( (\alpha(b + X) + \frac{1}{2}\alpha^2)m^2 G_t^X \right) \mid X \right]
\]

\[
= \mathbb{E}\left[\exp \left( m\sqrt{a^2 - (b + X)^2} - m\sqrt{a^2 - (b + X)^2} - 2 \left( \alpha(b + X) + \frac{1}{2}\alpha^2 \right) \right) \right],
\]

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

6.9 Generalised hyperbolic information process

Similar arguments lead to the construction of information processes based on other Lévy processes related to the inverse Gaussian distribution, including for example the hyperbolic process (Eberlein & Keller 1995, Eberlein, Keller & Prause 1998, Bingham & Kiesel 2001), and the generalised hyperbolic (GH) process (Eberlein 2001, Prause 1999).
In what follows below we construct a class of information processes based on the GH noise family.

6.9.1 Definition and properties

The GH process derives from the so-called generalised hyperbolic distribution introduced by Barndorff-Nielsen (1977). The GH distribution has the infinitely-divisible property (Barndorff-Nielsen & Halgreen 1977) and thus gives rise to a class of Lévy processes whose properties have been investigated extensively in the literature.

The GH family has four parameters, in terms of which the associated Lévy exponent is of the form

$$\psi_0(\alpha) = \ln \left[ \left( \frac{a^2 - b^2}{a^2 - (b + \alpha)^2} \right)^{\lambda/2} \frac{K_\lambda \left( \delta \sqrt{a^2 - (b + \alpha)^2} \right)}{K_\lambda \left( \delta \sqrt{a^2 - b^2} \right)} \right]. \quad (6.113)$$

Here $a > 0$, $|b| < a$, $\delta > 0$, and $\lambda \in \mathbb{R}$; and $K_\lambda$ denotes the modified Bessel function of the third order with index $\lambda$. Additionally, we have

$$\delta \geq 0 \quad \text{and} \quad |b| < a \quad \text{if} \quad \lambda > 0,$$

$$\delta > 0 \quad \text{and} \quad |b| < a \quad \text{if} \quad \lambda = 0,$$

$$\delta > 0 \quad \text{and} \quad |b| \leq a \quad \text{if} \quad \lambda < 0. \quad (6.114)$$

In what follows we shall write $\{H_t\}_{t \geq 0}$ for the generalised hyperbolic (GH) process. The density function for the value of the process at time $t = 1$ is given by

$$f(x) = \frac{(a^2 - b^2)^{\frac{\lambda}{2}} (\delta^2 + x^2)^{\frac{\lambda - 1}{2}} e^{bx}}{\sqrt{2\pi a^{\lambda - 1}} \frac{\delta^{\lambda} K_{\lambda}(\delta \sqrt{a^2 - b^2})}{\sqrt{a^2 - (b + \alpha)^2}}},$$

$$K_{\lambda - \frac{1}{2}}(a \sqrt{\delta^2 + x^2}). \quad (6.115)$$

The term “generalised hyperbolic process” is in fact something of a misnomer, since, as Eberlein (2001) points out, although the process has a generalised hyperbolic distribution at time $t = 1$, for general $t$ the distribution of $H_t$ is not in the generalised hyperbolic family. Nevertheless, on account of the infinitely divisible property of the GH distribution, the distribution of the GH process is completely determined by its marginal distribution at $t = 1$, this being of course a general property of Lévy processes. Thus we can use the parameters of the GH distribution (which is applicable to the value of the process at $t = 1$) to characterise the properties of the whole process.
The Lévy measure of the GH process is known to take the following form:

\[
\nu(dx) = \frac{\exp(bx)}{|x|} \left[ \int_0^\infty \frac{\exp(-|x|\sqrt{2u + a^2})}{\pi^2 u (J^2_\lambda(\delta\sqrt{2u}) + N^2_\lambda(\delta\sqrt{2u}))} \, du + \lambda \exp(-a|x|) \right]
\]

(6.116)

if \( \lambda \geq 0 \), and

\[
\nu(dx) = \frac{\exp(bx)}{|x|} \int_0^\infty \frac{\exp(-|x|\sqrt{2u + a^2})}{\pi^2 u (J^2_\lambda(\delta\sqrt{2u}) + N^2_\lambda(\delta\sqrt{2u}))} \, du
\]

(6.117)

if \( \lambda < 0 \). Here \( J_\lambda(x) \) is a Bessel function of the first kind and \( Y_\lambda(x) \) is a Bessel function of the second kind. The GH process has mean

\[
\mathbb{E}^P_0[H_t] = b\delta t \frac{K_{\lambda+1}(\delta t \sqrt{a^2 - b^2})}{K_\lambda(\delta t \sqrt{a^2 - b^2})},
\]

(6.118)

and variance

\[
\text{Var}^P_0[H_t] = \delta^2 t \left[ \frac{b^2}{a^2 - b^2} \left( \frac{K_{\lambda+2}(\zeta)}{K_\lambda(\zeta)} - \frac{K^2_\lambda(\zeta)}{K_\lambda(\zeta)} \right) + \frac{K_{\lambda+1}(\zeta)}{\zeta K_\lambda(\zeta)} \right],
\]

(6.119)

where \( \zeta = \delta \sqrt{a^2 - b^2} \).

The generalised hyperbolic process can be generated by use of the technique of time change. First, we consider a generalised inverse Gaussian (GIG) process with parameters \((a, b, \lambda)\). The GIG process is a generalisation of the IG process, and has the following Lévy exponent:

\[
\psi_0(\alpha) = \left( \frac{b^2}{b^2 - 2\alpha a} \right)^{\lambda/2} \frac{K_\lambda \left( a \sqrt{b^2 - 2\alpha} \right)}{K_\lambda(ab)}. \]

(6.120)

We then let \( \{F_t\} \) be a GIG process with parameters \((\delta, \sqrt{a^2 - b^2}, \lambda)\), and consider the process

\[
H_t = bF_t + B_t,
\]

(6.121)

where \( \{B_t\} \) is a standard Brownian motion. It follows then that

\[
\mathbb{E}^P \left[ e^{\alpha H_t} \right] = \mathbb{E}^P \left[ \mathbb{E}^P \left[ e^{\alpha F_t + \alpha B_t} \right] \bigg| F_t \right] = \mathbb{E}^P \left[ e^{(ab + \frac{1}{2} a^2)F_t} \right] = \left[ \frac{a^2 - b^2}{a^2 - (b + \alpha)^2} \right]^{\lambda/2} \frac{K_\lambda \left( \delta \sqrt{a^2 - (b + \alpha)^2} \right)}{K_\lambda \left( \delta \sqrt{a^2 - b^2} \right)}
\]

(6.122)
which shows that the process \( \{ H_t \} \) thus defined as above by use of a subordinator is indeed a GH process.

### 6.9.2 Generalised hyperbolic information

With the observations given in the foregoing material, we are able to deduce the following:

**Proposition 6.9.1.** Let the random variable \( X \) and the \( \mathbb{P} \)-Brownian motion \( \{ B_t \} \) be \( \mathbb{P} \)-independent, and assume that the process \( \{ F_t \} \) is \( \mathcal{F}^X \)-conditionally a generalised inverse Gaussian process with the parameter set \((\delta, \sqrt{a^2 - (b + X)^2}, \lambda)\), and that \( \{ F_t \} \) is \( \mathbb{P} \)-independent of \( \{ B_t \} \). Then the process \( \{ \xi_t \} \) defined by

\[
\xi_t = (b + X)F_t + B_{F_t}
\]  

(6.123)

is a Lévy information process with signal \( X \), generalized hyperbolic noise, and fiducial exponent (6.113).

**Proof.** To work out the conditional Lévy exponent of \( \xi_t \) we calculate as follows:

\[
\mathbb{E}^\mathbb{P}\left[ e^{a\xi_t} \right] = \mathbb{E}^\mathbb{P}\left[ \mathbb{E}^\mathbb{P}\left[ e^{(b+X)F_t + \alpha B_{F_t}} \middle| X, F_t \right] \right] \\
= \mathbb{E}^\mathbb{P}\left[ e^{(\alpha(b+X) + \frac{1}{2}a^2)F_t} \right] \\
= \int \left[ \left( \frac{a^2 - (b + x)^2}{a^2 - (b + x + \alpha)^2} \right)^{\lambda/2} \frac{K_\lambda \left( \delta \sqrt{a^2 - (b + x + \alpha)^2} \right)}{K_\lambda \left( \delta \sqrt{a^2 - (b + x)^2} \right)} \right]^t \pi(dx).
\]  

(6.124)

We see then that the conditional Lévy exponent of the GH information process is of the form

\[
\psi_X(\alpha) = \ln \left[ \left( \frac{a^2 - (b + X)^2}{a^2 - (b + X + \alpha)^2} \right)^{\lambda/2} \frac{K_\lambda \left( \delta \sqrt{a^2 - (b + X + \alpha)^2} \right)}{K_\lambda \left( \delta \sqrt{a^2 - (b + X)^2} \right)} \right].
\]  

(6.125)

This is consistent with the requirement that the conditional exponent should be of the form \( \psi_X(\alpha) = \psi_0(\alpha + X) - \psi_0(X) \).
6.10 Concluding remarks

Recent developments in the phenomenological representation of physical time series (Brody & Hughston 2002, 2005, 2006) and financial time series (Macrina 2006, Brody, Hughston & Macrina 2008a,b, Hoyle 2010, Hoyle, Hughston & Macrina 2011, Mackie 2012, Brody, Hughston & Mackie 2012, Brody, Hughston & Yang 2013a, 2013b, Brody & Hughston 2013) 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 sciences 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 theory of signal processing with Lévy information outlined in the foregoing material will find a variety of interesting and exciting applications.

I conclude this part of the thesis by remarking that the extension of the theory of signal processing beyond the Brownian and Poisson categories of processes for representing noise to the general Lévy class opens up an entirely new line of research in this area. The specific examples worked out above illustrate the fact that each Lévy process is different in capturing different types of noise that might arise in a variety of situations, and that provided that the structure of the signal is not overly complicated, there can be a great deal of analytic tractability. These observations are encouraging, and we hope that further progress will be made in this line of research.
Appendix A

Proof of joint Markov property satisfied by generators of commodity information filtration

For convenience of reference we restate the proposition 2.5.1 below, following which we provide the details of the proof.

**Proposition 2.5.1** The information process \( \{ \xi_t \} \) and the convenience dividend process \( \{ X_t \} \) are jointly Markovian, that is, the following relation holds:

\[
E \left[ \int_t^\infty P_uX_u du \ \bigg| \ {\xi_s}_{0 \leq s \leq t}, {X_s}_{0 \leq s \leq t} \right] = E \left[ \int_t^\infty P_uX_u du \ \bigg| \ \xi_t, X_t \right]. \tag{A.1}
\]

**Proof.** We define an alternative noisy observation process for the cumulative dividends by setting

\[
\eta_t = \sigma t \int_0^\infty P_uX_u du + B_t, \tag{A.2}
\]

which then implies that the main information process can be expressed in the following form:

\[
\xi_t = \eta_t - \sigma t \int_0^t P_uX_u du. \tag{A.3}
\]

It should thus be evident that the filtration is given by

\[
\mathcal{F}_t = \sigma (\{ X_s \}_{0 \leq s \leq t}, \{ \xi_s \}_{0 \leq s \leq t}) = \sigma (\{ X_s \}_{0 \leq s \leq t}, \{ \eta_s \}_{0 \leq s \leq t}). \tag{A.4}
\]
Next we observe that the information process $\{\eta_t\}$ has the Markovian property: that is to say, we have

$$\mathbb{Q}(\eta_T \leq x | \mathcal{F}_t^0) = \mathbb{Q}(\eta_T \leq x | \eta_t), \quad (A.5)$$

for all $t \leq T$ and for all $x \in \mathbb{R}$. To verify this it suffices to show that

$$\mathbb{Q}(\eta_t \leq x | \eta_s, \eta_{s_1}, \eta_{s_2}, \ldots, \eta_{s_k}) = \mathbb{Q}(\eta_t \leq x | \eta_s), \quad (A.6)$$

for any collection of times $t, s, s_1, s_2, \ldots, s_k$ such that $t \geq s \geq s_1 \geq s_2 \geq \cdots \geq s_k > 0$. Now it is a property of Brownian motion that for any times $t, s, s_1$ satisfying $t > s > s_1 > 0$ the random variables

$$B_t, \quad B_s/s - B_{s_1}/s_1$$

are independent. This fact arises as an ingredient of the theory of the Brownian bridge. More generally, if $s > s_1 > s_2 > s_3 > 0$, we find that the random variables

$$\frac{B_s}{s} - \frac{B_{s_1}}{s_1}, \quad \frac{B_{s_2}}{s_2} - \frac{B_{s_1}}{s_1}, \quad \frac{B_{s_3}}{s_3} - \frac{B_{s_2}}{s_2}, \ldots$$

are independent. Next we note that

$$\frac{\eta_s}{s} - \frac{\eta_{s_1}}{s_1} = \frac{B_s}{s} - \frac{B_{s_1}}{s_1}.$$

It follows therefore that

$$\mathbb{Q}(\eta_t \leq x | \eta_s, \eta_{s_1}, \eta_{s_2}, \ldots, \eta_{s_k}) = \mathbb{Q}(\eta_t \leq x | \eta_s, \frac{\eta_s}{s} - \frac{\eta_{s_1}}{s_1}, \frac{\eta_{s_1}}{s_1} - \frac{\eta_{s_2}}{s_2}, \ldots, \frac{\eta_{s_{k-1}}}{s_{k-1}} - \frac{\eta_{s_k}}{s_k})$$

$$= \mathbb{Q}(\eta_t \leq x | \eta_s, \frac{B_s}{s} - \frac{B_{s_1}}{s_1}, \frac{B_{s_2}}{s_2} - \frac{B_{s_1}}{s_1}, \ldots, \frac{B_{s_{k-1}}}{s_{k-1}} - \frac{B_{s_k}}{s_k}). \quad (A.7)$$

However, since $\eta_t$ and $\eta_s$ are both independent of the random variables

$$\frac{B_s}{s} - \frac{B_{s_1}}{s_1}, \quad \frac{B_{s_2}}{s_2} - \frac{B_{s_1}}{s_1}, \ldots, \frac{B_{s_{k-1}}}{s_{k-1}} - \frac{B_{s_k}}{s_k},$$

the claimed result follows. Let us now define

$$G_t = \sigma\left(\left\{\frac{\eta_t}{s} - \frac{\eta_s}{s}\right\}_{0 \leq s \leq t}\right). \quad (A.8)$$
Appendix A. Proof of joint Markov property

Note that $\eta_t$ and $\eta_T$ are independent from $\mathcal{G}_t$, and furthermore, that $\{X_s\}$ is independent of $\mathcal{G}_t$. Thus we conclude that the price is given by

$$P_t S_t = \mathbb{E} \left[ \int_t^\infty P_u X_u du \bigg| \mathcal{F}_t \right]$$

$$= \mathbb{E} \left[ \int_t^\infty P_u X_u du \bigg| \eta_t, \mathcal{G}_t, \{X_s\}_{0 \leq s \leq t} \right]$$

$$= \mathbb{E} \left[ \int_t^\infty P_u X_u du \bigg| \eta_t, \{X_s\}_{0 \leq s \leq t} \right]$$

$$= \mathbb{E} \left[ \int_t^\infty P_u X_u du \bigg| \eta_t, \{X_s\}_{0 \leq s \leq t} \right]. \quad (A.9)$$

On the other hand we observe that

$$\sigma(\{X_s\}_{0 \leq s \leq t}, \xi_t) = \sigma(X_t, \xi_t, \{b_{st}\}_{0 \leq s \leq t}), \quad (A.10)$$

and that the OU bridge $\{b_{st}\}_{0 \leq s \leq t}$ is independent of $\{X_u\}_{u \geq t}$. Thus $\{b_{st}\}$ is independent of $\xi_t$ and $\int_t^\infty P_u X_u du$.

Therefore, we have proved the claim and have deduced that

$$S_t = \frac{1}{P_t} \mathbb{E} \left[ \int_t^\infty P_u X_u du \bigg| \xi_t, X_t \right]. \quad (A.11)$$
Appendix B

Derivation of European commodity option pricing formula

In Chapter 3, we claimed in Proposition 3.1.1 that the initial price $C_0$ of a European style commodity option can be expressed in the form:

$$
C_0 = e^{-rT} \left[ \sqrt{\frac{\text{Var}[S_T]}{2\pi}} \exp \left( -\frac{(\mathbb{E}[S_T] - K)^2}{2\text{Var}[S_T]} \right) \right] + e^{-rT} \left[ (\mathbb{E}[S_T] - K) N \left( \frac{\mathbb{E}[S_T] - K}{\sqrt{\text{Var}[S_T]}} \right) \right], \quad (B.1)
$$

where $N(x)$ is the standard normal distribution function:

$$
N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left( -\frac{1}{2} z^2 \right) \, dz. \quad (B.2)
$$

The proof of this result is as follows.

Proof. Let $\mu = \mathbb{E}[S_T]$ and $\gamma^2 = \text{Var}[S_T]$. From (3.4) we have:

$$
C_0 = e^{-rT} \frac{1}{\sqrt{2\pi\gamma}} \int_{K}^{\infty} (z - \mu + \mu - K) \exp \left( -\frac{(z - \mu)^2}{2\gamma^2} \right) \, dz
$$

$$
= e^{-rT} \frac{1}{\sqrt{2\pi}} \int_{K}^{\infty} \frac{z - \mu}{\gamma} \exp \left( -\frac{(z - \mu)^2}{2\gamma^2} \right) \, dz
$$

$$
= e^{-rT} \frac{1}{\sqrt{2\pi}} \int_{K}^{\infty} \frac{z - \mu}{\gamma} \exp \left( -\frac{(z - \mu)^2}{2\gamma^2} \right) \, dz \quad (B.3)
$$

$$
+ e^{-rT} \frac{1}{\sqrt{2\pi}} \int_{K}^{\infty} \frac{\mu - K}{\gamma} \exp \left( -\frac{(z - \mu)^2}{2\gamma^2} \right) \, dz. \quad (B.4)
$$
In the term labelled (B.3), let \( p = (z - \mu)/\gamma \). We then have \( dz = \gamma dp \). Thus we have
\[
\frac{e^{-rT}}{\sqrt{2\pi}} \int_{\frac{K-\mu}{\gamma}}^{\infty} p \exp\left(-\frac{p^2}{2}\right) \gamma dp = \frac{e^{-rT} \gamma}{\sqrt{2\pi}} \int_{\frac{K-\mu}{\gamma}}^{\infty} p \exp\left(-\frac{p^2}{2}\right) dp. \tag{B.5}
\]

In (B.5), let \( t = p^2/2 \). Then we have \( dp = p^{-1}dt \). The range of integration changes to \( (K - \mu)^2 / 2\gamma^2 \leq t \leq \infty \). Then we obtain
\[
(B.5) = \frac{e^{-rT} \gamma}{\sqrt{2\pi}} \int_{\frac{(K-\mu)^2}{2\gamma^2}}^{\infty} e^{-t} dt = \frac{e^{-rT} \gamma}{\sqrt{2\pi}} \exp\left(-\frac{(K - \mu)^2}{2\gamma^2}\right). \tag{B.6}
\]

For the term labelled (B.4), we have
\[
e^{-rT} \frac{\mu - K}{\sqrt{2\pi} \gamma^2} \left[ \int_{-\infty}^{\infty} \exp\left(-\frac{(z - \mu)^2}{2\gamma^2}\right) dz - \int_{-\infty}^{K} \exp\left(-\frac{(z - \mu)^2}{2\gamma^2}\right) dz \right]
= e^{-rT} (\mu - K) \left[ 1 - \frac{1}{\sqrt{2\pi} \gamma^2} \int_{-\infty}^{K} \exp\left(-\frac{(z - \mu)^2}{2\gamma^2}\right) dz \right]
= e^{-rT} (\mu - K) \left[ 1 - \frac{1}{\sqrt{2\pi} \gamma^2} \int_{-\infty}^{\frac{K-\mu}{\gamma}} \exp\left(-\frac{p^2}{2}\right) dp \right]
= e^{-rT} (\mu - K) \left[ 1 - N\left(\frac{K - \mu}{\gamma}\right)\right]
= e^{-rT} (\mu - K) \left[N\left(\frac{\mu - K}{\gamma}\right)\right]. \tag{B.7}
\]

Therefore, the claim immediately follows from (B.6) + (B.7). \( \square \)
Appendix C

Constant parameter OU process vs time-inhomogeneous OU process

We would like to investigate whether the observations made on the behaviour of the model in the time homogeneous setup can provide useful insights regarding the behaviour of the model in the time inhomogeneous setup. That is to say, we would like to understand better the relationship between (i) a constant parameter OU process multiplied by a deterministic increasing function of time, and (ii) a time-dependent OU process.

We know that the dynamics of the constant parameter OU process are given by

\[ dX_t = \kappa(\theta - X_t)dt + \psi d\beta_t, \tag{C.1} \]

where \{\beta_t\} is a Brownian motion that is independent of \{B_t\}, \theta is the mean reversion level, \kappa is the mean reversion speed, and \psi is the dividend volatility. The closed-form solution of the stochastic differential equation above is:

\[ X_t = e^{-\kappa t}X_0 + \theta(1 - e^{-\kappa t}) + \psi e^{-\kappa t} \int_0^t e^{\kappa s} d\beta_s. \tag{C.2} \]

Now define a new process \{Y_t\} such that \( Y_t = \alpha_t X_t \). We have:

\[ Y_t \equiv \alpha_t X_t = \alpha_t e^{-\kappa t}X_0 + \alpha_t \theta(1 - e^{-\kappa t}) + \alpha_t \psi e^{-\kappa t} \int_0^t e^{\kappa s} d\beta_s. \tag{C.3} \]
We also know that the dynamics of the time-dependent OU process are given by

\[ dZ_t = \kappa_t (\theta_t - Z_t) dt + \psi_t dB_t, \]  

and that the closed-form solution for the process is:

\[ Z_t = e^{-f_t} \left[ Z_0 + \int_0^t e^{f_s} \kappa_s \theta_s ds + \int_0^t \psi_s e^{f_s} dB_s \right], \]

where \( f_t = \int_0^t \kappa_s ds \).

**Proposition C.0.1.** A constant parameter OU process \( \{X_t\} \), when multiplied by a deterministic increasing function of time, gives rise to a special case of the general time-inhomogeneous OU process \( \{Z_t\} \).

**Proof.** We are going to present the proof from two angles. First we shall deduce the proof from the closed-form solutions, and then we shall discuss the claim from SDE point of view. We know that the closed-form solutions for \( Y_t \) and \( Z_t \) are of the form:

\[ Y_t = \alpha_t e^{-\kappa t} \left[ X_0 + \theta (e^{\kappa t} - 1) + \psi e^{-\kappa t} \int_0^t e^{\kappa s} ds \right], \]  

and

\[ Z_t = e^{-f_t} \left[ Z_0 + \int_0^t e^{f_s} \kappa_s \theta_s ds + \int_0^t \psi_s e^{f_s} dB_s \right]. \]

We shall compare these equations term-by-term to show that the former is a special case of the latter. Consider the term outside of the bracket first. At the moment, we leave out the term \( \alpha_t \) and come back to it later. We know that

\[ f_t = \int_0^t \kappa_s ds, \]

and from the definition of an OU process, that the mean-reversion rate satisfies \( \kappa_t > 0, \forall t \). Therefore the integration \( f_t \) is positive and increasing as \( t \) increases. Thus \( e^{-f_t} \) is a strictly decreasing function of time \( t \). On the other hand, since \( \kappa > 0 \) from definition, we have \( e^{-\kappa t} \) is a strictly decreasing function of time \( t \).

Moving on to the terms inside the brackets, we can skip the \( X_0 \) term since it is constant. Given the assumptions that both \( \kappa_t > 0 \) and \( \theta_t > 0, \forall t \), it follows that

\[ \int_0^t e^{f_s} \kappa_s \theta_s ds > 0, \]
and that the integral is increasing as $t$ increases. On the other hand, since we have the constants $\kappa > 0$ and $\theta > 0$, it follows that $\theta(e^{\kappa t} - 1) > 0$ and is increasing in time.

For the third term inside the brackets, we have for both expressions a stochastic integral driven by the same Brownian motion. Under the assumption that $\psi_t > 0, \forall t$, we have that both integrands are greater than zero and are increasing with time.

Thus, for both expressions, the terms inside bracket behave in a similar manner (positive, increasing over time, driven by the same Brownian motion). Therefore, we can see that, with certain parameters fixed in $Z_t$, we can form a new process $Y_t$ with the desired characteristics.

Now we are going to use the SDE method to prove the claim. Let

$$Y_t = \alpha_t X_t.$$  \hfill (C.10)

From Ito’s lemma, we have:

$$dY_t = X_t d\alpha_t + \alpha_t \left[ \kappa (\theta - X_t) dt + \psi d\beta_t \right]$$

$$= X_t \alpha_t' dt + \alpha_t \kappa (\theta - X_t) dt + \alpha_t \psi d\beta_t$$

$$= \left[ \kappa \theta \alpha_t - \left( \kappa \theta \alpha_t - \alpha_t' \right) X_t \right] dt + \alpha_t \psi d\beta_t.$$  \hfill (C.11)

We can see that the above equation is in the form of a generic time-dependent OU process:

$$d\hat{Z}_t = \left( a_t - \beta_t \hat{Z}_t \right) dt + c_t d\beta_t.$$  \hfill (C.12)

Moreover, if we impose an extra condition for the parameters such that $\kappa \theta \alpha_t - \alpha_t' \neq 0$, we can re-arrange it into the form:

$$dY_t = \left( \kappa \theta \alpha_t - \alpha_t' \right) \left[ \kappa \theta \alpha_t - \alpha_t' - X_t \right] dt + \alpha_t \psi d\beta_t.$$  \hfill (C.13)

which has the structure of the time-dependent OU process defined in equation (C.4).

Now one can see that these two expressions can be written in the same form. \hfill \Box
Appendix D

Conditional variance of convenience dividend flow

The goal in this appendix is to work out an expression for the conditional variance $V_t$ of the future dividend flow in the commodity pricing model. The conditional variance can be expressed as follows:

$$V_t = E_t \left[ \left( \int_t^\infty e^{-ru} X_u du \right)^2 \right] - \left\{ E_t \left[ \int_t^\infty e^{-ru} X_u du \right] \right\}^2. \quad (D.1)$$

From the orthogonal decomposition of the OU process, we know that

$$\left( \int_t^\infty e^{-ru} X_u du \right)^2 = \left[ A_t + e^{rt} X_t \int_t^\infty e^{-(r+k)u} du \right]^2$$

$$= A_t^2 + 2e^{rt} X_t A_t \int_t^\infty e^{-(r+k)u} du + e^{2rt} X_t^2 \left[ \int_t^\infty e^{-(r+k)u} du \right]^2$$

$$= A_t^2 + \frac{2X_te^{-rt}}{r + \kappa} A_t + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}, \quad (D.2)$$

where

$$A_t = \int_t^\infty e^{-ru} \left( X_u - e^{-\kappa(u-t)} X_t \right) du. \quad (D.3)$$
The conditional expectation of the term \((\int_t^\infty e^{-ru}X_u du)^2\) is given by

\[
\mathbb{E}_t \left[ \left( \int_t^\infty e^{-ru}X_u du \right)^2 \right] = \mathbb{E}_t \left[ A_t^2 + \frac{2X_t e^{-rt}}{r + \kappa} A_t + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2} \right]
\]

\[
= \mathbb{E}_t \left[ A_t^2 \right] + \frac{2X_t e^{-rt}}{r + \kappa} \mathbb{E}_t \left[ A_t \right] + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}
\]

\[
= \mathbb{E}_t \left[ A_t^2 \right] + \frac{2X_t e^{-rt}}{r + \kappa} \mathbb{E}_t \left[ A_t \right] + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}
\]

\[
= \mathbb{E} \left\{ \left[ z_t(A + B) + (1 - z_t)A - z_t B \right]^2 \right\} + \frac{2X_t e^{-rt}}{r + \kappa} \mathbb{E}_t \left[ A_t \right] + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}
\]

\[
= \mathbb{E} \left\{ z_t^2(A + B)^2 + (1 - z_t)^2(A + B)^2 \right\} + \frac{2X_t e^{-rt}}{r + \kappa} \mathbb{E}_t \left[ A_t \right] + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}
\]

\[
= z_t^2(A + B)^2 + 2z_t(1 - z_t)(A + B)\mathbb{E}[A] + \frac{2X_t e^{-rt}}{r + \kappa} \mathbb{E}_t \left[ A_t \right] + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}
\]

\[
+ \frac{z_t^2}{\sigma^2 t} + \frac{2X_t e^{-rt}}{r + \kappa} \mathbb{E}_t \left[ A_t \right] + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}.
\]

(D.4)

The seventh equality above holds because \(A + B\) and \((1 - z_t)A - z_t B\) are independent and also \(\mathbb{E}[B] = 0\), since \(B = \frac{B_t}{\sigma^2 t}\) and \(B_t\) is a standard Brownian motion. The last equality is obtained since \(A\) and \(B\) are independent and \(\mathbb{E}[B] = 0\). Now consider

\[
\mathbb{E}_t \left[ \int_t^\infty e^{-ru}X_u du \right] = \mathbb{E}_t \left[ A_t \right] + X_t e^{-rt} \int_t^\infty e^{-(r + \kappa)u} du
\]

\[
= z_t(A + B) + (1 - z_t)\mathbb{E}[A] - z_t\mathbb{E}[B] + \frac{X_t e^{-rt}}{r + \kappa}
\]

\[
= z_t(A + B) + (1 - z_t)\mathbb{E}[A] + \frac{X_t e^{-rt}}{r + \kappa}.
\]

(D.5)

Therefore we deduce that

\[
\left\{ \mathbb{E}_t \left[ \int_t^\infty e^{-ru}X_u du \right] \right\}^2
\]

\[
= z_t^2(A + B)^2 + (1 - z_t)^2 \left\{ \mathbb{E}[A] \right\}^2 + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2} + 2z_t(1 - z_t)(A + B)\mathbb{E}[A]
\]

\[
+ 2z_t(A + B)\frac{X_t e^{-rt}}{r + \kappa} + 2(1 - z_t)\mathbb{E}[A] \frac{X_t e^{-rt}}{r + \kappa}.
\]

(D.6)
Finally, coming back to the expression for the conditional variance, we have

\[
V_t = \mathbb{E}_t \left[ \left( \int_t^{\infty} e^{-ru} X_u du \right)^2 \right] - \left\{ \mathbb{E}_t \left[ \int_t^{\infty} e^{-ru} X_u du \right] \right\}^2
\]

\[
= z_t^2 (A + B)^2 + 2z_t (1 - z_t)(A + B)\mathbb{E}_t[A] + (1 - z_t)^2 \mathbb{E}_t[A^2] + \frac{z_t^2}{\sigma^2 t}
\]

\[
+ 2X_t e^{-rt} \mathbb{E}_t[A_t] A + B + \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}
\]

\[
- z_t^2 (A + B)^2 - (1 - z_t)^2 \{\mathbb{E}_t[A]\}^2 - \frac{X_t^2 e^{-2rt}}{(r + \kappa)^2}
\]

\[
- 2z_t (1 - z_t)(A + B)\mathbb{E}_t[A] - 2z_t (A + B) X_t e^{-rt}
\]

\[
- 2 (1 - z_t) \mathbb{E}_t[A] \frac{X_t e^{-rt}}{r + \kappa}.
\]

(D.7)

After some cancellations and rearrangements of terms, we obtain

\[
V_t = (1 - z_t)^2 \left\{ \mathbb{E}_t[A^2] - \{\mathbb{E}_t[A]\}^2 \right\} + \frac{z_t^2}{\sigma^2 t} + \frac{2X_t e^{-rt}}{r + \kappa} \mathbb{E}_t[A_t] A + B
\]

\[
- \frac{2X_t e^{-rt}}{r + \kappa} [z_t (A + B) + (1 - z_t) \mathbb{E}_t[A]]
\]

\[
= (1 - z_t)^2 \mathbb{V}ar[A_t] + \frac{z_t^2}{\sigma^2 t} + \frac{2X_t e^{-rt}}{r + \kappa} [z_t (A + B) + (1 - z_t) \mathbb{E}_t[A]]
\]

\[
- \frac{2X_t e^{-rt}}{r + \kappa} [z_t (A + B) + (1 - z_t) \mathbb{E}_t[A]]
\]

\[
= (1 - z_t)^2 \mathbb{V}ar[A_t] + \frac{z_t^2}{\sigma^2 t}
\]

\[
= (1 - z_t)^2 \frac{\psi^2}{(r + \kappa)^2} \int_t^{\infty} e^{-2rs} ds + \frac{z_t^2}{\sigma^2 t}
\]

\[
= \frac{\psi^2 (1 - z_t)^2}{2r(r + \kappa) e^{2rt}} + \frac{z_t^2}{\sigma^2 t}
\]

\[
= \frac{\psi^2}{2r(r + \kappa) e^{2rt}} \left[ \frac{2r(r + \kappa)^2 e^{2rt}}{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t} \right]^2 + \frac{1}{\sigma^2 t} \left[ \frac{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t}{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t} \right]^2
\]

\[
= \frac{2r \psi^2 (r + \kappa)^2 e^{2rt}}{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t} \left[ \frac{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t}{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t} \right]^2
\]

\[
= \frac{\psi^2 (2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t)}{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t}
\]

\[
= \frac{2r(r + \kappa)^2 e^{2rt} + \sigma^2 \psi^2 t}{\sigma^2 t}
\]

(D.8)


