UNIVERSITY OF KWAZULU-NATAL

Applications of Symmetry Analysis of Partial Differential and Stochastic Differential Equations Arising from Mathematics of Finance

Felix Noyanim Nwobi

Applications of Symmetry Analysis of Partial Differential and Stochastic Differential Equations Arising from Mathematics of Finance

Felix Noyanim Nwobi
BSc. Hons, MSc. University of Nigeria, Nsukka

Submitted in fulfilment of the academic requirements for the degree of

Doctor of Philosophy in the School of Mathematical Sciences

University of KwaZulu-Natal

Durban, South Africa

November 2011

As the candidate's supervisors, we have approved this dissertation for submission.

Signed:	Professor PGL Leach	November 2011
Signed:	Dr. JG O'Hara	November 2011

Abstract

In the standard modeling of the pricing of options and derivatives as generally understood these days the underlying process is taken to be a Wiener Process or a Levy Process. The stochastic process is modeled as a stochastic differential equation. From this equation a partial differential equation is obtained by application of the Feynman-Kac Theorem. The resulting partial differential equation is of Hamilton-Jacobi-Bellman type.

Analysis of the partial differential equations arising from Mathematics of Finance using the methods of the Lie Theory of Continuous Groups has been performed over the last twenty years, but it is only in recent years that there has been a concerted effort to make full use of the Lie theory. We propose an extension of Mahomed and Leach's (1990) formula for the nth-prolongation of an nth-order ordinary differential equation to the nth-prolongation of the generator of an hyperbolic partial differential equation with p dependent and k independent variables. The symmetry analysis of this partial differential equation shows that the associated Lie algebra is $\{sl(2,R) \oplus W_3\} \oplus_s \infty A_1$ with 12 optimal systems.

A modeling approach based upon stochastic volatility for modeling prices in the deregulated Pennsylvania State Electricity market is adopted for application. We propose a dynamic linear model (DLM) in which switching structure for the measurement matrix is incorporated into a two-state Gaussian mixture/first-order autoregressive (AR (1)) configuration in a nonstationary independent process defined by time-varying probabilities. The estimates of maximum likeli-

hood of the parameters from the "modified" Kalman filter showed a significant mean-reversion rate of 0.9363 which translates to a half-life price of electricity of nine months. Associated with this mean-reversion is the high measure of price volatility at 35%.

Within the last decade there has been some work done upon the symmetries of stochastic differential equations. Here empirical results contradict earliest normality hypotheses on log-return series in favour of asymmetry of the probability distribution describing the process. Using the Akaike Information Criterion (AIC) and the Log-likelihood estimation (LLH) methods as selection criteria, the normal inverse Gaussian (NIG) outperformed four other candidate probability distributions among the class of Generalized Hyperbolic (GH) distributions in describing the heavy tails present in the process. Similarly, the Skewed Student's t (SSt) is the best fit for Bonny Crude Oil and Natural Gas log-returns. The observed volatility measures of these three commodity prices were examined. The Weibull distribution gives the best fit both electricity and crude oil data while the Gamma distribution is selected for natural gas data in the volatility profiles among the five candidate probability density functions (Normal, Lognormal, Gamma, Inverse Gamma and the Inverse Gaussian) considered.

Declaration

I, Felix Noyanim Nwobi, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then: (a). Their words have been rewritten, but the general information attributed to them has been referenced (b). Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References.

Signed:

Felix Noyanim Nwobi

November 2011

Dedication

To

1. My little daughter, Onyenye Ifeanyichukwu Nwobi, who was barely three weeks old when I started this adventure.

2. All those who seek knowledge without minding financial and logistic constraints.

Acknowledgments

For introducing me to symmetry analysis of partial differential equations, for his advice, encouragement and inspiration I am very grateful to Professor Peter G. L. Leach. He has the patience and possesses the rare but wonderful qualities of an editor. For the identification of our given partial differential equation that we analysed in Chapter Three and his numerous suggestions that improved the quality of this Thesis I appreciate Dr. John G. O'Hara.

The friendliness and wonderful efforts of the members of staff of the University of KwaZulu-Natal (UKZN) library particularly in the prompt sourcing and provision of research materials outside the UKZN libraries' system are highly acknowledged. Prominent among these staff are those in the interlibrary loan section of the Westville campus library especially, Mduduzi Zungu.

I take this unique opportunity to thank the following: Dr Kostantinos Andriopoulos (a.k.a. Kostis, my first ever discussion on Leach's lecture notes on Singularity and Symmetry Analyses was with him in my first three weeks of registration) from the University of Patras, Greece, Prof. John N. Nnadi (discussions on PDEs) from Abia State University, Uturu, Nigeria, Emma U. Ohaegbulem from Imo State University, Owerri, Nigeria, Dr. Alfred A. Nnadozie (wonderful friend) from Italy, Stelios Dimas (personal communication on the use of their package SYM) from the University of Patras, Dr. Boniface Mathadi, Patrick Ali and Dr. Nicolette Caister (had a session with her on SYM-MATHEMATICA) from School of Mathematical Sciences, UKZN.

My gratitude also goes to Bro. Johnson Nwokeiwu of Maris Brothers, Durban, and Dr. Remy Nnadozie and family. Remy introduced me to Peter Leach. William Shakespeare told us in his play *Julius Caesar* that "What touches us ourselves shall be last served". With love and affection I appreciate my wife, Ndidiamaka, for her unquantifiable level of endurance, sacrifice and deprivation. For my courage and perseverance, to God be the glory.

Contents

List of Tables

List of Figures

Chapter 1

Preliminaries

Sophus M Lie believed that any natural mathematical theory should be transparent, and that difficulties in mathematics usually arise not from the essence of the problem but from badly conceived definitions. -Yaglom (1988)

1.1 Introduction

We present this thesis in two parts. The first part discusses symmetry analysis as a procedure for finding solutions of differential equations and symmetries of our given partial differential equation. Through Lie point symmetries of this partial differential equation we find closed-form solutions necessary for modeling prices of electricity futures. In the second part we try to construct a "bridge" that links the given partial differential equation to numerical data and attempt to fit some probability distributions. Goodness-of-fit tests are intended to find or identify a class of probability distributions the random processes of which generated the data. In the following sections of this chapter we sketch some of the relevant definitions that are used in the two parts of the Thesis.

1.2 Overview of Symmetries

The perception of symmetry that lies at the core of our conscious life manifests in most abstract and physical situations and it makes nature have a sense of beauty. Due to the common occurrence of symmetries, most people intrinsically feel they understand its concept. The concept is a powerful one, finding applications in such diverse fields as art and science. Symmetry is universal, fascinating and of immense importance. Despite an astonishing variety of shapes (Cantwell 2002), all members of the animal kingdom possess body architectures that can be sorted into about 37 basic types. In Art, paintings by the Dutch artist M C Escher (Schattschneider D. 1990), for example, make extensive use of various symmetry operations. Today most floor tiles and textile designs are based upon the concepts of symmetry. Birkhoff (1933) in his fascinating work attempted to quantify the relationship between symmetry and beauty and as a result developed what he called the "Aesthetic Measure".

Although several definitions of symmetry exist depending upon the application for which it is intended, Weyl (1952) has it that an object is symmetrical if one can subject it to a certain operation and it remains exactly the same after that operation. The object is then said to be invariant with respect to the given operation. Leach (2006) defined symmetry as an operation that leaves invariant that upon which it operates. The symmetry properties of an object can usually be expressed in terms of a set of matrices each of which, when used to transform the various points comprising the object, leaves it unchanged in appearance. To classify the notion of symmetry and its mathematical description Cantwell examined the rotational and reflectional properties of a snowflake. Two types of symmetry can be distinguished, namely discrete and continuous symmetries. A discrete symmetry is one which must be performed as a single operation and cannot be broken up into parts. An example is provided by translations in the plane, as in, for example, floor tiles where specific pattern is repeated at finite intervals and hence invariant under finite translation of precise discreteness in the plane. The equilateral triangle is another example that is invariant under rotation about its centroid by any multiple of 120°. It has six distinct symmetries; the isosceles triangle has two, while the scalene (the triangle with three unequal sides) has only a trivial symmetry (Hydon 2000). As we see in this

Thesis, the number of point symmetries generated from a differential equation depends upon certain conditions specified for that equation.

A continuous symmetry on the other hand is one which depends upon the value of a parameter which can take all values in an interval of the real line. Invariance under a continuous transformation can be demonstrated by the rotation of a circle (or a disc) about an axis through its centre normal to the plane of the circle. No matter the angle through which the circle is rotated, the appearance of the circle remains unchanged. Since the angle can vary continuously, these rotations are known as continuous transformations.

Symmetries provide a systematic means to obtain an enriched understanding of physical phenomena and the associated equations. For instance, Leach and Andriopoulos (2005) and Naicker *et al* (2005) report that knowledge of symmetries of partial differential equations enables researchers to have a completely new way of looking at problems arising in Applied Mathematics especially in modeling problems of mathematical finance. From the definition of a symmetrical transformation one can deduce that every system of partial differential equations with topologically continuous solution sets admits symmetries. This plays a major role in many of the applications associated with differential equations. Such roles include the following:

(i) The deduction of new solutions from known ones;

(ii) The reduction of order of ordinary and partial differential equations;

(iii) The classification of special solutions;

(iv) The classification of families of equations;

(v) The construction of types of equations that admit a prescribed group of transformations;

(vi) The linearisation of equations by invertible transformations; the asymptotics of solutions; and

(vii) Benchmarks for the testing of numerical algorithms.

One of the most important uses of symmetries is their contribution to the reduction of the number of variables of a partial differential equation which is attributable to the interdepen-

dence of symmetry group techniques and the integrability of the equations. As a result of this reduction higher-order equations can usually be reduced to quadratures if there is a sufficient number of symmetries. It is important to recall that there are four approaches to the analysis of the system

$$u_i = f_i(x, u), \qquad i = 1, 2, ..., n. \tag{1.2.1}$$

These are the techniques of dynamical systems, numerical and computational mathematics, singularity analysis and symmetry analysis. The first two approaches Leach (2006) observes, are generally applicable while the latter two are particularly relevant to integrable systems because the results of the analyses provide information about the integrability or otherwise of the system. For the case of integrable systems they provide a procedure for finding the solution. Once the symmetries of the system are known, all other techniques of tackling the problem such as the numerical analysis can then be applied more effectively and with a better basic understanding of the problem.

The three major methods to find symmetries of a system are the classical Lie method (we introduce some of its concepts in section 1.4 and discuss its details in Chapters Two and Three), the nonclassical method and the direct method. The last is a subset of the nonclassical method. For the algorithms and comparisons of these methods the reader is referred to Arrigo *et al* (1993, 1994), Bluman and Cole (1974), Bluman and Kumei (1987), Clarkson (1989a,b; 1995), Clarkson and Kruskal (1989) and Olver (1987).

1.3 Symmetries

The evolution in time of different symmetry types were introduced in the following order (Leach 2003): Point, Contact, Generalized and Nonlocal symmetries. We briefly outline them in the following subsections.

1.3.1 Point symmetry

Let x denote the position of a general point of an object and, if

$$\Gamma : x \longrightarrow \hat{x}(x) \tag{1.3.1}$$

is any symmetry, then we assume that $\hat{x}$ is infinitely differentiable with respect to x. Moreover, since Γ^{-1} is also a symmetry, x is infinitely differentiable with respect to $\hat{x}$. Thus Γ is a (C^∞) diffeomorphism, that is a differentiable mapping which has a differentiable inverse (see, for instance, Cantwell (2002) and Hydon (2000)).

Consider the infinitesimal transformation for a function $f(x, u)$, x independent and u dependent variables written as

$$\bar{x} = x + \varepsilon\xi(x, u)$$
$$\bar{u} = u + \varepsilon\eta(x, u),$$

where ε is the parameter of smallness, the infinitesimal. When this transformation is written in terms of a differential operator

$$\Gamma = \xi(x, u)\frac{\partial}{\partial x} + \eta(x, u)\frac{\partial}{\partial u} \tag{1.3.2}$$

we refer to ξ and η as coefficient functions because of the way they occur in the definition of Γ in (1.3.2). Under the action of the infinitesimal transformation generated by Γ a function $f(x, u)$ becomes

$$\begin{aligned} \hat{f} &= (1 + \varepsilon\Gamma)f(x, u) \equiv f(x + \varepsilon\xi, u + \varepsilon\eta) \\ &= f(x, u) + \varepsilon\left(\xi\frac{\partial f}{\partial x} + \eta\frac{\partial f}{\partial u}\right). \end{aligned}$$

When the function $f(x, u)$ has a generator Γ under which it is invariant (unchanged), that is,

$$\hat{f}(\hat{x}, \hat{u}) = f(x, u),$$

or

$$\xi\frac{\partial f}{\partial x} + \eta\frac{\partial f}{\partial u} = 0,$$

then Γ is called the symmetry of $f(x,u)$ (see further illustrations in Sections (2.2) and (2.3) of this Thesis). If ξ and η are coefficient functions of x and u only, then we have a point transformation (1.3.2) and the symmetry is called a point symmetry.

For the generator to represent a symmetry of a differential equation, we need to extend the group generated by Γ into the derivatives of, say, the first derivative

$$\begin{aligned}
\frac{d\hat{u}}{d\hat{x}} &= \frac{d(u+\varepsilon\eta)}{d(x+\varepsilon\xi)} = \frac{du+\varepsilon d\eta}{dx+\varepsilon d\xi} \\
&= \left(\frac{du}{dx}+\varepsilon\frac{d\eta}{dx}\right)\left(1+\varepsilon\frac{d\xi}{dx}\right)^{-1} \\
&= \left(u^{'}+\varepsilon\eta^{'}\right)\left(1+\varepsilon\xi^{'}\right)^{-1} \\
&= \left(u^{'}+\varepsilon\eta^{'}\right)\left(1-\varepsilon\xi^{'}+\varepsilon^{2}\xi^{'2}-\cdots\right) \\
&= u^{'}+\varepsilon\left(\eta^{'}-u^{'}\xi^{'}\right)+O\left(\varepsilon^{2}\right),
\end{aligned}$$

where $O(\varepsilon^2)$ stands for the sum of all terms of second order or greater which is terminated at $O(\varepsilon)$ since ε is infinitesimal and prime denotes total differentiation with respect to x.

1.3.2 Contact symmetry

If, in addition to the dependent and independent variables in the point symmetry given in Subsection 1.3.1 above, the coefficient functions depend upon u', so that

$$\Gamma = \xi(x,u,u')\,\partial_x + \eta(x,u,u')\,\partial_u + \zeta(x,u,u')\,\partial_{u'}. \tag{1.3.3}$$

Since the first extension $\Gamma^{[1]}$ does not contain $u^{''}$ we require that, in

$$\begin{aligned}
\zeta(x,u,u') &= \eta'(x,u,u') - u'\xi'(x,u,u') \\
&= \frac{\partial\eta}{\partial x}+u'\frac{\partial\eta}{\partial u}+u''\frac{\partial\eta}{\partial u'}-u'\left(\frac{\partial\xi}{\partial x}+u'\frac{\partial\xi}{\partial u}+u''\frac{\partial\xi}{\partial u'}\right) \\
&= \frac{\partial\eta}{\partial x}-u'\frac{\partial\xi}{\partial x}+u'\left(\frac{\partial\eta}{\partial u}-u'\frac{\partial\xi}{\partial u}\right)+u''\left(\frac{\partial\eta}{\partial u'}-u'\frac{\partial\xi}{\partial u'}\right),
\end{aligned} \tag{1.3.4}$$

the last term is zero. This implies that Γ is a contact symmetry only if the coefficient of u'' is zero, that is

$$\frac{\partial \eta}{\partial u'} = u' \frac{\partial \xi}{\partial u'}.$$

1.3.3 Generalized symmetries

If the coefficient functions in Γ depend upon the derivatives possibly up to the highest admissible derivative[1], that is,

$$\Gamma = \xi\left(x, u, u', ...\right)\partial_x + \eta\left(x, u, u', ...\right)\partial_u. \tag{1.3.5}$$

This produces some technical problems in that (x, u) space is extended to $\left(x, u, u', \ldots\right)$ space. It was to overcome this problem that the concept of contact transformation was introduced. However, an nth-order transformation is of the form

$$\begin{aligned}\Gamma = {} & \xi\left(x, u, u', ..., u^{(n)}\right)\partial_x + \eta\left(x, u, u', ..., u^{(n)}\right)\partial_u + \zeta_1\left(x, u, u', ..., u^{(n)}\right)\partial_{u'} \\ & + ... + \zeta_n\left(x, u, u', ..., u^{(n)}\right)\partial_{u^{(n)}},\end{aligned} \tag{1.3.6}$$

where substitution for nth derivatives has been made with one of the nth-order equations. We note here that a point transformation is of order zero, contact symmetry is of first order, while that of the generalized symmetry is of the nth-order of the differential equation. For details of generalized symmetries and its historical evolution please see Olver (1993).

1.3.4 Nonlocal symmetry

Nonlocal symmetries are those symmetries in which the coefficient functions depend upon integrals containing the derivatives and the dependent (and the independent) variables of the differential equation. These symmetries are important as they have been linked with integrable models (Olver, 1993).

A *hidden* symmetry is a Lie point symmetry which appears in the given differential equation after a change of order and which does not have a point counterpart in the given equation. These

[1] In the case of an nth-order ordinary differential equation this means up to the $(n-1)$th derivative but this constraint falls away with partial differential equations.

hidden symmetries that manifest themselves as nonlocal symmetries of the original equation are suspected (Govinder and Leach, 1996) to be in their local form (where the transformations depend on the variables of the equation) as a point symmetry following a reduction (or increase) in the order of an equation. To summarize, the coefficient functions can depend upon integrals of integrands of which are functions of x, u and $u^{'}\left(u^{''},\ldots\right)$.

Two types of hidden symmetries are identifiable, viz. $TypeI$ and $TypeII$. A hidden symmetry of Type I occurs when the order of a given differential equation is increased, whereas decreasing the order of a given equation gives rise to a Type II hidden symmetry (Abraham-Shrauner and Guo, 1994). For instance, if a differential equation has p symmetries and its descendant after reduction of order produced $q(q > p)$ symmetries, then these $q - p$ symmetries were hidden and are referred to as Type II hidden symmetries. Hence equations possessing no Lie point symmetries may be reduced to quadratures when an increase in the order of the equation results in a Type I hidden symmetry (Abraham-Shrauner *et al* 1995). These nonlocal symmetries are referred to as "lost" because they cannot be determined via the infinitesimal generators of a differential equation. However, due to their content, omitting them would mean discarding a number of physical interesting solutions. For methods and procedures of finding these nonlocal symmetries, see, for instance, Abraham-Shrauner and Guo (1992), Abraham-Shrauner and Govinder (2006) and Anco and Bluman (1996).

1.4 Lie Group Theory

We define some of the key concepts that are used throughout this Thesis as powerful tools to guide computations. Details of items defined hereunder can be found in Bluman and Kumei (1989), Olver (1986) and Ovsyannikov (1982).

1.4.1 Group

A group G is a set of elements (numbers, vectors, octonions, etc.) with a law of composition ϕ between elements satisfying the following properties:

Closure Property: If X and Y are elements of G, then $\phi(X, Y)$ is an element of G.

Associative Property: For any elements X, Y and Z of G

$$\phi(X, \phi(Y, Z)) = \phi(\phi(X, Y), Z). \tag{1.4.1}$$

Identity Property: There exists a unique identity element I of G such that for any element X of G,

$$\phi(X, I) = X = \phi(I, X). \tag{1.4.2}$$

Inverse Element: For any element X of G there exists a unique element in G denoted by X^{-1} such that

$$\phi\left(X, X^{-1}\right) = I = \phi\left(X^{-1}, X\right), \tag{1.4.3}$$

the element X^{-1} is called the inverse of X.

If two elements X and Y of a group (in G) satisfy the property

$$\phi(X, Y) = \phi(Y, X), \tag{1.4.4}$$

they are said to commute.

1.4.2 Group of transformations

A set of transformations

$$\bar{x} = X(x, \varepsilon) \tag{1.4.5}$$

defined for each x in $D \subset \mathbb{R}$, depending upon the parameter ε lying in the set $S \subset \mathbb{R}$ with $\phi(\varepsilon, \delta)$ defining a composition of parameters ε and δ in S, forms a group of transformations on D if

(i) for each parameter ε in S the parameter is one-to-one onto D.

(ii) S, with the law of composition ϕ, forms a group.

(iii) $\bar{x} = x$ when $\varepsilon = I$, that is

$$X(x, I) = x. \tag{1.4.6}$$

(iv) If $\bar{x} = X(x, \varepsilon), \bar{\bar{x}} = X(\bar{x}, \delta)$, then

$$\bar{\bar{x}} = X(x, \phi(\varepsilon, \delta)). \tag{1.4.7}$$

1.4.3 Lie Group of transformations

A one-parameter (ε) Lie group of transformations is a group of transformations which, in addition to the properties stated above, satisfies the following conditions:

(i) ε is a continuous parameter, that is, S is an interval in $\mathbb{R}$ (without loss of generality $\varepsilon = 0$ corresponds to the identity element I).

(ii) X is infinitely differentiable with respect to x in D and an analytic function of ε in S.

(iii) $\phi(\varepsilon, \delta)$ is an analytic function of ε and δ and $\varepsilon \in S, \quad \delta \in S$.

1.4.4 Lie algebra of operators

An r-parameter Lie transformation group has associated r group operators $\Gamma_1, \ldots, \Gamma_r$, which are linearly independent and form an r-dimensional vector space over $\mathbb{R}$ with the additional structure of closure under the operation of taking a Lie Bracket. Let

$$\Gamma_\alpha = \xi_{1,\alpha}(x) \frac{\partial}{\partial x_1} + ... + \xi_{n,\alpha}(x) \frac{\partial}{\partial x_n}, \quad \alpha = 1, ..., r$$

and

$$\Gamma_\beta = \xi_{1,\beta}(x) \frac{\partial}{\partial x_1} + ... + \xi_{n,\beta}(x) \frac{\partial}{\partial x_n}, \quad \beta = 1, ..., r,$$

be any two operators. Then the first-order differential operator,$[\Gamma_\alpha, \Gamma_\beta]_{LB}$, is defined by

$$[\Gamma_\alpha, \Gamma_\beta]_{LB} = \sum_{i=1}^{n} (\Gamma_\alpha(\xi_{i,\beta}) - \Gamma_\beta(\xi_{i,\alpha})) \frac{\partial}{\partial x_i} = \Gamma_\alpha \Gamma_\beta - \Gamma_\beta \Gamma_\alpha \in L. \tag{1.4.8}$$

The Lie Bracket $[\Gamma_\alpha, \Gamma_\beta]_{LB}$ ($= [\Gamma_\alpha, \Gamma_\beta]$ from now on without loss of generality) satisfies the following axioms:

(i) Bilinearity: $[k_1\Gamma_\alpha + k_2\Gamma_\beta, \Gamma_\gamma] = k_1[\Gamma_\alpha, \Gamma_\gamma] + k_2[\Gamma_\beta, \Gamma_\gamma]$
where k_1 and k_2 are constants.

(ii) Skew Symmetric: $[\Gamma_\alpha, \Gamma_\beta] = -[\Gamma_\beta, \Gamma_\alpha]$.

(iii) Jacobi Identity:$[\Gamma_\alpha, [\Gamma_\beta, \Gamma_\gamma]] + [\Gamma_\beta, [\Gamma_\gamma, \Gamma_\alpha]] + [\Gamma_\gamma, [\Gamma_\alpha, \Gamma_\beta]] = 0$
for all vectors $\Gamma_\alpha, \Gamma_\beta, \Gamma_\gamma \subset L$.

Any vector space of operators satisfying the above three axioms is called a *Lie algebra* of operators. A Lie algebra of operators contains all the information necessary to reconstruct a Lie group. If $[\Gamma_\alpha, \Gamma_\beta] = 0$, then the generators Γ_α and Γ_β are said to *commute*. In particular every generator commutes with itself, that is, $[\Gamma_\alpha, \Gamma_\alpha] = 0$. If all the elements (generators) of L (basis) commute, then L is called an *Abelian Lie algebra.*
The finite-dimensional Lie algebra L^r is usually indicated by the *vector basis* $\{\Gamma_i\}$ *in the space* L^r. In this case and in accordance with the axiom of bilinearity the operation of the vector basis in L^r is fully defined by the table of Lie Brackets, that is, by an $r \times r$ square matrix in which the Lie Bracket $[\Gamma_i, \Gamma_j]\,(i, j = 1, ..., r)$ is the (i, j)th element of the square matrix (see, for example, Table 2.1 of Chapter Two).
NB: The dimension dim L of the Lie algebra is the dimension of the vector space L. We therefore use the symbol L^r to denote an $r-$dimensional Lie algebra.

1.4.5 Linear combination

Suppose we let $\Gamma \in L^r$ be any operator and c_k be any constant. Then $\Gamma = c_1\Gamma_1 + \cdots + c_k\Gamma_k$ is called a linear combination of the $\Gamma_k, r = 1, ..., k$. For example, if we let $\Gamma_1 = \frac{\partial}{\partial x}$, $\Gamma_2 = \frac{\partial}{\partial u}$, $\Gamma_3 = x\frac{\partial}{\partial x} - u\frac{\partial}{\partial u}$ then we refer to $\Gamma = c_1\Gamma_1 + c_2\Gamma_2 + c_3\Gamma_3$ as a linear combination of these operators.

1.4.6 Basis of the vector space

Let L^r be a finite-dimensional Lie algebra and suppose that $\Gamma_\alpha = \xi_{i\alpha}(x)\dfrac{\partial}{\partial x_i}, \alpha = 1, ..., r$ be a basis of the vector space L^r . In particular $[\Gamma_\alpha, \Gamma_\beta] \in L^r$. Hence $[\Gamma_\alpha, \Gamma_\beta] = C^k_{\alpha\beta}\Gamma_k, \quad \alpha, \beta, k = 1, ..., r$. The constant coefficients $C^k_{\alpha\beta}$ are called the *structure constants* of the algebra L^r .

1.4.7 Optimal system

Let the finite Lie algebra L^r be spanned by the operators Γ provide a possibility to find invariant solutions of a differential equation based on one-dimensional subalgebras of the algebra L^r i.e., on any operator $\Gamma \in L^r$. However, there are infinite number of one-dimensional subalgebras of L^r and since an arbitrary operator can be written from L^r as a linear combination of the operators and hence depends on r arbitrary constants[2]. The desire to minimize the search for invariant solutions by finding nonequivalent branches of solutions leads to the concept of an optimal system. The set of representatives of all classes of similar operators $\Gamma \in L^r$ is an optimal system of one-dimensional subalgebras. Similarly, when all invariant solutions are obtained in principle by constructing the invariant solution for each member of the optimal system of subalgebras, the set of invariant solutions obtained in this way is an optimal system of invariant solutions (Ibragimov 2009).

1.4.8 Linear span

Suppose we have operators $\Gamma_1, \Gamma_2, ..., \Gamma_s$. Then their linear span is denoted by $\langle \Gamma_1, \Gamma_2, ..., \Gamma_s \rangle$. For example, given L^r with basis $\Gamma_\alpha = \xi_{i\alpha}(x)\dfrac{\partial}{\partial x_i}, \alpha = 1, ..., r$, the span is denoted by $L_\alpha = \langle \Gamma_1, \Gamma_2, ..., \Gamma_r \rangle$.

[2]L. V. Ovsyannikov (1982) introduced the concept of optimal system of subalgebras (in order to make the problem of infinite number of $1-$ dimensional subalgebras manageable) by noting that if two subalgebras are similar, i.e., connected with each other by a transformation of symmetry group, then their corresponding invariant solutions are connected with each other by the same transformation.

1.4.9 Subalgebra

Suppose that L is a Lie algebra. A subspace $K \subset L$ of a vector space L is called a *subalgebra* of the Lie algebra L if K is closed under the Lie Bracket $[K, K] \subset K$. In other words a Lie algebra L^q is a subalgebra of $L^r (q < r)$ if $L^q \subset L^r$.

1.4.10 Ideal

The subalgebra L^q is called an *ideal of subalgebras* of L^r if, for any $\Gamma_\alpha \in L^q$ and $\Gamma_\beta \in L^r$ the Lie Bracket $[\Gamma_\alpha, \Gamma_\beta] \in L^q$. For illustrative examples of subalgebras the reader is referred to Cantwell (2002, p.128).

1.4.11 Solvable Lie algebras

Consider the Lie algebra corresponding to $\Gamma_\alpha, \Gamma_\beta$ and Γ_γ. If the sequence of subalgebras $L^0, L^1 = \Gamma_\gamma, L^2 = \Gamma_\alpha, \Gamma_\gamma, L^3 = \Gamma_\alpha, \Gamma_\beta, \Gamma_\gamma$ has the property that each item in the sequence is an ideal of the next item, then the subalgebra is *solvable.*

Definition 1.4.1 (Cantwell)**.** The Lie algebra L^q is a q-dimensional solvable Lie algebra if there exists a chain of subalgebras

$$L^0 \subset L^1 \subset L^2 \subset \cdots \subset L^{q-1} \subset L^q \tag{1.4.9}$$

such that L^k is a k-dimensional Lie algebra and L^{k-1} is an ideal of L^k for $k = 1, 2, ..., q$. Here L^0 is the null ideal with no operators.

1.4.12 Lie equation

Consider the group transformation $\bar{x}_i = f_i(x, \varepsilon), i = 1, ..., n$ in n-dimensional space with generator $\Gamma = \xi_i(x) \dfrac{\partial}{\partial x_i}$, where $\xi_i(x) = \dfrac{\partial f_i(x, \varepsilon)}{\partial \varepsilon} \mid_{\varepsilon=0}$ is defined by integrating the following general differential equation called the Lie equation, $\dfrac{\partial \bar{x}_i}{\partial \varepsilon} = \xi_i(\bar{x}), \bar{x}_i \mid_{\varepsilon=0} = x_i$. It is reported in Yaglom (1988) that Lie's main result is the proof that it is always possible to assign to a continuous

group (Lie group) a corresponding Lie algebra and vice versa. Thus for the real special linear group, $SL(n, R)$, there is a corresponding Lie algebra, $sl(n, R)$ and for the special orthogonal group, $SO(n, R)$, there is the corresponding special orthogonal algebra $so(n, R)$.

1.4.13 Symmetry group

Mathematical objects such as functions, differential equations, surfaces etc are closely related to the concept of a group as well as invariance and symmetry. Let G be a set of invertible transformations T. Then any given object M on which T acts on does not change, that is, $T : M \longrightarrow M$. Mathematically $T : M \longrightarrow M$ contains the following transformation,

(i) Identity $\boldsymbol{I}$.

(ii) Inverse T^{-1} and

(iii) Product T_1T_2,

where T_1 and $T_2 \in G$. G is called a group or more precisely a symmetry group of the object M.

1.4.14 Extended Lie group transformations

An extended Lie group of transformations of a partial differential equation is a continuous group of transformations which act on an extended space of variables that include the parameters of the equation in addition to independent and dependent variables. An extended group of transformations represents a particular case of an equivalence group that preserves the class of partial differential equations that have the same differential structure but with arbitrary functions having different forms. The approach to find these equivalence transformation groups with the use of the Lie infinitesimal technique was introduced by Ovsyannikov (1982). He suggested using Lie's infinitesimal criterion in the properly extended space of variables including dependent and independent variables, arbitrary functions and their derivatives. I. S. Akhatov and his group in 1989 further developed Ovsyannikov's original method. The generalization of this idea has appeared in several research papers, see, for instance, Romano and Torrisi (1999)

and the references therein. The transformations in the extended space of variables obtained by the addition of parameters to the list of independent variables have been used in the context of the renormalization group (RG) symmetries by Kovalev *et al*(1998) and Shirkov and Kovalev (2001).

1.5 Probability Concepts

In this section we present some necessary definitions and sketch descriptions of probability and probability distributions that are applied in Chapters Four and Five. Much information on probability, its theory and modeling abounds in the mathematical and statistical literature. In particular, the reader may refer to Feller (1957), Ross (2000), Dineen (2005) and Shreve (2004) for details.

1.5.1 Probability spaces

Let Ω be a nonempty set and let $\mathcal{F}$ be a collection of subsets of Ω.

Definition 1.5.1. A σ-algebra (sometimes called a σ-field) is a collection $\mathcal{F}$ of subsets of Ω with the following properties:

(i)
$$\phi, \Omega \in \mathcal{F}; \tag{1.5.1}$$

(ii) If
$$A \in \mathcal{F}, \quad \text{then} \quad A^{'} \in \mathcal{F}; \tag{1.5.2}$$

(iii) If
$$A_1, A_2, ... \in \mathcal{F}, \quad \text{then} \quad \bigcup_{k=1}^{\infty} A_k, \bigcap_{k=1}^{\infty} A_k \in \mathcal{F}. \tag{1.5.3}$$

The points in $\mathcal{F}$ being subsets of Ω are called $\mathcal{F}$-events or $\mathcal{F}$-*measurable sets.* A pair $(\Omega, \mathcal{F})$ is called a *measurable space.*

Definition 1.5.2. If $\mathcal{F}$ is a $\sigma-$algebra of subsets of Ω, then $\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ is a *probability measure* if

(i)

$$\mathbb{P}(\phi) = 0 \quad \text{and} \quad \mathbb{P}(\Omega) = 1; \tag{1.5.4}$$

(ii)

$$A_1, A_2, ... \in \mathcal{F}, \quad \text{then} \quad \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_k). \tag{1.5.5}$$

It therefore follows that, if $A, B \in \mathcal{F}$, then $A \subseteq B$ implies that $\mathbb{P}(A) \subseteq \mathbb{P}(B)$.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space.* We note here that the probability space is the proper setting for mathematical theory. This means that we must firstly carefully identify an appropriate $(\Omega, \mathcal{F}, \mathbb{P})$ whenever we try to solve problems.
Having defined the elements in the σ-algebra from a set-theoretic viewpoint, we now consider them as events. Associated with each event is information, which in the financial world, increases as time increases. A sample space $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$ is the set of all possible outcomes of some experiment, ε , while the σ-algebra $\mathcal{F}$ represents the events that are observed and can be recorded when the experiment is performed. In other words it is the information we receive upon performing the experiment. Thus after the experiment we can observe whether or not $A = \{\omega_{j1}, \omega_{j2}, ..., \omega_{jm}; j = 1, 2, ..., N\} \in \mathcal{F}$ occurred. If $\mathcal{F}_1$ and $\mathcal{F}_2$ are two σ-algebras on Ω, then $\mathcal{F}_1 \subset \mathcal{F}_2$ if and only if $\mathcal{F}_2$ contains more information than $\mathcal{F}_1$.

Definition 1.5.3. Let $(\Omega, \mathcal{F})$ be a measurable space.

(i) A discrete filtration on $(\Omega, \mathcal{F})$ is an increasing sequence of $\sigma-$algebras $(\mathcal{F})_{k=1}^{\infty}$ such that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_i \subset \cdots \subset \mathcal{F}$.

(ii) A continuous filtration on $(\Omega, \mathcal{F})$ is a set of $\sigma-$algebras $(\mathcal{F})_{t \in I}$, where I is an interval in $\mathbb{R}$ such that for all $t, s \in I, t < s$, we have $\mathcal{F}_t \subset \mathcal{F}_s \subset \mathcal{F}$.

We call $\mathcal{F}_k$ (respectively $\mathcal{F}_t$) the history up to time k (respectively time t).

1.5.2 Random variables

Unlike the point of time of the impact on the ground of a stone dropped from certain altitude being known before execution of the experiment (Newton's Laws), quantities of complex systems (such as stocks, commodity prices etc) are nondeterministic. However, their values may be predicted under uncertainties. Contrary to the falling stone, data which cannot be described successfully by a deterministic mechanism can be modeled by random variables.

Definition 1.5.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is a real-valued function X defined on Ω with the property that for every Borel subset $\mathcal{B}$ of $\mathbb{R}$ the subset of Ω given by

$$\{X \in \mathcal{B}\} = \{\omega \in \Omega; X(\omega) \in \mathcal{B}\}$$

is in the $\sigma-$algebra $\mathcal{F}$.

For properties of random variables the reader is referred to the references listed at the beginning of this section. A random variable X is a numerical quantity the value of which is determined by the random experiment choosing $\omega \in \Omega$. The properties of probability measures $\mathbb{P}(\mathcal{B})$ for every Boral subset $\mathcal{B}$ of $\mathbb{R}$. Denoting the distribution measure of X under $\mathbb{P}$ by μ_X, we have for the set of all probabilities,

$$\mu_X[a,b] = \mathbb{P}(\omega_i : a \leq X_\omega \leq b), \quad -\infty < a \leq b < \infty \tag{1.5.6}$$

a measure that determines the distribution of X. In other words the distribution is defined by the probabilities of all events which depend upon X.
We can describe the distribution function of a random variable in terms of its *cumulative distribution function* (cdf)

$$\mathbb{F}(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R} \tag{1.5.7}$$

The $\mathbb{F}(x)$ is monotonically increasing and converges for $x \longrightarrow -\infty$ to 0 and for $x \longrightarrow \infty$ to 1. If there is a function, p, such that the probabilities can be computed by means of an integral

$$\mathbb{P}(a \leq X \leq b) = \int_a^b p(x)\, dx, \tag{1.5.8}$$

p is called the *probability density* or, simply, density of X. When the cumulative function is a primitive of p,

$$\mathbb{F}(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} p(y)\, dy. \tag{1.5.9}$$

Thus $p(x)$ is a measure of the likelihood that X takes values close to x and the pdf

$$f(x) = \frac{\partial}{\partial x}\mathbb{F}(x), \quad \text{i.e.,} \quad \phi(x)\, dx = \Phi'(x)\, dx. \tag{1.5.10}$$

The most important family of distributions with densities is the family of *normal distribution.* It is characterised by two parameters the mean, μ, and the variance,σ^2. The density is given by

$$f_X\left(x; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \tag{1.5.11}$$

for $-\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 \geq 0$, and

$$f_Z(z; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}. \tag{1.5.12}$$

The distribution with density (1.5.12) is called the standard normal distribution for which the mean and variance of Z are zero and one respectively and that in (1.5.11) shows that X is a normal random variable distributed as $X \sim N(\mu, \sigma^2)$.

Closely related to the normal distribution is the log-normal distribution that is very important in modeling commodity prices. Let X be a positive random variable with natural logarithm of which, $\ln(X) \sim N(\mu, \sigma^2)$. We say that X is log-normally distributed with parameters μ and σ^2. Its cumulative distribution function follows from (1.5.9) as

$$\mathbb{F}(x) = \mathbb{P}(\ln X \leq x) = \Phi\left(\frac{\ln X - \mu}{\sigma}\right), \quad x > 0. \tag{1.5.13}$$

Hence

$$f_X\left(\ln x; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right\}. \tag{1.5.14}$$

Most other probability distributions especially the continuous types are either special cases or derivable from the normal distribution.

1.5.3 Moments of random variables

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *first moment* known as the *mathematical expectation* or the *mean* $\mathbb{E}(X)$ of a real random variable X is a measure for the location of the distribution of X. If X has a density $p(x)$, then its expectation is defined as

$$\mathbb{E}(X) = \begin{cases} \int_{-\infty}^{\infty} xp(x)dx & \text{if } x \text{ is continuous} \\ \sum_{i=1}^{\infty} x_i p(x_i) & \text{if } x \text{ is discrete,} \end{cases} \tag{1.5.15}$$

where $\int_{\Omega} X_{(\omega)} d\mathbb{P}(\omega)$ is a Lebesgue integral. A measure of the dispersion of a random variable X around its mean is given by the *variance* $\mathbb{V}(X)$ as

$$\mathbb{V}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 p(x)\, dx. \tag{1.5.16}$$

The log-normally distributed random variable X defined in (1.5.14) has mean

$$\mathbb{E}(X) = e^{\mu + \frac{1}{2}\sigma^2} \tag{1.5.17}$$

and variance

$$\mathbb{V}(X) = e^{2\mu+\sigma^2} \left(e^{\sigma^2} - 1\right). \tag{1.5.18}$$

Details of other relevant definitions, theorems and axioms can be found in the references cited earlier.

1.6 Stochastic Processes

Definition 1.6.1. A *stochastic process* $X(t)$ is a family of random variables $\{X_t(\gamma), t \in T, \gamma \in \Omega\}$, i.e., for each t in the index set T , $X(t)$ is a random variable. Here we interpret t as time and call $X(t)$ the state of the process at time t.

Since a stochastic process is a family of random variables, its specification is analogous to that for random vectors.

1.7 Objectives

This Thesis is intended to address some fundamental issues facing contemporary mathematics and statistics−issues of flexibility, adaptability and applicability to our environment. It is known that mathematical models grow out of equations that determine how a system changes from one state to the next and/or how one variable depends upon the value/state of other variables (state equations). It is also known that statistical models include the characterisation of numerical data, estimation of the probabilistic future behaviour of a system based upon its past behaviour, extrapolation, interpolation of data based on some goodness-of-fit, error estimates of observations and spectral analysis of data or model generated output. A well-known academic statistician, George Box, once said that "All models are wrong, but some are useful." We see how some of these mathematical models can be turned into useful tools for solving societal problems. Mathematicians are justifiably attracted by the beauty of their subject which Bertrand Russell characterised as cold and austere like the beauty of sculpture. While context obscures structure in mathematics, context provides meaning in data analysis. There is truism in mathematical theorems whereas statistical methods are sometimes effective when used with skill. These points are demonstrated in this Thesis.

Chapter 2

Lie Point Symmetries of Evolution Equations

2.1 Introduction

In this Chapter we introduce the notion of evolution equations and their method of solution through symmetry analysis. We give a typical example of an evolution equation relevant to this study in order to appreciate the algorithm involved in finding solutions through the symmetry method. We start by defining the evolution partial differential equation.

Definition 2.1.1 (Evolution equation)**.** An evolution partial differential equation is an equation involving an unknown function of several variables that includes time, t, as one of the independent variables.

A second-order evolution partial differential equation in one dependent and two independent variables is an equation of the form

$$F\left(x,t,u,u_x,u_t,u_{xx}\right)=0, \quad (x,t)\in D, \tag{2.1.1}$$

where, as is indicated, the independent variables x and t lie in some given domain $D\in\mathbb{R}^2$. By this definition, u is the dependent variable while the $x-t$ domain D in $\mathbb{R}^2$, on which the problem is defined, is called the space-time domain. By a solution of (2.1.1) we mean a twice

continuously differentiable function $u = u(x,t)$, defined on D which, when substituted into (2.1.1), reduces (2.1.1) to an identity on the domain D. The function $u = u(x,t)$, is assumed to belong to a set of all twice continuously differentiable functions on $\mathbb{R}$ that vanish at infinity so that calculation of its first and second derivatives and the substitution of these derivatives into (2.1.1) is defined.

There are four standard approaches to the solution of an evolution partial differential equation (in fact the general differential equation, whether linear or nonlinear). These approaches which have extensive literatures as we have observed in Section 1.2 of Chapter 1. The singularity analysis is synonymous with Painlevé while the symmetry analysis is associated with Sophus Lie. In this Thesis we concentrate on the method of symmetry analysis of evolution partial differential equations (hereinafter referred to as partial differential equations without loss of generality).

2.2 Lie Groups of Transformations

The idea of a group has been introduced in Section 1.4. Consider $u = u(x,t)$, that is to say

$$f = f(x,t,u). \tag{2.2.1}$$

If $f : X \longrightarrow u$ is a smooth function from $X \simeq \mathbb{R}^p$ to $u \simeq \mathbb{R}^q$ so that

$$u = f(x) = \left(f'(x), ..., f^q(x)\right)$$

is a partial differential equation with q dependent variables (here u is the dependent variable) and p independent variables. We let $p = 2$ such that x and t here represent the two independent variables. Then the general vector field on $X \times U \simeq \mathbb{R}^2 \times \mathbb{R}$. Here a point transformation is a diffeomorphism

$$\Gamma : (x,t,u) \mapsto \left(\hat{x}(x,t,u), \hat{t}(x,t,u), \hat{u}(x,t,u)\right). \tag{2.2.2}$$

This transformation maps the surface $u = u(x,t)$ to the surface parametrized by x and t as follows

$$\begin{aligned} \hat{x} &= \hat{x}(x,t,u) \\ \hat{t} &= \hat{t}(x,t,u) \\ \hat{u} &= \hat{u}(x,t,u). \end{aligned} \tag{2.2.3}$$

The infinitesimal transformations of these variables in Taylor series for the Lie group action (see Cantwell (2002) and Hydon (2000) for details) are

$$\left.\begin{aligned} \hat{x} = x + \varepsilon\xi(x,t,u) + o(\varepsilon^2) = x + \varepsilon\Gamma x + \cdots \\ \hat{t} = t + \varepsilon\tau(x,t,u) + o(\varepsilon^2) = t + \varepsilon\Gamma t + \cdots \\ \hat{u} = u + \varepsilon\eta(x,t,u) + o(\varepsilon^2) = u + \varepsilon\Gamma u + \cdots \end{aligned}\right\} \tag{2.2.4}$$

with the vector fields which span the associated Lie algebra, called the *generators* of the *infinitesimal transformation* (2.2.4), Γ, so that

$$\Gamma = \xi(x,t,u)\frac{\partial}{\partial x} + \tau(x,t,u)\frac{\partial}{\partial t} + \eta(x,t,u)\frac{\partial}{\partial u} \tag{2.2.5}$$

and

$$\left.\begin{aligned} \hat{x} = (1+\varepsilon\Gamma)x = x + \varepsilon\xi \\ \hat{t} = (1+\varepsilon\Gamma)t = t + \varepsilon\tau \\ \hat{u} = (1+\varepsilon\Gamma)u = u + \varepsilon\eta \end{aligned}\right\}. \tag{2.2.6}$$

In particular the components of Γ at (x,t,u) are

$$\xi(x,t,u) = \left.\frac{\partial\hat{x}}{\partial\varepsilon}\right|_{\varepsilon=0}, \quad \tau(x,t,u) = \left.\frac{\partial\hat{t}}{\partial\varepsilon}\right|_{\varepsilon=0}, \quad \eta(x,t,u) = \left.\frac{\partial\hat{u}}{\partial\varepsilon}\right|_{\varepsilon=0}. \tag{2.2.7}$$

We recall that ξ, τ and η occur in the definition of Γ and are called *coefficient functions,* that is, the components of the *tangent vector* (*vector field*) Γ are exactly ξ, τ and η. The operator Γ is called the generator of the infinitesimal transformation.

Example 2.2.1. Consider the transformation variables (x,u) in the $xu-$ plane (say, the rotational symmetry of a circle by an arbitrary angle, ε mathematically as a transformation) given by

$$\begin{aligned} \hat{x} &= x\cos\varepsilon - u\sin\varepsilon \\ \hat{u} &= x\sin\varepsilon + u\cos\varepsilon, \end{aligned} \tag{2.2.8}$$

where $\varepsilon \in \mathbb{R}$. The object moves under the action of the transformation while the reference axis remains fixed. The transformation (2.2.8) can be written as

$$\begin{pmatrix} \hat{x} \\ \hat{u} \end{pmatrix} = (1 - \varepsilon J) \begin{pmatrix} x \\ u \end{pmatrix}, \tag{2.2.9}$$

where (2.2.9) is the 2×2 symplectic matrix and

$$J := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.2.10}$$

A reflectional symmetry is also expressed in a similar way. The rotations have determinant of $J = +1$ while the reflections have determinant of $J = -1$. Note that for each ε these transformations constitute a rotation. Hence this family of transformations forms a Lie group known as a *rotation group*. The image point $(\hat{x}, \hat{u})$ can be determined by rotating the radius vector to the source point (x, u) counterclockwise through the angle, ε. The identity element is given by $\varepsilon = 0$ and the inverse transformation by $-\varepsilon$.

2.3 The Infinitesimal Transformations

In this Section we are interested in determining the nth extension of the generator of an infinitesimal transformation. We have looked at a specific example, namely, invariance under rotation and reflection in a plane. We now consider the general infinitesimal transformation which is another way to look at a transformation group.

2.3.1 One dependent and one independent variable

We begin by assuming that for a system $u = f(x)$ there is an arbitrary point for the case of one dependent variable, u, and an independent variable, x. Let the infinitesimal transformation be

$$\left.\begin{aligned} \hat{x} &= x + \varepsilon\xi \\ \hat{u} &= u + \varepsilon\phi \end{aligned}\right\}, \tag{2.3.1}$$

where ξ and ϕ are arbitrary functions. It so happens (Leach (2006), Bluman and Kumei (1989), Bluman and Anco (2002), Dresner (1999) and Stephani (1989)) that, if the differential operator

$$\Gamma = \xi(x,u)\frac{\partial}{\partial x} + \phi(x,u)\frac{\partial}{\partial u} \tag{2.3.2}$$

is a generator of an infinitesimal point transformation, then its extension up to the nth derivative is

$$\Gamma^{[n]} = \xi\frac{\partial}{\partial x} + \phi\frac{\partial}{\partial u} + \phi'\frac{\partial}{\partial u'} + \cdots + \phi^{(n)}\frac{\partial}{\partial u^{(n)}}, \tag{2.3.3}$$

where the coefficients $\phi^{(n)}\left(x,u,u^{'},...,u^{(n)}\right)$ are given recursively by

$$\phi^{(n)} = \frac{d\phi^{(n-1)}}{dx} - u^{(n)}\frac{d\xi}{dx}, \quad n = 1,2,..., \tag{2.3.4}$$

and $\phi^{(0)} = \phi\left(x,u\right)$. Leach (2006) extended the generator to the nth derivative by deriving the following

$$\left.\begin{aligned}
\frac{d\hat{u}}{d\hat{x}} &= u' + \varepsilon\left(\phi' - \xi' u'\right)\\
\frac{d^2\hat{u}}{d\hat{x}^2} &= u'' + \varepsilon\left(\phi'' - 2\xi' u'' - \xi'' u'\right)\\
\frac{d^3\hat{u}}{d\hat{x}^3} &= u''' + \varepsilon\left(\phi'''' - 3\xi' u''' - 3\xi'' u'' - \xi''' u'\right)\\
\frac{d^4\hat{u}}{d\hat{x}^4} &= u^{iv} + \varepsilon\left(\phi^{iv} - 4\xi' u^{iv} - 6\xi'' u''' - 4\xi''' u'' - \xi^{iv} u'\right)
\end{aligned}\right\}. \tag{2.3.5}$$

He observed that the coefficients of the derivatives of ξ are just the binomial coefficients less the first one and then generalised (2.3.5) to

$$\frac{d^n\hat{u}}{d\hat{x}^n} = u^{(n)} + \varepsilon\left[\phi^{(n)} - \sum_{i=1}^{n}\left(\begin{array}{c} n \\ i \end{array}\right)\xi^{(i)}u^{(n-i+1)}\right]. \tag{2.3.6}$$

To enable Γ deal with derivatives Mahomed and Leach (1990) gave the nth prolongation of the generator of an nth-order ordinary differential equation compactly as

$$\Gamma^{[n]} = \Gamma + \sum_{i=1}^{n}\left\{\phi^{(i)} - \sum_{j=1}^{i}\left(\begin{array}{c} i \\ j \end{array}\right)u^{(i+1-j)}\xi^{(j)}\right\}\frac{\partial}{\partial u^{(i)}}. \tag{2.3.7}$$

2.3.2 Generalization to p dependent and k independent variables

If (2.3.2) is written in vector notation such as

$$\Gamma = \boldsymbol{\lambda}\left(\boldsymbol{\nu},\boldsymbol{u}\right)\frac{\partial}{\partial \boldsymbol{\nu}} + \boldsymbol{\phi}\left(\boldsymbol{\nu},\boldsymbol{u}\right)\frac{\partial}{\partial \boldsymbol{u}}, \tag{2.3.8}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_k) \cong (\xi, \theta, ..., \tau)$, $\boldsymbol{\nu} = (\nu_1, \nu_2, ..., \nu_k) \cong (x, r, ..., t)$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, ..., \phi_p)$, $\boldsymbol{u} = (u_1, u_2, ..., u_p)$. In this we require that the coefficient functions $\boldsymbol{\lambda}$ and $\boldsymbol{\phi}$ depend upon the independent and dependent variables, $\boldsymbol{\nu}$ and $\boldsymbol{u}$ respectively. For simplicity and without causing confusion we drop the "boldness" for the vectors without losing their original meanings. In line with Leach (2006)

$$\begin{aligned}
\frac{d\hat{u}}{d\hat{\nu}} &= \frac{d(u+\varepsilon\phi)}{d(\nu+\varepsilon\lambda)} = \frac{du+\varepsilon d\phi}{d\nu+\varepsilon d\lambda} \\
&= \frac{\dfrac{du}{d\nu}+\varepsilon\dfrac{d\phi}{d\nu}}{\dfrac{d\nu}{d\nu}+\varepsilon\dfrac{d\lambda}{d\nu}} \\
&= \frac{\dfrac{du}{dx}+\dfrac{du}{dr}+\cdots+\dfrac{du}{dt}+\varepsilon\left(\dfrac{d\phi}{dx}+\dfrac{d\phi}{dr}+\cdots+\dfrac{d\phi}{dt}\right)}{1+\varepsilon\left(\dfrac{d\xi}{dx}+\dfrac{d\theta}{dr}+\cdots+\dfrac{d\tau}{dt}\right)} \\
&= \frac{\sum_{i=1}^{k}\dfrac{du}{d\nu_i}+\varepsilon\sum_{i=1}^{k}\dfrac{d\phi}{d\nu_i}}{1+\varepsilon\sum_{i=1}^{k}\dfrac{d\lambda}{d\nu_i}} \\
&= \frac{u_x+u_r+\cdots+u_t+\varepsilon(\phi_x+\phi_r+\cdots+\phi_t)}{1+\varepsilon(\xi_x+\theta_r+\cdots+\tau_t)} \\
&= \frac{u'+\varepsilon\phi'}{(1+\varepsilon\lambda')} = (u'+\varepsilon\phi')(1+\varepsilon\lambda')^{-1} \\
&= (u'+\varepsilon\phi')\left(1-\varepsilon\lambda'+\varepsilon^2\lambda'^2-\cdots\right) \\
&= (u'+\varepsilon\phi')(1-\varepsilon\lambda') = u'+\varepsilon\phi'-\varepsilon\lambda'u'-\varepsilon^2\lambda'\phi' \\
&= u'+\varepsilon\phi'-\varepsilon\lambda'u'-o(\varepsilon) \\
&= u'+\varepsilon(\phi'-\lambda'u'). \qquad (2.3.9)
\end{aligned}$$

Notice that $u' = u_x + u_r + \cdots + u_t$, $\quad \lambda' = \xi_x + \theta_r + \cdots + \tau_t$ and $\phi' = \phi_x + \phi_r + \cdots + \phi_t$ and we continue with this notation throughout this derivation.

The second extension ($n = 2$) is derived as follows

$$\begin{aligned}
\frac{d^2\hat{u}}{d\hat{\nu}^2} &= \frac{d}{d\hat{\nu}}\left(\frac{d\hat{u}}{d\hat{\nu}}\right) \\
&= \frac{d}{d\hat{\nu}}\left[u' + \varepsilon\left(\phi' - \lambda' u'\right)\right] = \frac{du' + \varepsilon d\left(\phi' - \lambda' u'\right)}{d\hat{\nu} + \varepsilon d\lambda} \\
&= \frac{u'' + \varepsilon\left(\phi'' - \lambda'' u' - \lambda' u''\right)}{1 + \varepsilon\lambda'} \\
&= \left[u'' + \varepsilon\left(\phi'' - \lambda'' u' - \lambda' u''\right)\right]\left(1 - \varepsilon\lambda'\right) \\
&= u'' - \varepsilon\lambda' u'' + \varepsilon\left(\phi'' - \lambda'' u' - \lambda' u''\right) - \varepsilon^2\lambda'\left(\phi'' - \lambda'' u' - \lambda' u''\right) \\
&= u'' - \varepsilon\lambda' u'' + \varepsilon\left(\phi'' - \lambda'' u' - \lambda' u''\right) + o(\varepsilon) \\
&= u'' + \varepsilon\left(\phi'' - 2\lambda' u'' - \lambda'' u'\right). \qquad (2.3.10)
\end{aligned}$$

Here $u'' = u_{xx} + u_{rr} + \cdots + u_{tt}$, $\lambda'' = \xi_{xx} + \theta_{rr} + \cdots + \tau_{tt}$ and $\phi'' = \phi_{xx} + \phi_{rr} + \cdots + \phi_{tt}$.

For the third extension ($n = 3$), we proceed as follows:

$$\begin{aligned}
\frac{d^3\hat{u}}{d\hat{\nu}^3} &= \frac{d}{d\hat{\nu}}\left(\frac{d^2\hat{u}}{d\hat{\nu}^2}\right) \\
&= \frac{d}{d\hat{\nu}}\left[u'' + \varepsilon\left(\phi'' - 2\lambda' u'' - \lambda'' u'\right)\right] = \frac{d\left[u'' + \varepsilon\left(\phi'' - 2\lambda' u'' - \lambda'' u'\right)\right]}{d\left(\hat{\nu} + \varepsilon\lambda\right)} \\
&= \frac{u''' + \varepsilon\left(\phi''' - 2\lambda'' u'' - 2\lambda' u''' - \lambda''' u' - \lambda'' u''\right)}{1 + \varepsilon\lambda'} \\
&= \left[u''' + \varepsilon\left(\phi''' - 3\lambda'' u'' - 2\lambda' u''' - \lambda''' u'\right)\right]\left(1 - \varepsilon\lambda'\right) \\
&= u''' - \varepsilon\lambda' u''' + \varepsilon\left(\phi''' - 3\lambda'' u'' - 2\lambda' u''' - \lambda''' u'\right) - \varepsilon^2\lambda'\left(\phi''' - 3\lambda'' u'' - 2\lambda' u''' - \lambda''' u'\right) \\
&= u''' - \varepsilon\lambda' u''' + \varepsilon\left(\phi''' - 3\lambda'' u'' - 2\lambda' u''' - \lambda''' u'\right) + o(\varepsilon) \\
&= u''' + \varepsilon\left(\phi''' - 3\lambda' u''' - 3\lambda'' u'' - \lambda''' u'\right). \qquad (2.3.11)
\end{aligned}$$

Also $u''' = u_{xxx} + u_{rrr} + \cdots + u_{ttt}$, $\quad \lambda''' = \xi_{xxx} + \theta_{rrr} + \cdots + \tau_{ttt}$ and $\phi''' = \phi_{xxx} + \phi_{rrr} + \cdots + \phi_{ttt}$.

The fourth ($n = 4$) extension is similarly obtained as

$$\frac{d^4\hat{u}}{d\hat{\nu}^4} = u^{iv} + \varepsilon\left(\phi^{iv} - 4\lambda' u^{iv} - 6\lambda'' u''' - 4\lambda''' u'' - \lambda^{iv} u'\right). \qquad (2.3.12)$$

To deal with the infinitesimal transformations of equations and functions involving derivatives we need the extensions of the generator Γ. We therefore indicate that Γ has been extended by

writing

$$\begin{aligned}\Gamma^{[1]} &= \Gamma + (\phi' - \lambda' u') \frac{\partial}{\partial u'} \\ \Gamma^{[2]} &= \Gamma^{[1]} + (\phi'' - 2\lambda' u'' - \lambda'' u') \frac{\partial}{\partial u''}\end{aligned}$$

in the case of one independent variable which generalises to equation (2.3.7). For the general k independent variables, equation (2.3.9) through (2.3.12) yields

$$\Gamma^{[n]} = \Gamma + \sum_{\nu} \left[D_\nu^n \left(\phi - \sum_{i=1}^{k} \lambda_i u_i \right) + \sum_{i=1}^{k} \lambda_i u_{\nu, i^{(n)}} \right] \frac{\partial}{\partial u_\nu}, \tag{2.3.13}$$

where $\sum_\nu$ sums for each independent variable, D_ν^n denotes the nth total derivative (for the desired nth prolongation) with respect to ν. All four vectors are now understood as

$$\nu = \{\nu_i\}_{i=1,2,...,k} \Rightarrow \nu_1 = x, \nu_2 = r, ..., \nu_k = t.$$

Case A: Two independent variables, x and t, and one dependent variable, u

Recall (2.3.8)

$$\Gamma = \xi(x,t,u) \frac{\partial}{dx} + \tau(x,t,u) \frac{\partial}{dt} + \phi(x,t,u) \frac{\partial}{\partial u}. \tag{2.3.14}$$

The second extension using (2.3.13) is

$$\begin{aligned}\Gamma^{[2]} = \Gamma &+ D_x(\phi - \xi u_x - \tau u_t) + \xi u_{xx} + \tau u_{xt} \\ &+ D_t(\phi - \xi u_x - \tau u_t) + \xi u_{tx} + \tau u_{tt} \Big\} n = 1 \\ &+ D_x D_x(\phi - \xi u_x - \tau u_t) + \xi u_{xxx} + \tau u_{xxt} \\ &+ D_t D_x(\phi - \xi u_x - \tau u_t) + \xi u_{txx} + \tau u_{txt} \\ &+ D_t D_t(\phi - \xi u_x - \tau u_t) + \xi u_{ttx} + \tau u_{ttt} \Big\} n = 2.\end{aligned} \tag{2.3.15}$$

In (2.3.15) only the inner part (.) of the [..] in (2.3.13) is taken as the rest eventually vanishes on simplification. That aspect can be denoted by $\phi_\nu^{(n)}$ as explained hereunder. Equation (2.3.15) can then be written as

$$\Gamma^{[2]} = \Gamma + \underbrace{\phi_x \frac{\partial}{\partial u_x} + \phi_t \frac{\partial}{\partial u_t}}_{n=1} + \underbrace{\phi_{xx} \frac{\partial^2}{\partial u_x^2} + \phi_{xt} \frac{\partial^2}{\partial u_x u_t} + \phi_{tt} \frac{\partial^2}{\partial u_t^2}}_{n=2}$$

$$\phi_x^{(1)} = D_x(\phi - \xi u_x - \tau u_t),$$

$$\phi_t^{(1)} = D_t\left(\phi - \xi u_x - \tau u_t\right),$$

$$\phi_{xx}^{(2)} = D_x D_x\left(\phi - \xi u_x - \tau u_t\right) = D_x^2\left(\phi - \xi u_x - \tau u_t\right)$$

and similarly,

$$\phi_{xt}^{(2)} = D_x D_t\left(\phi - \xi u_x - \tau u_t\right),$$

$$\phi_{tt}^{(2)} = D_t D_t\left(\phi - \xi u_x - \tau u_t\right).$$

On expansion these total derivatives give

$$D_x\left(\phi - \xi u_x - \tau u_t\right) = \frac{\partial \phi}{\partial x} + \left(\frac{\partial \phi}{\partial u} - \frac{\partial \xi}{\partial x}\right) u_x - \frac{\partial \tau}{\partial x} u_t - \frac{\partial \xi}{\partial u} u_x^2 - \frac{\partial \tau}{\partial u} u_x u_t \tag{2.3.16}$$

$$D_t\left(\phi - \xi u_x - \tau u_t\right) = \frac{\partial \phi}{\partial t} + \left(\frac{\partial \phi}{\partial u} - \frac{\partial \tau}{\partial t}\right) u_t - \frac{\partial \xi}{\partial t} u_x - \frac{\partial \tau}{\partial u} u_t^2 - \frac{\partial \xi}{\partial u} \xi_u u_x u_t \tag{2.3.17}$$

$$\begin{aligned} D_{xx}\left(\phi - \xi u_x - \tau u_t\right) = \; & \frac{\partial^2 \phi}{\partial x^2} + \left(2\frac{\partial^2 \phi}{\partial x \partial u} - \frac{\partial^2 \xi}{\partial x^2}\right) u_x - \frac{\partial^2 \tau}{\partial x^2} u_t + \left(\frac{\partial^2 \phi}{\partial u^2} - 2\frac{\partial^2 \xi}{\partial x \partial u}\right) u_x^2 \\ & - 2\frac{\partial^2 \tau}{\partial x \partial u} u_x u_t - \frac{\partial^2 \xi}{\partial u^2} u_x^3 - \frac{\partial^2 \tau}{\partial u^2} u_x^2 u_t + \left(\frac{\partial \phi}{\partial u} - 2\frac{\partial \xi}{\partial x}\right) u_{xx} \\ & - 2\frac{\partial \tau}{\partial x} u_{xt} - 3\frac{\partial \xi}{\partial u} u_x u_{xx} - \frac{\partial \tau}{\partial u} u_{xx} u_t - 2\frac{\partial \tau}{\partial u} u_x u_{xt} \end{aligned} \tag{2.3.18}$$

$$\begin{aligned} D_{tt}\left(\phi - \xi u_x - \tau u_t\right) = \; & \frac{\partial^2 \phi}{\partial t^2} + \left(2\frac{\partial^2 \phi}{\partial t \partial u} - \frac{\partial^2 \tau}{\partial t^2}\right) u_t - \frac{\partial^2 \xi}{\partial t^2} u_x + \left(\frac{\partial^2 \phi}{\partial u^2} - 2\frac{\partial^2 \tau}{\partial t \partial u}\right) u_t^2 \\ & - 2\frac{\partial^2 \xi}{\partial x \partial u} u_x u_t - \frac{\partial^2 \tau}{\partial u^2} u_t^3 - \frac{\partial^2 \xi}{\partial u^2} u_x u_t^2 + \left(\frac{\partial \phi}{\partial u} - 2\frac{\partial \tau}{\partial t}\right) u_{tt} \\ & - 2\frac{\partial \xi}{\partial t} u_{xt} - 3\frac{\partial \tau}{\partial u} u_t u_{tt} - \frac{\partial \xi}{\partial u} u_x u_{tt} - 2\frac{\partial \xi}{\partial u} u_t u_{xt} \end{aligned} \tag{2.3.19}$$

$$\begin{aligned} D_{xt}\left(\phi - \xi u_x - \tau u_t\right) = \; & - \; \frac{\partial \xi}{\partial t} u_{xx} + \frac{\partial^2 \phi}{\partial x \partial t} + \left(\frac{\partial^2 \phi}{\partial t \partial u} - \frac{\partial^2 \xi}{\partial x \partial t}\right) u_x + \left(\frac{\partial^2 \phi}{\partial x \partial u} - \frac{\partial^2 \tau}{\partial x \partial t}\right) u_t \\ & + \; \left(\frac{\partial \phi}{\partial u} - \frac{\partial \xi}{\partial x} - \frac{\partial \tau}{\partial t}\right) u_{xt} - \frac{\partial^2 \xi}{\partial t \partial u} u_x^2 + \left(\frac{\partial^2 \phi}{\partial u^2} - \frac{\partial^2 \tau}{\partial t \partial u} - \frac{\partial^2 \xi}{\partial x \partial u}\right) u_x u_t \\ & - \; \frac{\partial \tau}{\partial x} u_{tt} - \frac{\partial^2 \tau}{\partial x \partial u} u_t^2 - \frac{\partial^2 \xi}{\partial u^2} u_x^2 u_t - \frac{\partial^2 \tau}{\partial u^2} u_x u_t^2 - 2\frac{\partial \xi}{\partial u} u_x u_{xt} - 2\frac{\partial \tau}{\partial u} u_t u_{xt} \\ & - \; \frac{\partial \tau}{\partial u} u_x u_{tt} - \frac{\partial \xi}{\partial u} u_t u_{xx}. \end{aligned} \tag{2.3.20}$$

Case B: Three independent variables, x, r and t

$$\Gamma = \xi(x,r,t,u)\frac{\partial}{dx} + \theta(x,r,t,u)\frac{\partial}{\partial r} + \tau(x,r,t,u)\frac{\partial}{\partial t} + \phi(x,r,t,u)\frac{\partial}{\partial u}. \tag{2.3.21}$$

The first and second extensions are

$$\begin{aligned}
\Gamma^{[2]} &= \Gamma + \sum_v \left[D_v(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_x + \theta u_r + \tau u_t\right]\frac{\partial}{\partial u_v} \\
&= \Gamma + \left.\begin{aligned} &D_x(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{xx} + \theta u_{xr} + \tau u_{xt} \\ &+D_r(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{rx} + \theta u_{rr} + \tau u_{rt} \\ &+D_t(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{tx} + \theta u_{tr} + \tau u_{tt} \end{aligned}\right\} n=1 \\
&\left.\begin{aligned} &+D_xD_x(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{xxx} + \theta u_{xxr} + \tau u_{xxt} \\ &+D_xD_r(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{xrx} + \theta u_{xrr} + \tau u_{xrt} \\ &+D_xD_t(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{xtx} + \theta u_{xtr} + \tau u_{xtt} \\ &+D_rD_r(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{rrx} + \theta u_{rrr} + \tau u_{rrt} \\ &+D_rD_t(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{rtx} + \theta u_{rtr} + \tau u_{rtt} \\ &+D_tD_t(\phi - \xi u_x - \theta u_r - \tau u_t) + \xi u_{ttx} + \theta u_{ttr} + \tau u_{ttt} \end{aligned}\right\} n=2 \\
&= \Gamma + \underbrace{\phi_x\frac{\partial}{\partial u_x} + \phi_r\frac{\partial}{\partial u_r} + \phi_t\frac{\partial}{\partial u_t}}_{n=1} \\
&+ \underbrace{\phi_{xx}\frac{\partial}{\partial u_{xx}} + \phi_{xr}\frac{\partial}{\partial u_{xr}} + \phi_{rr}\frac{\partial}{\partial u_{rr}} + \phi_{xt}\frac{\partial}{\partial u_{xt}} + \phi_{rt}\frac{\partial}{\partial u_{rt}} + \phi_{tt}\frac{\partial}{\partial u_{tt}}}_{n=2}.
\end{aligned} \tag{2.3.22}$$

Notice the action of $\sum_\nu [\ \ ]$ on the entire [.] and the vector ν. The behaviour of D_ν^n is also clear, where $n = 1$ in the first extension and $n = 2$, etc. The portion $D_\nu^n(\phi - \xi u_x - \theta u_r - \tau u_t)$ is the relevant portion since others eventually cancel after factorization. Influence of n and v are noticeable only on ϕ as we see

$$\begin{aligned}
\phi_x^{(1)} &= D_x(\phi) - u_x D_x(\xi) - u_r D_x(\theta) - u_t D_x(\tau) \\
&= \frac{\partial\phi}{\partial x} + \left(\frac{\partial\phi}{\partial u} - \frac{\partial\xi}{\partial x}\right)u_x - \frac{\partial\theta}{\partial x}u_r - \frac{\partial\tau}{\partial x}u_t
\end{aligned} \tag{2.3.23}$$

$$\begin{aligned}
\phi_r^{(1)} &= D_r(\phi) - u_x D_r(\xi) - u_r D_r(\theta) - u_t D_r(\tau) \\
&= \frac{\partial\phi}{\partial r} + \left(\frac{\partial\phi}{\partial u} - \frac{\partial\theta}{\partial r}\right)u_r - \frac{\partial\xi}{\partial r}u_x - \frac{\partial\tau}{\partial r}u_t
\end{aligned} \tag{2.3.24}$$

$$\begin{aligned}\phi_t^{(1)} &= D_t(\phi) - u_x D_t(\xi) - u_r D_t(\theta) - u_t D_t(\tau) \\ &= \frac{\partial \phi}{\partial t} + \left(\frac{\partial \phi}{\partial u} - \frac{\partial \tau}{\partial t}\right) u_t - \frac{\partial \xi}{\partial t} u_x - \frac{\partial \theta}{\partial t} u_r\end{aligned} \tag{2.3.25}$$

$$\begin{aligned}&\phi_{xx}^{(2)} = D_x(\phi_x) - u_{xx} D_x(\xi) - u_{xr} D_x(\theta) - u_{xt} D_x(\tau) \\ &= \frac{\partial^2 \phi}{\partial x^2} + \left(2\frac{\partial^2 \phi}{\partial x \partial u} - \frac{\partial^2 \xi}{\partial x^2}\right) u_x - \frac{\partial^2 \theta}{\partial x^2} u_r - 2\frac{\partial \tau}{\partial x} u_{xt} + \left(\frac{\partial \phi}{\partial u} - 2\frac{\partial \xi}{\partial x}\right) u_{xx} \\ &- \frac{\partial^2 \tau}{\partial x^2} u_t - 2\frac{\partial \theta}{\partial x} u_{xr}\end{aligned} \tag{2.3.26}$$

$$\begin{aligned}&\phi_{xr}^{(2)} = \phi_{rx}^{(2)} = D_r(\phi_x) - u_{rx} D_r(\xi) - u_{rr} D_r(\theta) - u_{rt} D_r(\tau) \\ &= \frac{\partial^2 \phi}{\partial x \partial r} + \left(\frac{\partial^2 \phi}{\partial r \partial u} - \frac{\partial^2 \theta}{\partial x \partial r}\right) u_x + \left(\frac{\partial^2 \phi}{\partial x \partial u} - \frac{\partial^2 \xi}{\partial x \partial r}\right) u_r + \left(\frac{\partial \phi}{\partial u} - \frac{\partial \xi}{\partial x} - \frac{\partial \theta}{\partial r}\right) u_{xr} \\ &- \frac{\partial \theta}{\partial x} u_{rr} - \frac{\partial \xi}{\partial r} u_{xx} - \frac{\partial \tau}{\partial r} u_{xt} - \frac{\partial \tau}{\partial x} u_{rt} - \frac{\partial^2 \tau}{\partial x \partial r} u_t\end{aligned} \tag{2.3.27}$$

$$\begin{aligned}&\phi_{xt}^{(2)} = \phi_{tx}^{(2)} = D_x(\phi_t) - u_{xx} D_x(\xi) - u_{xr} D_x(\theta) - u_{xt} D_x(\tau) \\ &= \frac{\partial^2 \phi}{\partial x \partial t} + \left(\frac{\partial^2 \phi}{\partial t \partial u} - \frac{\partial^2 \theta}{\partial x \partial t}\right) u_x + \left(\frac{\partial^2 \phi}{\partial x \partial u} - \frac{\partial^2 \xi}{\partial x \partial t}\right) u_t + \left(\frac{\partial \phi}{\partial u} - \frac{\partial \xi}{\partial x} - \frac{\partial \tau}{\partial t}\right) u_{xt} \\ &- \frac{\partial \tau}{\partial x} u_{tt} - \frac{\partial \xi}{\partial t} u_{xx} - \frac{\partial \theta}{\partial t} u_{xr} - \frac{\partial \theta}{\partial x} u_{rt} - \frac{\partial^2 \theta}{\partial x \partial t} u_r\end{aligned} \tag{2.3.28}$$

$$\begin{aligned}&\phi_{rr}^{(2)} = D_r(\phi_r) - u_{rx} D_r(\xi) - u_{rr} D_r(\theta) - u_{rt} D_r(\tau) \\ &= \frac{\partial^2 \phi}{\partial r^2} + \left(2\frac{\partial^2 \phi}{\partial r \partial u} - \frac{\partial^2 \theta}{\partial r^2}\right) u_r - \frac{\partial^2 \xi}{\partial r^2} u_x - 2\frac{\partial \tau}{\partial r} u_{rt} + \left(\frac{\partial \phi}{\partial u} - 2\frac{\partial \theta}{\partial r}\right) u_{rr} \\ &- \frac{\partial^2 \tau}{\partial r^2} u_t - 2\frac{\partial \xi}{\partial r} u_{xr}\end{aligned} \tag{2.3.29}$$

$$\begin{aligned}&\phi_{rt}^{(2)} = \phi_{tr}^{(2)} = D_r(\phi_t) - u_{rx} D_r(\xi) - u_{rr} D_r(\theta) - u_{rt} D_r(\tau) \\ &= \frac{\partial^2 \phi}{\partial r \partial t} + \left(\frac{\partial^2 \phi}{\partial t \partial u} - \frac{\partial^2 \phi}{\partial r \partial t}\right) u_r + \left(\frac{\partial^2 \phi}{\partial r \partial u} - \frac{\partial^2 \phi}{\partial r \partial t}\right) u_t + \left(\frac{\partial \phi}{\partial u} - \frac{\partial \theta}{\partial r} - \frac{\partial \tau}{\partial t}\right) u_{rt} \\ &- \frac{\partial \theta}{\partial t} u_{rr} - \frac{\partial \tau}{\partial r} u_{tt} - \frac{\partial \xi}{\partial t} u_{xt} - \frac{\partial \xi}{\partial r} u_{xt} - \frac{\partial^2 \xi}{\partial r \partial t} u_x\end{aligned} \tag{2.3.30}$$

$$\begin{aligned}&\phi_{tt}^{(2)} = D_t(\phi_t) - u_{tx} D_t(\xi) - u_{tr} D_t(\theta) - u_{tt} D_t(\tau) \\ &= \frac{\partial^2 \phi}{\partial t^2} + \left(2\frac{\partial^2 \phi}{\partial t \partial u} - \frac{\partial^2 \tau}{\partial t^2}\right) u_t - \frac{\partial^2 \xi}{\partial t^2} u_x - 2\frac{\partial \theta}{\partial t} u_{rt} + \left(\frac{\partial \phi}{\partial u} - 2\frac{\partial \tau}{\partial t}\right) u_{tt} \\ &- \frac{\partial^2 \theta}{\partial t^2} u_r - 2\frac{\partial \xi}{\partial t} u_{xr}.\end{aligned} \tag{2.3.31}$$

Proposition 2.3.1. *Let $f : V \longrightarrow U$ be a smooth function from $V \simeq \mathbb{R}^k$ to $U \simeq \mathbb{R}^p$ so that the general vector field on $V \times U \simeq \mathbb{R}^k \times \mathbb{R}^p$. Then the nth extension of the generator Γ is given by*

$$\Gamma^{[n]} = \Gamma + \sum_{\nu} \left[D_\nu^n \left(\phi - \sum_{i=1}^{k} \lambda_i u_i \right) + \sum_{i=1}^{k} \lambda_i u_{\nu, i^{(n)}} \right] \frac{\partial}{\partial u_\nu}$$

where $\Gamma, \nu, \lambda, \phi, u, \sum_\nu$ and D_ν^n are as defined under equations (2.3.8) and (2.3.13). With this proposition we have succeeded in establishing a formula for the nth prolongation of a generator Γ with k independent variables and p dependent variables.

2.4 Group-invariance of Differential Equations

Symmetry groups of partial differential equations can be used to reduce the total number of variables (dependent and independent) in an equation (Clarkson, 1995) during an attempt to obtain the solution to the original equation. Each attempt of a group reduction results in an introduction of new variables through which the number of independent variables decreases by one. Thus for partial differential equations, like the heat equation, admitting only two independent variables, a single reduction transforms it into an ordinary differential equation which is generally simpler to solve than the original partial differential equation and its solution still maintains the characteristics of the particular partial differential equation.

The Lie method has the advantage that the new variables can be determined through the symmetries of the partial differential equation. The solutions obtained by this method were referred to by Olver and Rosenau (1987) as *group-invariant solutions* (also called *similarity solutions* especially when scale-invariance is involved) of the partial differential equation. The likes of Bluman and Kumei (1989), Dresner (1999), Ibragimov (1995) and Stephani (1989) are all in unison that this method is one of the best known systematic methods for the simplification and solution of partial differential equations. The beauty of the Lie method is that it enables a systematic approach to the determination of particular solutions. It is the combination of the Lie algebra of the differential equation and compatibility with boundary/initial conditions which provides the possibility of obtaining the solution. We may also seek various forms of solution such as similarity solutions, *traveling-wave solutions*, *separable solutions*, *et cetera*.

Consider the nth-order partial differential equation (2.1.1) for $n \geq 2$ that admits a one-parameter Lie group of point transformations with the infinitesimal generator (2.3.8), where $\lambda(\nu, u)$ is not identically zero.

Definition 2.4.1 (Bluman and Anco)**.** $u = \Theta(\lambda)$ is an invariant solution of (2.1.1) resulting from its admitted point symmetry with the infinitesimal generator (2.3.8) if and only if

(i) $u = \Theta(\lambda)$ is an invariant surface of (2.3.8) and

(ii) $u = \Theta(\lambda)$ solves (2.1.1).

It follows that we obtain the invariant surface conditions for reduction by solving the corresponding characteristic equation $\Gamma(u - \Theta(\lambda)) = 0$ when $u = \Theta(\lambda)$, i.e. from

$$\Gamma = \xi(x, \cdots, t, u)\frac{\partial}{dx} + \cdots + \tau(x, \cdots, t, u)\frac{\partial}{\partial t} + \eta(x, \cdots, t, u)\frac{\partial}{\partial u} \tag{2.4.1}$$

$$\frac{dx}{\xi(x, \cdots, t, u)} = \cdots = \frac{dt}{\tau(x, \cdots, t, u)} = \frac{du}{\eta(x, \cdots, t, u)}$$

The invariance requirement is determined by

$$\Gamma^{[2]}(u - \Theta(\lambda))|_{u=\Theta(\lambda)} = 0 \tag{2.4.2}$$

extended to the second jet space, parameterized by $(x, t, u, u_x, u_t, u_{xx}, u_{xt}, u_{tt})$. Equation (2.4.2) is a polynomial equation in a set of independent functions of the derivatives of u. As the equation must be true for arbitrary values of these independent functions, their coefficients must vanish and this leads to an overdetermined linear system of equations known as the *determining equations* for the coefficients $\xi(x, t, u)$, $\tau(x, t, u)$ and $\eta(x, t, u)$. For known functions, ξ, τ and η, invariant solutions u corresponding to (2.2.4) satisfy the invariant surface condition

$$\Delta = \xi(x, t, u)\frac{\partial}{dx} + \tau(x, t, u)\frac{\partial}{\partial t} + \eta(x, t, u)\frac{\partial}{\partial u} = 0$$

which, when solved as a first-order partial differential equation by the method of characteristics, yields the functional form of the similarity solution in terms of an arbitrary function, i.e.,

$$u = \phi(x, t, \phi(z)), \quad z = z(x, t),$$

where ϕ is an arbitrary function of an invariant z for the symmetry. The substitution of this functional form into $\Delta(x, t, u, u_x, u_t, u_{xx}, u_{xt}, u_{tt}) = 0$ produces a quotient ordinary differential equation which can be solved for the function $\phi(z)$.

2.5 Lie's Algorithm

The computation of Lie point symmetries follow a particular sequence usually called the *Lie algorithm*. The determination of symmetries of a given differential equation involves setting up and solving an associated system of linear homogeneous partial differential equations called *determining equations*. We discuss how determining equations arise from symmetry problems and illustrate this by outlining the derivation of such equations using the heat equation as an example. Lie (1881) gave the group classification of linear second-order partial differential equations with two independent variables and developed methods of their integration. In Lie's classification (Gazizov and Ibragimov (1998)) all parabolic equations admitting the symmetry group of the highest order reduce to the heat equation. By this statement the heat equation clearly becomes the benchmark for accessing other parabolic partial differential equations. This property of the heat equation is revisited in Chapter Three of this Thesis when we map our working partial differential equation to the linear heat equation.

2.5.1 New Solutions from old ones

Definition 2.5.1. A symmetry of a given partial differential equation is a transformation which maps every solution of the system to another solution of the same equation (i.e., it maps the solution set of the equation into itself).

Symmetries of certain differential equations are obvious, such as that of the Laplace equation which is not discussed here. The next example illustrates the fact that symmetries of differential equations such as that of the linear heat equation are not always obvious.

Example 2.5.2 (The heat Equation)**.** Consider the linear heat equation

$$\frac{\partial u}{\partial t} - h\frac{\partial^2 u}{\partial x^2} = 0, \tag{2.5.1}$$

where u represents the temperature of the medium at time, t, and x the only spatial variable represents distance while h is a constant representing the diffusivity of the medium. Equation (2.5.1) has been studied for nearly two centuries as a model of the flow (or diffusion) of heat

in a continuous medium. It is one of the most successful and widely used models in Applied Mathematics and a considerable body of theory on its properties and solution abound. Without loss of generality (2.5.1) can be suitably rescaled to

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \tag{2.5.2}$$

which has an unobvious symmetry $(x, t, u) \longmapsto \left(\hat{x}, \hat{t}, \hat{u}\right)$ given by[1]

$$\hat{x} = \frac{x}{1-\varepsilon t}, \quad \hat{t} = \frac{t}{1-\varepsilon t}, \quad \hat{u} = u\sqrt{1-\varepsilon t}\exp\left(\frac{-\varepsilon x^2}{4\left(1-\varepsilon t\right)}\right) \tag{2.5.3}$$

for $\varepsilon \in \mathbb{R}$. By calculating how u_t and u_{xx} transform under (2.5.2) and demonstrating that $\frac{\hat{\partial}}{\partial \hat{t}} = \frac{\hat{\partial}^2}{\partial \hat{x}^2}, (i.e \quad \hat{u}_t = \hat{u}_{\hat{x}\hat{x}})$ when $\frac{\partial}{\partial t} = \frac{\partial^2}{\partial x^2}$, it is straightforward but algebraically tedious to verify that (2.5.2) is a symmetry of (2.5.1). Symmetries can be used to generate nonobvious solutions from obvious ones, a property which can be exploited in application (Olver, 1993). Take the trivial solution $u = 1$ (of (2.5.1)); it represents a plane in $(x, t, u) -$ space. The symmetry (2.5.2) maps $u = 1$ into a family of surfaces in (x, t, u) −space, representing solutions $\hat{u} = \hat{\psi}\left(\hat{x}, \hat{t}\right)$ of $\hat{u}_{\hat{t}} = \hat{u}_{\hat{x}\hat{x}}$, see Figure 2.5.1.

Inverting (2.3.2) we obtain

$$x = \frac{\hat{x}}{1-\varepsilon\hat{t}}, \quad t = \frac{\hat{t}}{1-\varepsilon\hat{t}} \quad and \quad u = \hat{u}\sqrt{1-\varepsilon\hat{t}}\exp\left(\frac{-\varepsilon\hat{x}^2}{4\left(1-\varepsilon\hat{t}\right)}\right). \tag{2.5.4}$$

So under the action of the above symmetry the solution $u = 1$ maps to

$$\hat{u}\sqrt{1-\varepsilon\hat{t}}\exp\left(\frac{\varepsilon\hat{x}^2}{4\left(1-\varepsilon\hat{t}\right)}\right) = 1 \tag{2.5.5}$$

or, equivalently,

$$u = \frac{1}{\sqrt{1+\varepsilon t}}\exp\left(\frac{-\varepsilon x^2}{4\left(1+\varepsilon t\right)}\right) \tag{2.5.6}$$

after dropping the caret and by letting $\varepsilon = -\varepsilon$.

[1] Actually the finite transformation corresponds to some symmetry.

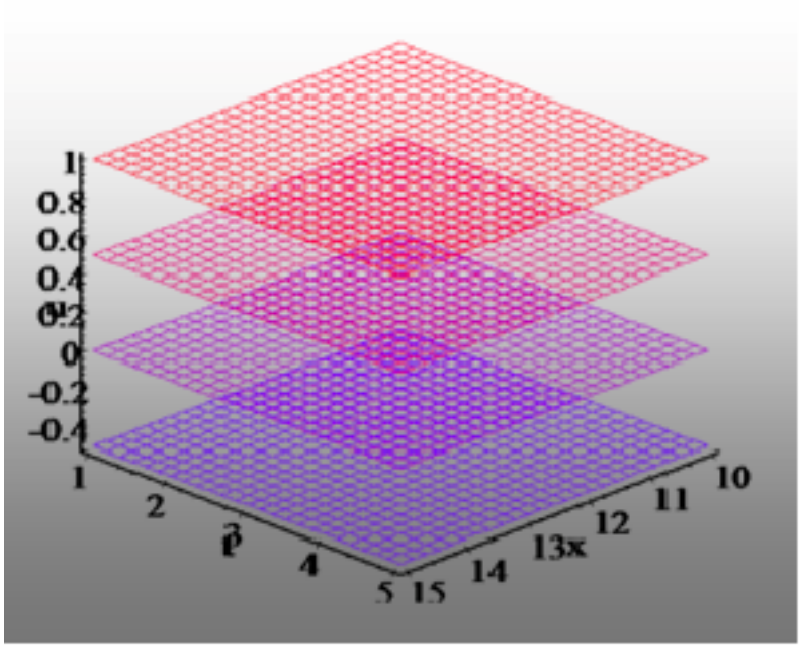


Figure 2.5.1: The solution surface of the trivial solution of the heat equation for $u = 1$

2.5.2 Computation of Lie Point Symmetries

We demonstrate in this subsection a systematic procedure due to Lie for obtaining related partial differential equations which, if solved fully, yield all infinitesimal symmetries[2]

Example 2.5.3. Consider the heat equation of Example 2.5.1

$$\frac{\partial u}{\partial t} - h\frac{\partial^2 u}{\partial x^2} = 0.$$

Recall the differential operator of (2.3.14)

$$\Gamma = \xi\left(x,t,u\right)\frac{\partial}{\partial x} + \tau\left(x,t,u\right)\frac{\partial}{\partial t} + \phi\left(x,t,u\right)\frac{\partial}{\partial u}.$$

Then after the application of its second extension (2.3.15) and the observation of the infinitesimal criterion for invariance, i.e.,

$$\Gamma^{[2]}\left(\frac{\partial u}{\partial t} - h\frac{\partial^2 u}{\partial x^2}\right)\Big|_{\frac{\partial u}{\partial t} - h\frac{\partial^2 u}{\partial x^2} = 0} = 0, \tag{2.5.7}$$

[2]It is important to note however that certain difficulties preclude the possibilities of solving the determining equations for all infinitesimal symmetries.

we obtain

$$\phi_t = h\phi_{xx} \tag{2.5.8}$$

whenever $u_t = hu_{xx}$ is satisfied. Substitution of hu_{xx} for u_t wherever u_t occurs in (2.5.8) (recalling (2.3.17) and (2.3.18)) yields

$$\begin{aligned}
&\eta_t + u_{xx}\left(\eta_u - \tau_t\right) - \xi_t u_x - \tau_u u_{xx}^2 - \xi_u u_x u_{xx} - h\eta_{xx} - hu_x\left(2\eta_{xu} - \xi_{xx}\right) \\
&+h\tau_{xx}u_{xx} - hu_x^2\left(\eta_{uu} - 2\xi_{xu}\right) + 2hu_x u_{xx}\tau_{xu} + h\xi_{uu}u_x^3 + h\tau_{uu}u_x^2 u_{xx} - hu_{xx}\left(\eta_u - 2\xi_x\right) \\
&+2h\tau_x u_{xt} + 3h\xi_u u_x u_{xx} + h\tau_u u_{xx}^2 + 2h\tau_u u_x u_{xt} = 0.
\end{aligned} \tag{2.5.9}$$

Equation (2.5.9) is satisfied if and only if all the coefficients of the powers of the derivatives of u are identically zero since ξ, τ, and η are functions of x, t and u only. This leads to the following set of coupled linear partial differential equations

$$\left.\begin{aligned}
&1 : \eta_t - h\eta_{xx} = 0 \\
&u_x : 2h\eta_{xu} - h\xi_{xx} + \xi_t = 0 \\
&u_x^2 : \eta_{uu} - 2\xi_{xu} = 0 \\
&u_x^3 : \xi_{uu} = 0 \\
&u_{xx} : 2\xi_x + h\tau_{xx} - \tau_t = 0 \\
&u_{xx}^2 : \tau_u = 0 \\
&u_x u_{xx} : \xi_u + h\tau_{xu} = 0 \\
&u_x^2 u_{xx} : \tau_{uu} = 0 \\
&u_{xt} : \tau_x = 0 \\
&u_x u_{xt} : \tau_u = 0
\end{aligned}\right\}. \tag{2.5.10}$$

Solution of these determining equations leads to

$$\left.\begin{aligned}
\xi\left(x,t,u\right) &= \tfrac{1}{2}c_2 x + c_3 xt + c_4 + c_5 t \\
\tau\left(x,t,u\right) &= c_1 + c_2 t + c_3 t^2 \\
\eta\left(x,t,u\right) &= -\left[c_3\left(\tfrac{1}{2}t + \tfrac{1}{4h}x^2\right) + \tfrac{1}{2h}c_5 x - c_6\right]u + f\left(x,t\right)
\end{aligned}\right\}. \tag{2.5.11}$$

The arbitrary constants, $c_1, ..., c_6$, in (2.5.11) indicate that the system has a six-parameter Lie group of infinitesimal operators which we list hereunder

$$\Gamma_1 = \frac{\partial}{\partial x} \tag{2.5.12}$$

$$\Gamma_2 = \frac{\partial}{\partial t} \tag{2.5.13}$$

$$\Gamma_3 = u\frac{\partial}{\partial u} \tag{2.5.14}$$

$$\Gamma_4 = \frac{1}{2}x\frac{\partial}{\partial x} + t\frac{\partial}{\partial t} \tag{2.5.15}$$

$$\Gamma_5 = t\frac{\partial}{\partial x} - \frac{1}{2h}xu\frac{\partial}{\partial u} \tag{2.5.16}$$

$$\Gamma_6 = xt\frac{\partial}{\partial x} + t^2\frac{\partial}{\partial t} - \left(\frac{1}{4h}x^2 + \frac{1}{2}t\right)u\frac{\partial}{\partial u} \tag{2.5.17}$$

$$\Gamma_\infty = f(x,t)\frac{\partial}{\partial u}, \tag{2.5.18}$$

where $f(x,t)$ is any solution of the original heat equation reflecting its linearity. Thus this equation admits an infinite-dimensional Lie symmetry algebra. The associated Lie algebra of (2.5.12)–(2.5.18) is given by Dimas *et al* (2009) as $\{sl(2,R) \oplus_s W_3\} \oplus_s \infty A_1$, where W_3 is the three-dimensional Heisenberg-Weyl algebra.

By the definition of the Lie Brackets of (1.4.8) in Section (1.4) we construct Table 2.5.1, where its elements are the structure constants arising from the commutation relations among pairs of operators in (2.5.12)–(2.5.17). Possession of a sufficient number of point symmetries is a rare phenomenon among differential equations and those partial differential equations modeled from natural applications such as those arising from financial mathematics often belong to this category. However, most of them are linked through a coordinate transformation to the heat equation (Bluman and Kumei (1989)) in $1+1$ dimensions. In fact it has been shown in Bluman and Cole (1974) that the heat equation is the only polynomial partial differential equation of the second order with two independent parameters invariant under the finite group of the heat equation itself. These are some of the reasons why the heat equation is relevant to this study and therefore used to demonstrate the algorithm.

Although the method of determining point symmetries of differential equations is entirely algorithmic, manual calculations always involve tedious computations and any slip in handling of the algebra results in inaccurate results and sometimes frustration. Fortunately the advent of symbolic manipulation of packages has virtually eliminated the strenuous algebra involved in the application of Lie algorithm. There are several programs available for the prosecution of

Table 2.5.1: Table of Lie Brackets for (2.5.12)−(2.5.17)

$[\Gamma_i, \Gamma_j]$	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6
Γ_1	0	0	0	$\frac{1}{2}\Gamma_1$	$-\frac{1}{2h}\Gamma_3$	Γ_5
Γ_2	0	0	0	Γ_2	Γ_1	$-\frac{1}{2}\Gamma_3 + \Gamma_4$
Γ_3	0	0	0	0	0	0
Γ_4	$-\frac{1}{2}\Gamma_1$	$-\Gamma_2$	0	0	$\frac{1}{2}\Gamma_5$	Γ_6
Γ_5	$\frac{1}{2h}\Gamma_3$	$-\Gamma_1$	0	$-\frac{1}{2}\Gamma_5$	0	0
Γ_6	$-\Gamma_5$	$\frac{1}{2}\Gamma_3 - \Gamma_4$	0	$-\Gamma_6$	0	0

the task of finding symmetries of differential equations. Some of them that are implemented in the MATHEMATICA environment are the Program LIE[3] (Head 1993, 1996) and Sherring *et al* (1997). Some of the stand-alone programs are REDUCE (Schwarz 1982) and Nucci (1990, 1996) while other computer implementation packages that are problem specifics (specific to some differential equations) are those of Hereman and Nuseir (1997), Fushchych and Kornyak (2001) and Champagne *et al* (1991). Hereman (1994, 1996) reviews symbolic manipulation programs that have been in use and especially, those adaptable for Lie group analysis of differential equations. One of the most recent packages used interactively with MATHEMATICA is the package SYM by Dimas and Tsoubelis (2005, 2006) and this was used in Chapter Three of this Thesis.

2.6 Chapter Summary

In 1990 Mahomed and Leach published their derivation (in correction of the error in Krause and Michel (1990)) of the nth prolongation of an ordinary differential equation of order n, see for

[3]The program LIE was introduced by Alan Head as a stand-alone PC program for the analysis of differential equations and written in MUMATH (Woof and Hodgkinson, 1987), a symbolic mathematics language for IBM-type PCs. Sinkala (2006) observed that sometimes LIE fails to find the symmetries of certain differential equations automatically and so he applied LIE interactively with MATHEMATICA in what he referred to as MATHEMATICA-assisted computation of Lie point symmetries.

instance, equation (2.3.7) of section 2.3. We have extended this result to the nth prolongation of the general k independent variables and p dependent variables of nth-order partial differential equation. This result was formally stated in Proposition 2.3.1. This proposed formula is expected to take care of prolongations in ordinary differential equations. This formula was applied in Section 2.5 of this Chapter for the heat equation and in Chapter Three in which we analysed the partial differential equation for pricing of electricity future contracts. We reviewed the Lie algorithm and applied it in the determination of Lie point symmetries of the heat equation.

Chapter 3

Symmetry Analysis of a Commodity-Pricing Model

3.1 Introduction

We perform a complete symmetry analysis of the electricity pricing model and its variants. Over the last few decades there has been a great interest in the modeling and analysis of problems arising in commodity markets. Some of these problems are modeled in terms of evolution partial differential equations. A number of studies have been devoted to the use of symmetry techniques for partial differential equations arising in the field of Financial Mathematics, see, for example, Ibragimov and Gazizov (1998), Goard (2000), Chou and Li (2001) and Sinkala *et al* (2008a,b). The general form of our model (the partial differential equation under the equivalent martingale measure $\mathbb{Q}$) is

$$\tfrac{1}{2}\sigma^2\nu_t\frac{\partial^2 y(t,\nu_t)}{\partial\nu_t^2}+\left(\kappa\theta-\left(\kappa+\lambda_\nu\right)\nu_t\right)\frac{\partial y(t,\nu_t)}{\partial\nu_t}+\frac{\partial y(t,\nu_t)}{\partial t}=k_1\nu_t y(t,\nu_t) \tag{3.1.1}$$

with boundary condition for the value at maturity (Kellerhals 2004, p.192)

$$y\left(T,\nu_t\right)=exp\left(k_1\nu_T\right) \tag{3.1.2}$$

For simplicity and without loss of generality we let $k_1 = k, \nu_t = x, \lambda_\nu = \lambda$ so that $y(\nu_t, t) = u(x,t)$. Hence (3.1.1) becomes

$$\frac{\partial u}{\partial t} + \tfrac{1}{2}\sigma^2 x \frac{\partial^2 u}{\partial x^2} + (\kappa\theta - (\kappa + \lambda)\, x) \frac{\partial u}{\partial x} - kxu = 0 \tag{3.1.3}$$

The package SYM (Dimas and Tsoubelis 2005, 2006) was used with MATHEMATICA 6.0 for the results which we present below.

3.2 The Lie Point Symmetries of (3.1.3)

As we saw in Chapter Two, the determination of the Lie point symmetries means finding the functions $\xi(x,t,u), \tau(x,t,u)$ and $\eta(x,t,u)$ such that the symmetry conditions are met. The symmetry conditions lead to a system of linear partial differential equations for the independent and dependent variables which eventually split into many more equations since they (independent and dependent variables) are independent of the derivatives of the dependent variable. However, the coefficients of these variables do depend upon these derivatives.

The basis operators of the Lie algebra under the subcase for which $\alpha = \kappa + \lambda$ and $\sigma^2 = 4\theta\kappa/3$ are given as

$$\Gamma_1 = \frac{\partial}{\partial t} \tag{3.2.1}$$

$$\Gamma_2 = u\frac{\partial}{\partial u} \tag{3.2.2}$$

$$\Gamma_{3,4} = e^{\pm\frac{1}{2}\phi t}\left[\sqrt{x}\frac{\partial}{\partial x} - \frac{1}{4\theta\kappa\sqrt{x}}\left(2\theta\kappa \pm 3x\left(\alpha \pm \phi\right)\right) u\frac{\partial}{\partial u}\right] \tag{3.2.3}$$

$$\Gamma_{5,6} = e^{\pm\phi t}\left[x\frac{\partial}{\partial x} \pm \frac{1}{\phi}\frac{\partial}{\partial t} + \frac{3}{4\theta\kappa\phi}\left(\phi x - \theta\kappa\right)\left(\alpha \pm \phi\right) u\frac{\partial}{\partial u}\right] \tag{3.2.4}$$

and

$$\Gamma_\infty = f(x,t)\,\frac{\partial}{\partial u}, \tag{3.2.5}$$

where $\phi = \sqrt{\kappa^2 + \lambda^2 + 2\kappa\lambda + 2k\sigma^2}$ and $f(x,t)$ is any solution of (3.1.3). The basis symmetries in (3.2.1)$-$(3.2.5) may be represented for convenience as

$$\Gamma_1 = \frac{\partial}{\partial t} \tag{3.2.6}$$

$$\Gamma_2 = u\frac{\partial}{\partial u} \tag{3.2.7}$$

$$\Gamma_3 = e^{\frac{1}{2}\phi t}\left[\sqrt{x}\frac{\partial}{\partial x} - \left(\frac{1}{2\sqrt{x}} + \mathcal{A}\sqrt{x}\right)u\frac{\partial}{\partial u}\right] \tag{3.2.8}$$

$$\Gamma_4 = e^{-\frac{1}{2}\phi t}\left[\sqrt{x}\frac{\partial}{\partial x} - \left(\frac{1}{2\sqrt{x}} - \mathcal{B}\sqrt{x}\right)u\frac{\partial}{\partial u}\right] \tag{3.2.9}$$

$$\Gamma_5 = e^{\phi t}\left[x\frac{\partial}{\partial x} + \frac{1}{\phi}\frac{\partial}{\partial t} + (\mathcal{A}x - \mathcal{C})\,u\frac{\partial}{\partial u}\right] \tag{3.2.10}$$

$$\Gamma_6 = e^{-\phi t}\left[x\frac{\partial}{\partial x} - \frac{1}{\phi}\frac{\partial}{\partial t} + (\mathcal{B}x - \mathcal{D})\,u\frac{\partial}{\partial u}\right] \tag{3.2.11}$$

and

$$\Gamma_\infty = f(x,t)\frac{\partial}{\partial u}, \tag{3.2.12}$$

where $\mathcal{A} = \frac{3}{4\theta\kappa}(\alpha+\phi), \mathcal{B} = \frac{3}{4\theta\kappa}(\alpha-\phi), \mathcal{C} = \frac{3}{4\phi}(\alpha+\phi)$ and $\mathcal{D} = \frac{3}{4\phi}(\alpha-\phi)$. The nongeneric Lie point symmetries $\Gamma_1, \Gamma_3 - \Gamma_6$, comprise two groups: the first group of symmetries Γ_1, Γ_5 and Γ_6 constitutes the Lie algebra $sl(2,R)$ and the second group Γ_3 and Γ_4 correspond to the solution symmetries of the one-dimensional free particle. The former is characteristic of an equation arising from finance. It is important to note that the Lie algebra of point symmetries for (3.2.6)-(3.2.12) spanned by the vectors Γ_1 (translation in t), Γ_2 (dilatation in u), Γ_3 and Γ_4(Galilean boost), Γ_5 and Γ_6 (local symmetries) and Γ_∞ is an additional infinite-dimensional subalgebra in which $f(x,t)$ is the solution of (3.1.3) and reflects its linearity. The associated Lie algebra of the above six-parameter Lie group of infinitesimal operators is $\{sl(2,R) \oplus W_3\} \oplus_s \infty A_1$, where W_3 is the three-dimensional Heisenberg-Weyl algebra implied by the commutation relations given in Table 3.2.1, when the solution symmetry is omitted. Note that the symbol $\oplus$ is used when all elements of the first subalgebra have zero Lie Brackets with all elements of the second. According to Andriopoulos (2008) it is important to note that the knowledge of the algebra of a given differential equation is vital. Not only does one perceive better the internal structure of the symmetries the differential equation possesses but also, provided the Lie algebra is either a well-studied one or consists of well-known subalgebras, one can disern those which are, for example, solvable and proceed to an order of reduction using one symmetry at a time.

Table 3.2.1: Table of Lie Brackets for (3.2.6)−(3.2.11)

$[\Gamma_i, \Gamma_j]$	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6
Γ_1	0	0	$\frac{\phi}{2}\Gamma_3$	$-\frac{\phi}{2}\Gamma_4$	$\phi\Gamma_5$	$-\phi\Gamma_6$
Γ_2	0	0	0	0	0	0
Γ_3	$-\frac{\phi}{2}\Gamma_3$	0	0	$-\frac{3\phi}{\theta\kappa}\Gamma_2$	0	$-\Gamma_4$
Γ_4	$\frac{\phi}{2}\Gamma_4$	0	$\frac{3\phi}{\theta\kappa}\Gamma_2$	0	$-\Gamma_3$	0
Γ_5	$-\phi\Gamma_5$	0	0	Γ_3	0	$\frac{32}{\phi}\Gamma_1 - \frac{24\alpha}{\phi}\Gamma_2$
Γ_6	$\phi\Gamma_6$	0	Γ_4	0	$-\frac{32}{\phi}\Gamma_1 + \frac{24\alpha}{\phi}\Gamma_2$	0

3.2.1 Adjoint Representation of (3.2.6)−(3.2.11)

The problem of deriving the minimal combination (optimal systems) of subalgebras spanned by these operators is equivalent to finding an optimal system of Lie symmetries (or group invariant solutions). This is possible because there is a connection between the Lie group and the adjoint representation of Lie algebra. We construct the adjoint representation to define an equivalence relation on one-dimensional subalgebras which is generated through $\Gamma_i, i = 1, ..., 6$, by summing the Lie series given in Olver (1993) as

$$Ad\left(\exp\left(\varepsilon\Gamma_i\right)\right)\Gamma_j = \sum_{n=0}^{\infty} \frac{\varepsilon^n}{n!}(Ad\Gamma_i)^n = \Gamma_j - \varepsilon\left[\Gamma_i, \Gamma_j\right] + \frac{\varepsilon^2}{2}\left[\Gamma_i, \left[\Gamma_i, \Gamma_j\right]\right] - \cdots$$

with reference to the table of Lie Brackets (Table 3.2.1), where $[\Gamma_i, \Gamma_j] = \Gamma_i\Gamma_j - \Gamma_j\Gamma_i$, is the Lie Bracket and $\varepsilon \in \mathbb{R}$. However, we adopt the global matrix (discussed in Section 3.3) of the adjoint transformations instead, from which we deduce that

$$Ad(j) = A\left(j, \varepsilon\right)\Gamma_i, \tag{3.2.13}$$

where $A\left(j, \varepsilon\right)$ is some $r \times r$ (here $r = 6$) matrix corresponding to each generator Γ_j The transpose of (3.2.13) yields the $ij-$th entry of Table 3.2.2.

Table 3.2.2: Actions of the Adjoint representation of (3.2.6)−(3.2.11)

Ad	$\widetilde{\Gamma_1}$	$\widetilde{\Gamma_2}$	$\widetilde{\Gamma_3}$	$\widetilde{\Gamma_4}$	$\widetilde{\Gamma_5}$	$\widetilde{\Gamma_6}$
$\widetilde{\Gamma_1}$	Γ_1	Γ_2	$e^{-\frac{\phi}{2}\varepsilon}\Gamma_3$	$e^{\frac{\phi}{2}\varepsilon}\Gamma_4$	$e^{-\phi\varepsilon}\Gamma_5$	$e^{\phi\varepsilon}\Gamma_6$
$\widetilde{\Gamma_2}$	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6
$\widetilde{\Gamma_3}$	$\Gamma_1+\frac{\phi}{2}\varepsilon\Gamma_3$	Γ_2	Γ_3	$\Gamma_4+\frac{3\phi}{\theta\kappa}\varepsilon\Gamma_2$	Γ_5	$\Gamma_6+\varepsilon\Gamma_4$
$\widetilde{\Gamma_4}$	$\Gamma_1-\frac{\phi}{2}\varepsilon\Gamma_4$	Γ_2	$\Gamma_3-\frac{3\phi}{\theta\kappa}\varepsilon\Gamma_2$	Γ_4	$\Gamma_5+\varepsilon\Gamma_3$	Γ_6
$\widetilde{\Gamma_5}$	$\Gamma_1+\phi\varepsilon\Gamma_5$	Γ_2	Γ_3	$\Gamma_4+\varepsilon\Gamma_3$	Γ_5	$\Gamma_6-\frac{32}{\phi}\varepsilon\Gamma_1+\frac{24\alpha}{\phi}\varepsilon\Gamma_2$
$\widetilde{\Gamma_6}$	$\Gamma_1-\phi\varepsilon\Gamma_6$	Γ_2	$\Gamma_3+\varepsilon\Gamma_4$	Γ_4	$\Gamma_5+\frac{32}{\phi}\varepsilon\Gamma_1-\frac{24\alpha}{\phi}\varepsilon\Gamma_2$	Γ_6

3.2.2 Construction of invertible mapping: Transformation of (3.1.3) to the heat equation

Bluman and Cole (1974) and Bluman (1980, 1983) proved that the heat equation is the only polynomial partial differential equation of the second order in two independent variables invariant under the finite group of the heat equation itself. Bluman and Kumei (1989), in Chapter Six, provide not only the framework for the existence and construction of a transformation between two (linear or nonlinear) partial differential equations but an algorithm to determine whether the necessary and sufficient conditions for the linearisation of a partial differential equation are satisfied. In this Subsection we construct a transformation which maps our given partial differential equation (3.1.3) to a target partial differential equation, the heat equation. If such a mapping exists, it is necessary that any infinitesimal generator admitted by (3.1.3) be mapped into an infinitesimal generator admitted by the heat equation. The necessary and sufficient conditions for the existence of an invertible mapping are now stated from Bluman and Kumei (1989).

Theorem 3.2.1 (Bluman and Kumei)**.** *In the case of one dependent variable, u, a mapping μ defines an invertible mapping from $\left(x,u,u_{(1)},...,u_{(p)}\right)-$space to $\left(z,w,w_{(1)},...,w_{(p)}\right)-$space for any fixed p if and only if u, is a one-to-one contact transformation of the form*

$$z=\phi\left(x,u,u_{(1)}\right), \tag{3.2.14}$$

$$w = \psi\left(x, u, u_{(1)}\right), \tag{3.2.15}$$

$$w_{(1)} = \psi_{(1)}\left(x, u, u_{(1)}\right). \tag{3.2.16}$$

Note that, if ϕ and ψ are independent of $u_{(1)}$, then (3.2.14)−(3.2.16) define a point transformation.

Theorem 3.2.2 (Necessary conditions for the existence of an invertible mapping)**.** *If there exists an invertible transformation μ which maps a given nonlinear partial differential equation $R\{x,t,u\}$ to a linear system of partial differential equation $S\{\boldsymbol{z}, w\}$, then*

(a) the mapping must be a point transformation of the form

$$z_j = \phi_j(x,t,u), \quad j = 1,2, \tag{3.2.17}$$

$$w = \varphi(x,t,u); \tag{3.2.18}$$

(b) $R\{x,t,u\}$ must admit an infinite-parameter Lie group of point transformations having infinitesimal generator

$$\Gamma = \xi(x,t,u)\frac{\partial}{\partial x} + \tau(x,t,u)\frac{\partial}{\partial t} + \eta(x,t,u)\frac{\partial}{\partial u} \tag{3.2.19}$$

with $\xi(x,t,u), \tau(x,t,u)$ and $\eta(x,t,u)$ characterised by

$$\xi(x,t,u) = \alpha(x,t,u)F(x,t,u) \tag{3.2.20}$$

$$\tau(x,t,u) = \beta(x,t,u)F(x,t,u) \tag{3.2.21}$$

$$\eta(x,t,u) = \varphi(x,t,u)F(x,t,u), \tag{3.2.22}$$

where $\alpha(x,t,u), \beta(x,t,u)$ and $\varphi(x,t,u)$ are specific functions of (x,t,u) and F is an arbitrary solution of some linear system of partial differential equations

$$L[\Gamma]F = 0 \tag{3.2.23}$$

with $L[\Gamma]$ representing a linear differential operator depending upon independent variables

$$\Gamma = (\Gamma_1(x,t,u), \Gamma_2(x,t,u)) \tag{3.2.24}$$

of the same order as the order of the partial differential equation $R\{x,t,u\}$.

Theorem 3.2.3 (Sufficient conditions for the existence of an invertible mapping)**.** *Let a given nonlinear system of partial differential equation* $R\{x,t,u\}$ *admit an infinitesimal generator (3.2.20) the coefficients of which are of the form (3.2.20)−(3.2.22) with* F *being an arbitrary solution of a linear system, (3.2.19), with specific independent variables (3.2.24). If the linear homogeneous system of* m *first-order partial differential equations for scalar* Φ,

$$\alpha(x,t,u)\frac{\partial\Phi}{\partial x}+\beta(x,t,u)\frac{\partial\Phi}{\partial t}+\varphi(x,t,u)\frac{\partial\Phi}{\partial u}=0 \tag{3.2.25}$$

has two functionally independent solutions, $\Gamma_1(x,t,u)$ *and* $\Gamma_2(x,t,u)$, *and the linear first-order partial differential equation,*

$$\alpha(x,t,u)\frac{\partial\psi}{\partial x}+\beta(x,t,u)\frac{\partial\psi}{\partial t}+\varphi(x,t,u)\frac{\partial\psi}{\partial u}=1, \tag{3.2.26}$$

has the solution,

$$\psi=(\psi_1(x,t,u),\psi_2(x,t,u)),$$

then the invertible mapping μ *given by*

$$z_1=\phi_1(x,t,u)=\Gamma_1(x,t,u), \tag{3.2.27}$$

$$z_2=\phi_2(x,t,u)=\Gamma_2(x,t,u), \tag{3.2.28}$$

$$w=\psi(x,t,u) \tag{3.2.29}$$

transforms $R\{x,t,u\}$ *to a linear partial differential equation* $S\{\mathbf{z},w\}$

$$L[\mathbf{z}]\,w=g(\mathbf{z})$$

for some nonhomogeneous term $g(\mathbf{z})$.

NB: $\mathbf{z}$ in Theorems 3.3.2 and 3.3.3 is in fact $\mathbf{z}=(z_1,z_2)$.

We apply these theorems to the construction of an invertible point mapping by finding the point transformation which relates our given equation to the heat equation. Since the generators Γ_4 and Γ_6 of (3.2.9)−(3.2.11) commute, we use them to construct a transformation that maps (3.1.3) invertibly to the heat equation. Let $X_1:=\Gamma_4$ and $X_2:=\Gamma_6$. From the vectors of the infinitesimal symmetries of (3.1.3)

$$X_1=\xi_{11}(x,t)\frac{\partial}{\partial x}+\xi_{12}(x,t)\frac{\partial}{\partial t}+f_1(x,t)\,u\frac{\partial}{\partial u} \tag{3.2.30}$$

$$X_2 = \xi_{21}(x,t)\frac{\partial}{\partial x} + \xi_{22}(x,t)\frac{\partial}{\partial t} + f_2(x,t)\,u\frac{\partial}{\partial u}, \tag{3.2.31}$$

where

$$\xi_{11} = \sqrt{x}e^{\frac{1}{2}\phi t} \qquad \xi_{12} = 0 \qquad f_1 = \left(\frac{1}{2\sqrt{x}} - \mathcal{B}\sqrt{x}\right)ue^{\frac{1}{2}\phi t}$$

$$\xi_{21} = xe^{-\phi t} \qquad \xi_{22} = \frac{1}{\phi}e^{-\phi t} \qquad f_2 = -\left(\mathcal{B}x - \mathcal{D}\right)e^{-\phi t}$$

To verify that the Jacobian of the transformation is nonzero, we find the determinant

$$J = \begin{vmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{vmatrix} = \frac{\sqrt{x}}{\phi}e^{-\frac{3}{2}\phi t} \neq 0, \quad \phi \neq 0.$$

Since $J \neq 0$, then the existence of an invertible mapping of the form

$$\left.\begin{aligned} z &= \alpha(x,t) \\ \tau &= \varphi(x,t) \\ \omega &= \nu(x,t)\,u \end{aligned}\right\} \tag{3.2.32}$$

is guaranteed for the mapping of (3.1.3) into a constant partial differential equation. The mapping (3.2.32) must satisfy the following necessary conditions

$$\xi_{11}\alpha_x + \xi_{12}\alpha_t = 1; \qquad \xi_{21}\alpha_x + \xi_{22}\alpha_t = 0 \tag{3.2.33}$$

$$\xi_{11}\varphi_x + \xi_{12}\varphi_t = 0; \qquad \xi_{21}\varphi_x + \xi_{22}\varphi_t = 1 \tag{3.2.34}$$

$$\xi_{11}\nu_x + \xi_{12}\nu_t + f_1\nu = 0; \qquad \xi_{21}\nu_x + \xi_{22}\nu_t + f_2\nu = 0 \tag{3.2.35}$$

Substitution of the values of the $\xi_{ij's}$ into (3.2.33)$-$(3.2.35) and solution of the resulting simultaneous systems yield

$$\left.\begin{aligned} \sqrt{x}e^{-\frac{1}{2}\phi t}\alpha_x + 0 &= 1 \\ xe^{-\phi t}\alpha_x - \tfrac{1}{\phi}e^{-\phi t}\alpha_t &= 0 \end{aligned}\right\} \quad \begin{aligned} \alpha_x &= \tfrac{1}{\sqrt{x}}e^{\frac{1}{2}\phi t} \\ \alpha_t &= -\phi\sqrt{x}e^{\frac{1}{2}\phi t} \end{aligned} \tag{3.2.36}$$

$$\left.\begin{aligned} \sqrt{x}e^{-\frac{1}{2}\phi t}\varphi_x + 0 &= 0 \\ xe^{-\phi t}\varphi_x - \tfrac{1}{\phi}e^{-\phi t}\varphi_t &= 1 \end{aligned}\right\} \quad \begin{aligned} \varphi_x &= 0 \\ \varphi_t &= -\phi e^{\phi t} \end{aligned} \tag{3.2.37}$$

$$\left.\begin{aligned} \sqrt{x}e^{-\frac{1}{2}\phi t}\nu_x + 0 &= \left(\tfrac{1}{2\sqrt{x}} - \mathcal{B}\sqrt{x}\right)e^{-\frac{1}{2}\phi t}u \\ xe^{-\phi t}\nu_x - \tfrac{1}{\phi}e^{-\phi t}\nu_t &= \left(\mathcal{B}x - \mathcal{D}\right)e^{-\phi t} \end{aligned}\right\} \quad \begin{aligned} \nu_x &= -\tfrac{1}{\phi}\left(2\mathcal{B}x - 1\right)u \\ \nu_t &= \tfrac{\phi}{2}\left(\mathcal{D} - \mathcal{B}x + 1\right)u. \end{aligned} \tag{3.2.38}$$

From (3.2.36)

$$\alpha(x,t) = \int \alpha_x dx = 2\sqrt{x}e^{\frac{1}{2}\phi t} + a(t).$$

$$\therefore a(t) = A_1 \Longrightarrow \alpha(x,t) = \mathcal{K} + \frac{2\sqrt{x}}{\phi}(2\phi - 1)e^{\frac{1}{2}\phi t}.$$

From (3.2.37)

$$\varphi(x,t) = \int \varphi_x dx = b(t)$$

$$\frac{\partial}{\partial t}\varphi(x,t) = \varphi_t \Longrightarrow \dot{b}(t) = -\phi e^{\phi t}, \quad b(t) = \mathcal{L} - e^{\phi t}$$

.

$$\therefore \varphi(x,t) = \mathcal{L} - e^{\phi t}.$$

From (3.2.38)

$$\nu(x,t) = \int \nu_x dx = -u\mathcal{B}x + \frac{1}{2}u\ln x + c(t)$$

$$\frac{\partial}{\partial t}\nu(x,t) = \nu_t \Longrightarrow 0 + \dot{c}(t) = \frac{\phi}{2}(\mathcal{D} - \mathcal{B}x + 1)u \quad and \quad c(t) = \frac{\phi}{2}(\mathcal{D} - \mathcal{B}x + 1)ut + \mathcal{M}$$

$$\therefore \ \nu(x,t) = \mathcal{M}\exp\left\{\frac{\phi}{2}(\mathcal{D} - \mathcal{B}x + 1)t - \mathcal{B}x + \frac{1}{2}\ln x\right\},$$

where $\mathcal{K}, \mathcal{L}$ and $\mathcal{M}$ are arbitrary constants. Without loss of generality we may let $\mathcal{K} = \mathcal{L} = 0$ and $\mathcal{M} = 1$ to have (3.2.39)−(3.2.41).

$$z = \frac{2}{\phi}\sqrt{x}(2\phi - 1)e^{-\frac{1}{2}\phi t}, \tag{3.2.39}$$

$$\tau = -e^{\phi t} \tag{3.2.40}$$

and

$$w(z,\tau) = u(x,t)\exp\left\{\frac{\phi}{2}(\mathcal{D} - \mathcal{B}x + 1)\,\mathrm{t} - \mathcal{B}x + \frac{1}{2}\ln x\right\}. \tag{3.2.41}$$

We now map (3.1.3) invertibly to the heat equation by using the transformation (3.2.39)−(3.2.41). Using calculus and especially the principle of the chain rule, we obtain

$$\frac{\partial u}{\partial t} = \left\{\frac{x^{-\frac{1}{2}}}{\phi}(2\phi - 1)e^{\frac{1}{2}t}\frac{\partial w}{\partial z} - \phi e^{\phi t}\frac{\partial w}{\partial \tau} - \frac{\phi}{2}\sqrt{x}(\mathcal{B}x - \mathcal{D} - 1)w\right\}e^{\psi} \tag{3.2.42}$$

$$\frac{\partial u}{\partial x} = \left\{\frac{x^{-\frac{1}{2}}}{\phi}(2\phi - 1)e^{\frac{1}{2}t}\frac{\partial w}{\partial z} - \frac{x^{-\frac{1}{2}}}{2}(\phi\mathcal{B}tx + 2\mathcal{B}x - 1)w\right\}e^{\psi} \tag{3.2.43}$$

and

$$\frac{\partial^2 u}{\partial x^2} = \left\{ \begin{array}{l} \frac{x^{-1}}{\phi^2}\left(2\phi-1\right)^2 e^{\phi t}\frac{\partial^2 w}{\partial z^2} - \frac{x^{-\frac{3}{2}}}{2\phi}\left(2\phi-1\right)e^{\frac{1}{2}\phi t}\frac{\partial w}{\partial z} \\ +\frac{1}{4}\left[\left(\phi\mathcal{B}tx+2\mathcal{B}x-1\right)x^{-1}-x^{-\frac{3}{2}}\right]\left(\phi\mathcal{B}tx+2\mathcal{B}x-1\right)w \end{array} \right\} e^{\psi} \tag{3.2.44}$$

where $\psi = \frac{\phi}{2}\left(\mathcal{D}-\mathcal{B}x+1\right)t - \mathcal{B}x$. Substituting (3.2.29)−(3.2.31) into (3.1.3) we obtain the linear heat equation

$$\frac{\partial w}{\partial \tau} - \mathcal{H}\frac{\partial^2 w}{\partial z^2} = 0, \tag{3.2.45}$$

where

$$\mathcal{H} = \frac{2\theta\kappa}{3\phi^3}(2\phi-1)^2.$$

3.3 Construction of an Optimal System of One-dimensional Subalgebras

It is well known that reduction of the independent variables of a partial differential equation by one is possible using any linear combination of our symmetry generators $\Gamma_i, i = 1, ..., 6$. We now construct a set of minimal combinations known as an optimal system for which the commutators of the admitted symmetries given in Table 3.2.1 are an essential component.

3.3.1 Construction of linear transformation

As explained in Olver (1993), the infinite-dimensionl subalgebra Γ_∞ does not lead to group invariant solutions. Consequently it is not considered in the classification problem. We need to classify the generators for a particular algebra. This classification applies to every differential equation with a Lie algebra whether ordinary or partial differential equation. The approaches for the classification of symmetry generators can be found in Bluman and Kumei (1989), Ibragimov (1999), Ibragimov *et al* (1991) and Hydon (2000). Recall that any Lie algebra is closed under the operation of taking the Lie Bracket. We split Γ and $\tilde{\Gamma}$ into components as follows:

$$\Gamma = \kappa^i \Gamma_i \tag{3.3.1}$$

and

$$\tilde{\Gamma} = \kappa^i \tilde{\Gamma}_i = \kappa^i e^{-\varepsilon \Gamma_j} \Gamma_i e^{\varepsilon \Gamma_j}, \tag{3.3.2}$$

where

$$\tilde{\Gamma}_i = \left(A\left(j, \varepsilon\right)\right)_i^m \Gamma_m. \tag{3.3.3}$$

$A\left(j, \varepsilon\right)$ is some matrix corresponding to each generator Γ_j. To obtain $A\left(j, \varepsilon\right)$ Hydon used the generator $\tilde{\Gamma}_i$. which is the solution of the initial-value problem

$$\frac{d\tilde{\Gamma}_i}{d\varepsilon} = -e^{-\varepsilon \Gamma_j} \left[\Gamma_j, \Gamma_i\right] e^{\varepsilon \Gamma_j} = C_{ij}^k \tilde{\Gamma}_k$$

$$\tilde{\Gamma}_i \Big|_{\varepsilon=0} = \Gamma_i$$

Wherefore from (3.3.3)

$$\frac{d\left(A\left(j, \varepsilon\right)\right)_i^m}{d\varepsilon} \Gamma_m = C_{ij}^k \left(A\left(j, \varepsilon\right)\right)_k^m \Gamma_m,$$

$$\left(A\left(j, 0\right)\right)_i^m \Gamma_m = \Gamma_i.$$

The generators Γ_m are linearly independent and so

$$\frac{d\left(A\left(j, \varepsilon\right)\right)_i^m}{d\varepsilon} = C_{ij}^k \left(A\left(j, \varepsilon\right)\right)_k^m, \quad \left(A\left(j, 0\right)\right)_i^m = \delta_i^m, \tag{3.3.4}$$

where the structure constants C_{ij}^k can be used to define the matrix $C\left(j\right)$ with the following relationship

$$\left(C\left(j\right)\right)_i^k = C_{ij}^k. \tag{3.3.5}$$

Using (3.3.4) we have the matrix differential equation

$$\frac{dA\left(j, \varepsilon\right)}{d\varepsilon_j} = C\left(j\right) A\left(j, \varepsilon\right), \quad A\left(j, 0\right) = I, \tag{3.3.6}$$

with the general solution

$$A(j, \varepsilon_j) = \exp\left[\varepsilon_j C(j)\right] = \sum_{n=0}^{\infty} C(j)^n \frac{\varepsilon_j^n}{n!}. \tag{3.3.7}$$

The necessary (and sufficient) condition for invariants of $I\left(\kappa\right)$ such that

$$I\left(\kappa \exp\left[\varepsilon_j C\left(j\right)\right]\right) = I\left(\kappa\right) \quad \forall j, \varepsilon, \tag{3.3.8}$$

for which

$$\left.\frac{\partial I\left(\kappa \exp\left[\varepsilon_j C(j)\right]\right)}{\partial \varepsilon}\right|_{\varepsilon=0} = 0 \tag{3.3.9}$$

is[1]

$$K\left(\kappa\right) \nabla I\left(\kappa\right) = \mathbb{O}. \tag{3.3.10}$$

When we consider the nonvanishing two-dimensional Lie algebra the basis of which is presented as the elements of the Lie Brackets in Table 3.2.1, the only nonzero structure constants C_{ij}^{k} with $j=1$ are $C_{31}^{3}=-\phi/2, C_{41}^{4}=\phi/2, C_{51}^{5}=-\phi, C_{61}^{6}=\phi$ and for $j=3$ are $C_{13}^{3}=\phi/2, C_{43}^{2}=3\phi/\theta\kappa, C_{63}^{4}=1$ and so on. When we apply (3.3.5), we obtain

$$C\left(1\right)=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{\phi}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\phi}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\phi & 0 \\ 0 & 0 & 0 & 0 & 0 & \phi \end{pmatrix}, \quad C(2)=\mathbb{O}_{6\times 6},$$

$$C\left(3\right)=\begin{pmatrix} 0 & 0 & \frac{\phi}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{3\phi}{\theta\kappa} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad C(4)=\begin{pmatrix} 0 & 0 & 0 & \frac{-\phi}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{-3\phi}{\theta\kappa} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$C\left(5\right)=\begin{pmatrix} 0 & 0 & 0 & 0 & \phi & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{-32}{\phi} & \frac{24\alpha}{\phi} & 0 & 0 & 0 & 0 \end{pmatrix}, \quad C\left(6\right)=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -\phi \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{32}{\phi} & \frac{-24\alpha}{\phi} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

[1]Using MATHEMATICA we generate the matrix $A\left(j,\varepsilon_j\right)$ using $C\left(j\right)$ with the command *MatrixExp*$\left[\varepsilon_j \left\{\{...\},...,\{...\}\right\}\right]$, where the $\{...\}$ represents the ith row of $C\left(j\right)$.

When we utilize (3.3.7), the exponentiated matrices $\exp\left[\varepsilon_j C(j)\right] = A(j,\varepsilon_j)$ are

$$A\left(1,\varepsilon_1\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-\frac{\phi\varepsilon_1}{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{\frac{\phi\varepsilon_1}{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{-\phi\varepsilon_1} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{\phi\varepsilon_1} \end{pmatrix}, \quad A\left(2,\varepsilon_2\right) = I_{6\times 6},$$

$$A\left(3,\varepsilon_3\right) = \begin{pmatrix} 1 & 0 & \frac{\phi}{2}\varepsilon_3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{3\phi}{\theta\kappa}\varepsilon_3 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{3\phi}{2\theta\kappa}\varepsilon_3^2 & 0 & \varepsilon_3 & 0 & 1 \end{pmatrix}, \quad A\left(4,\varepsilon_4\right) = \begin{pmatrix} 1 & 0 & 0 & \frac{-\phi}{2}\varepsilon_4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{-3\phi}{\theta\kappa}\varepsilon_4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{-3\phi}{2\theta\kappa}\varepsilon_4^2 & \varepsilon_4 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$A\left(5,\varepsilon_5\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & \phi\varepsilon_5 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\varepsilon_5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{-32}{\phi}\varepsilon_5 & \frac{24\alpha}{\phi}\varepsilon_5 & 0 & 0 & -16\varepsilon_5^2 & 1 \end{pmatrix},$$

$$A\left(6,\varepsilon_6\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & -\phi\varepsilon_6 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-\varepsilon_6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{32}{\phi}\varepsilon_6 & \frac{-24\alpha}{\phi}\varepsilon_6 & 0 & 0 & 1 & -16\varepsilon_6^2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The transformations generated by the $A\left(j,\varepsilon\right)'s$ are

$$E_i = \left(\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6\right)\left[A(j,\varepsilon_j)\right] \tag{3.3.11}$$

$$
\begin{aligned}
E_1 &= \left\{\lambda_1, \lambda_2, e^{-\frac{\phi}{2}\varepsilon_1}\lambda_3, e^{\frac{\phi}{2}\varepsilon_1}\lambda_4, e^{-\phi\varepsilon_1}\lambda_5, e^{\phi\varepsilon_1}\lambda_6\right\} \\
E_2 &= \{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6\} \\
E_3 &= \left\{\lambda_1, \lambda_2 + \frac{3\phi}{\theta\kappa}\lambda_4\varepsilon_3 + \frac{3\phi}{2\theta\kappa}\lambda_6\varepsilon_3^2, \lambda_3 + \frac{\phi}{2}\lambda_1\varepsilon_3, \lambda_4 + \lambda_6\varepsilon_3, \lambda_5, \lambda_6\right\} \\
E_4 &= \left\{\lambda_1, \lambda_2 - \frac{3\phi}{\theta\kappa}\lambda_3\varepsilon_4 - \frac{3\phi}{2\theta\kappa}\lambda_5\varepsilon_4^2, \lambda_3 + \lambda_5\varepsilon_4, \lambda_4 - \frac{\phi}{2}\lambda_1\varepsilon_4, \lambda_5, \lambda_6\right\} \\
E_5 &= \left\{\lambda_1 - \frac{32}{\phi}\lambda_6\varepsilon_5, \lambda_2 + \frac{24\alpha}{\phi}\lambda_6\varepsilon_5, \lambda_3 - \lambda_4\varepsilon_5, \lambda_4, \lambda_5 + \phi\lambda_1\varepsilon_5 - 16\lambda_6\varepsilon_5^2, \lambda_6\right\} \\
E_6 &= \left\{\lambda_1 + \frac{32}{\phi}\lambda_5\varepsilon_6, \lambda_2 - \frac{24\alpha}{\phi}\lambda_5\varepsilon_6, e^{-\varepsilon_2}\lambda_3, \lambda_4, \lambda_5, \lambda_6 - \phi\lambda_1\varepsilon_6 - 16\lambda_5\varepsilon_6^2\right\}.
\end{aligned}
$$

3.3.2 One functionally invariant solutions

The invariance condition (3.3.10) is not of full rank and may be solved by the method of characteristics. This means that the transformations $(E_1) - (E_6)$ have precisely one functional invariant. The integration of the equations

$$
E_i\,(D) = 0 \qquad i = 1, ..., 6,
$$

$$
\lambda_5 + \phi\lambda_1\varepsilon_5 - 16\lambda_6\varepsilon_5^2 = 0,
$$

$$
D = (\phi\lambda_1)^2 + 64\lambda_5\lambda_6. \tag{3.3.12}
$$

The invariant (3.3.12) (which is simply the discriminant of the quadratic in any E_i) simplifies further calculations immediately. The solution of the quadratic in E_5 is

$$
\varepsilon_5 = \frac{\lambda_1^2 \pm \sqrt{D}}{32\lambda_6}. \tag{3.3.13}
$$

As D is quadratic in the components of λ_1, rescaling can only multiply D by a positive constant. Hence we must consider the three distinct conditions: $D > 0, D = 0$ and $D < 0$.

CASES

Case I: The Case $D > 0$ and $\lambda_6 = 0$

We bifurcate this case into the following subcases namely: (a) $\lambda_5 \neq 0$ and (b) $\lambda_5 = 0$

(a) $\lambda_6 = 0$ and $\lambda_5 \neq 0$

Consider the vector $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$. Then by subcase (a) we have

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, 0), \tag{3.3.14}$$

where $\lambda_5 \neq 0$. Using λ_5 to reduce the vector given above in E_4, we obtain $\lambda_3 + \lambda_5\varepsilon_4 = 0$. Therefore $\varepsilon_4 = -\lambda_3/\lambda_5 \Rightarrow \lambda_3 = 0$. From E_6 $\lambda_1 + 32\lambda_5\varepsilon_6/\phi = 0$ and so $\varepsilon_6 = -\phi\lambda_1/32\lambda_5 \Rightarrow \lambda_1 = 0$ and $\lambda_2 - 24\alpha\lambda_5\varepsilon_6/\phi = 0$ from which $\varepsilon_6 = \phi\lambda_2/2\alpha\lambda_5 \Rightarrow \lambda_2 = 0$. The above vector therefore reduces to the form

$$(0, 0, 0, \lambda_4, \lambda_5, 0). \tag{3.3.15}$$

a(i) We can make $\lambda_5 = \pm 1$ to obtain the following representative for the optimal system

$$\Gamma_4 + \Gamma_5 \quad \textit{and} \quad \Gamma_4 - \Gamma_5. \tag{3.3.16}$$

a(ii) If we consider Γ_4 in (3.3.15) and divide through by Γ_5, we have the reduced vector as

$$(0, 0, 0, k, 1, 0). \tag{3.3.17}$$

If $\lambda_4 = 0$ in (3.3.15), we obtain a reduced vector given by

$$(0, 0, 0, 0, \lambda_5, 0).$$

These operations produce the following representative for the optimal system

$$k\Gamma_4 + \Gamma_5 \quad \text{and} \quad \Gamma_5. \tag{3.3.18}$$

(b) $\lambda_6 = 0$, and $\lambda_5 = 0$

In this subcase we have the vector

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, 0, 0). \tag{3.3.19}$$

b(i) Suppose $\lambda_1 \neq 0$. Then from E_3, $\lambda_3 + \frac{1}{2}\phi\lambda_1\varepsilon_3 = 0$. Therefore $\varepsilon_3 = -2\lambda_3/\phi\lambda_1 \Longrightarrow \lambda_3 = 0$. Again from E_4, $\lambda_4 - \frac{1}{2}\phi\lambda_1\varepsilon_4 = 0$. So that $\varepsilon_4 = 2\lambda_4/\phi\lambda_1 \Longrightarrow \lambda_4 = 0$. Hence we now have the vector

$$(\lambda_1, \lambda_2, 0, 0, 0, 0). \tag{3.3.20}$$

From (3.3.20) we have another representative for the optimal system

$$\Gamma_1 + \Gamma_2. \tag{3.3.21}$$

b(ii) Suppose $\lambda_1 = 0$ in (3.3.19). We then consider the vector

$$(0, \lambda_2, \lambda_3, \lambda_4, 0, 0)\,. \tag{3.3.22}$$

Based on the vector (3.3.22) we consider $\lambda_4 \neq 0$, then using the transformation E_3, $\quad \lambda_2 + 3\phi\lambda_4\varepsilon_3/\theta\kappa + 3\phi\lambda_6\varepsilon_3^2/2\theta\kappa = 0$ and, because $\lambda_6 = 0$ already, $\varepsilon_3 = -\theta\kappa\lambda_2/3\phi\lambda_4$. This implies that $\lambda_2 = 0$ and from transformation E_5, $\quad \lambda_3 - \lambda_4\varepsilon_5 = 0$ so that $\varepsilon_5 = \lambda_3/\lambda_4 \Longrightarrow \lambda_3 = 0$. Hence we now obtain a new reduced vector

$$(0, 0, 0, \lambda_4, 0, 0)\,. \tag{3.3.23}$$

The vector (3.3.23) provides the optimal system representative given by

$$\Gamma_4. \tag{3.3.24}$$

When we consider $\lambda_4 = 0$ in (3.3.22), we produce the reduced vector

$$(0, \lambda_2, \lambda_3, 0, 0, 0)\,. \tag{3.3.25}$$

$b\,(ii')$ If $\lambda_3 \neq 0$ in (3.3.26), then from E_4, $\quad \lambda_2 - 3\phi\lambda_3\varepsilon_4/\theta\kappa = 0$ (since $\lambda_5 = 0$ already) $\varepsilon_4 = \theta\kappa\lambda_2/3\phi\lambda_3 \Longrightarrow \lambda_2 = 0$ If we take into account the possibility that $\lambda_3 = 0$, we then have the new reduced vectors

$$(0, \lambda_2, 0, 0, 0, 0) \text{ and } (0, 0, \lambda_3, 0, 0, 0) \tag{3.3.26}$$

which yield a representation for the optimal systems given by

$$\Gamma_2 \text{ and } \Gamma_3. \tag{3.3.27}$$

Case II: The Case $D = 0$ and $\lambda_6 \neq 0$

For this condition we use (3.3.27), i.e., $\varepsilon_5 = \lambda_1^2/32\lambda_6$ which implies that $\lambda_1 = 0$. This procedure yields the new vector

$$(0, \lambda_2, \lambda_3, \lambda_4, 0, \lambda_6)\,. \tag{3.3.28}$$

Furthermore, if $\lambda_1 = 0$, we have $D = 62\lambda_5\lambda_6$ but $D = 0$ yields $\lambda_5 = 0$ since $\lambda_6 \neq 0$ and we obtain the same vector as (3.3.28). When we use $E_3, \lambda_4 + \lambda_6\varepsilon_3 = 0$, where $\varepsilon_3 = -\lambda_4/\lambda_6$ so that $\lambda_4 = 0$. Similarly from $E_5, \varepsilon_5 = -\phi\lambda_2/24\alpha\lambda_6 \Rightarrow \lambda_2 = 0$. The new vector is

$$(0, 0, \lambda_3, 0, 0, \lambda_6)\,. \tag{3.3.29}$$

The representative for the optimal system is

$$\Gamma_3 + \Gamma_6. \tag{3.3.30}$$

Case III: The Case $D < 0$ and $\lambda_6 \neq 0$

$D < 0$ means that $(\phi\lambda_1)^2 + 62\lambda_5\lambda_6 < 0$, that is $\lambda_5 \neq 0$. We can apply the transformations E_3, E_4 and E_5 respectively, as follows

$\lambda_4 + \lambda_6\varepsilon_3 = 0, \quad i.e \quad \varepsilon_3 = -\lambda_4/\lambda_6 \Longrightarrow \lambda_4 = 0$

$\lambda_3 + \lambda_5\varepsilon_4 = 0, \quad i.e., \quad \varepsilon_4 = -\lambda_3/\lambda_5 \Longrightarrow \lambda_3 = 0.$

$\lambda_1 - 32\lambda_6\varepsilon_5/\phi = 0, \quad i.e., \quad \varepsilon_5 = \varphi\lambda_1/32\lambda_6 \Longrightarrow \lambda_1 = 0.$

$\lambda_2 + 24\alpha\lambda_6\varepsilon_5/\phi = 0, \quad i.e., \quad \varepsilon_5 = -\phi\lambda_2/24\alpha\lambda_6 \Longrightarrow \lambda_2 = 0.$

The above operations reduce the vector to

$$(0, 0, 0, 0, \lambda_5, \lambda_6)\,. \tag{3.3.31}$$

The components λ_5 and λ_6 of the vector (3.3.31) do not have a common sign since $D < 0$ and $\lambda_1 = 0$ already. For $D < 0$ we need $\lambda_5\lambda_6 < 0$. The condition is that either $\lambda_5 = +1$ and $\lambda_6 = -1$ or $\lambda_5 = -1$ and $\lambda_6 = +1$. Vector (3.3.31) then yields the representation for the optimal system as

$$\Gamma_5 - \Gamma_6 \text{ and } -\Gamma_5 + \Gamma_6. \tag{3.3.32}$$

Case IV: The Case $D > 0$ and $\lambda_6 \neq 0$

Since D is an invariant under the transformations $E_1, E_3 - E_6$, the condition $D > 0$ shows that either $\lambda_1 \neq 0$ or $\lambda_5 \neq 0$. Another possibility is that $\lambda_1 \neq 0$ and $\lambda_5 = 0$ since $\lambda_6 \neq 0$ and condition $D > 0$ is still met, or $\lambda_1 = 0$ and both λ_5 and λ_6 have the same signs such that $\lambda_5\lambda_6 > 0$ (when $\lambda_1 = 0$).

$\lambda_6 \neq 0, \quad \lambda_1 \neq 0 \quad \text{and} \quad \lambda_5 \neq 0$

From $E_3 - E_6$, respectively, we have

$\lambda_4 + \lambda_6\varepsilon_3 = 0, \quad i.e., \quad \varepsilon_3 = -\lambda_4/\lambda_6 \Longrightarrow \lambda_4 = 0$

$\lambda_3 + \lambda_5\varepsilon_5 = 0, \quad i.e., \quad \varepsilon_5 = -\lambda_3/\lambda_5 \Longrightarrow \lambda_3 = 0$

$\lambda_2 + 24\alpha\lambda_6\varepsilon_5/\phi = 0, \quad i.e., \quad \varepsilon_5 = -\phi\lambda_2/24\alpha\lambda_6 \Longrightarrow \lambda_2 = 0$

$\lambda_1 + 32\lambda_5\varepsilon_6/\phi = 0, \quad i.e., \quad \varepsilon_6 = -\phi\lambda_1/32\lambda_6 \Longrightarrow \lambda_1 = 0$

for which we obtain a reduced vector of the form

$$(0, 0, 0, 0, \lambda_5, \lambda_6) \tag{3.3.33}$$

such that the representative for the optimal system is

$$-\Gamma_5 - \Gamma_6 \quad \text{and} \quad \Gamma_5 + \Gamma_6. \tag{3.3.34}$$

- **Summary of the optimal system (3.3.16)−(3.3.34)**

$$\Gamma_2, \quad \Gamma_3, \quad \Gamma_4, \quad \Gamma_5,$$
$$\Gamma_1 + \Gamma_2, \quad \Gamma_3 + \Gamma_6, \quad \Gamma_4 + \Gamma_5, \quad \Gamma_4 - \Gamma_5,$$
$$\Gamma_5 + \Gamma_6, \quad \Gamma_5 - \Gamma_6, \quad -\Gamma_5 + \Gamma_6, \quad -(\Gamma_5 + \Gamma_6), \quad k\Gamma_4 + \Gamma_5.$$

3.3.3 Symmetry reductions and invariant solutions

Having obtained an optimal system of generators, we can use the method of group invariant solutions to calculate the associated invariant solutions. We then start by noting that the invariance condition for the operator Γ_2 is $u = u(t)$ and provides a trivial solution that $u =$ constant.

A: $\Gamma = \Gamma_1 + \Gamma_2$

$$\Gamma = \Gamma_1 + \Gamma_2 = \frac{\partial}{\partial t} + u\frac{\partial}{\partial u}. \tag{3.3.35}$$

The associated Lagrange's system of (3.3.35) is

$$\frac{dt}{1} = \frac{dx}{0} = \frac{du}{u}$$

$$\frac{dt}{1} = \frac{dx}{0} \Rightarrow D_1 = x \tag{3.3.36}$$

$$\frac{dt}{1} = \frac{du}{u}$$
$$D_2 = ue^{-t}. \tag{3.3.37}$$

Invariant checks

(a) $\Gamma(D_1) = (\Gamma_1 + \Gamma_2)(D_1) = \left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial u}\right)(x) = 0.$

Hence the invariance condition is satisfied.

(b) $\Gamma(D_2) = (\Gamma_1 + \Gamma_2)(D_2) = \left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial u}\right)(ue^{-t})$

$$= \frac{d}{dt}ue^{-t} + u\frac{d}{du}ue^{-t} = -ue^{-t} + ue^{-t} = 0.$$

The invariance condition is also satisfied.

Working with D_1 and D_2 and designating one of them as a function of the other we have

$$D_2 = \varphi(D_1), \text{ i.e., } ue^{-t} = \varphi(x) \Rightarrow u = e^t\varphi(x). \tag{3.3.38}$$

Substituting (3.3.38) and the partial derivatives of u with respect to t and x into our given equation we obtain

$$\frac{2}{3}\theta\kappa x\varphi'' + (\theta\kappa - \alpha x)\varphi' + (1 - kx)\varphi = 0, \tag{3.3.39}$$

where $\alpha = \kappa + \lambda$ and $\sigma^2 = 4\theta\kappa/3$.

B: $\Gamma = \Gamma_4 = e^{-\frac{1}{2}\phi t}\left[\sqrt{x}\frac{\partial}{\partial x} - \frac{1}{4\theta\kappa\sqrt{x}}(2\theta\kappa - 3x(\alpha - \phi))u\frac{\partial}{\partial u}\right]$

The associated Lagrange's system is

$$\frac{dt}{0} = \frac{dx}{\sqrt{x}} = \frac{-4\theta\kappa\sqrt{x}du}{u(2\theta\kappa - 3x(\alpha - \phi))}$$

$$\frac{dt}{0} = \frac{dx}{\sqrt{x}} \Rightarrow dt = 0 \Rightarrow D_1 = t \tag{3.3.40}$$

$$\frac{dx}{\sqrt{x}} = \frac{-4\theta\kappa\sqrt{x}du}{u(2\theta\kappa - 3x(\alpha - \phi))}, \quad \text{i.e.,} \quad \frac{-(2\theta\kappa - 3x(\alpha - \phi))dx}{4\theta\kappa x} = \frac{du}{u}$$

$$D_2 = ue^{-\frac{3x}{4\theta\kappa}(\alpha - \phi)}x^{1/2} \quad \text{or} \quad D_2 = u\exp\left(-\ln\sqrt{x} + \frac{3x}{4\theta\kappa}(\alpha - \phi)\right). \tag{3.3.41}$$

Invariant checks:

(a) $\Gamma(D_1) = \Gamma_4(D_1) = e^{-\frac{1}{2}\phi t}\left[\sqrt{x}\frac{\partial}{\partial x} - \frac{1}{4\theta\kappa\sqrt{x}}\left(2\theta\kappa - 3x\left(\alpha-\phi\right)\right)u\frac{\partial}{\partial u}\right](t) = 0.$

This operator satisfies the invariance condition.

(b)

$$\Gamma(D_2) = \Gamma_4(D_2) = e^{-\frac{1}{2}\phi t}\left[\sqrt{x}\frac{\partial}{\partial x} - \frac{1}{4\theta\kappa\sqrt{x}}\left(2\theta\kappa - 3x\left(\alpha-\phi\right)\right)u\frac{\partial}{\partial u}\right]\left(ue^{-\frac{3x}{4\theta\kappa}(\alpha-\phi)}x^{\frac{1}{2}}\right) = 0.$$

This operator also satisfies the invariant condition.

Making D_2 a function of D_1 we have

$$D_2 = \varphi\left(D_1\right).$$

We obtain the invariant

$$u\exp\left\{-\frac{3x}{4\theta\kappa}\left(\alpha-\phi\right)\right\}\sqrt{x} = \varphi\left(\mathrm{t}\right) \;\Rightarrow u = \exp\left\{\frac{3x}{4\theta\kappa}\left(\alpha-\phi\right)\right\}x^{-1/2}\;\varphi\left(\mathrm{t}\right). \tag{3.3.42}$$

Differentiating (3.3.42) partially with respect to t and x we obtain

$$u_t = \exp\left\{\frac{3x}{4\theta\kappa}\left(\alpha-\phi\right)\right\}x^{-1/2}\varphi' \tag{3.3.43}$$

$$u_x = \frac{3}{4\theta\kappa}\left(\alpha-\phi\right)x^{-1/2}\exp\left\{\frac{3x}{4\theta\kappa}\left(\alpha-\phi\right)\right\}\varphi - \frac{1}{2}x^{-1/2}\exp\left\{\frac{3x}{4\theta\kappa}\left(\alpha-\phi\right)\right\}\varphi \tag{3.3.44}$$

$$u_{xx} = \left[\left(\frac{3}{4\theta\kappa}\left(\alpha-\phi\right)\right)^2 - \frac{3}{4\theta\kappa}\left(\alpha-\phi\right)x + \frac{3}{4}x^{-2}\right]x^{-1/2}\exp\left\{\frac{3x}{4\theta\kappa}\left(\alpha-\phi\right)\right\}\varphi. \tag{3.3.45}$$

Substitution of (3.3.43)−(3.3.45) into our given equation (3.1.3) gives

$$\varphi' + \left\{\left[\frac{3x}{8\theta\kappa}\left(\alpha-\phi\right) - \frac{x^2}{2} + \frac{3}{4\theta\kappa}\left(\theta-\alpha x\right)\right]\left(\alpha-\phi\right) + \frac{\theta\kappa}{2x} - \frac{1}{2}\left(\theta\kappa-\alpha x\right) - kx\right\}\varphi = 0, \tag{3.3.46}$$

where $\alpha = \kappa + \lambda$ and $\sigma^2 = 4\theta\kappa/3$.
Let $A = \left[\frac{3x}{8\theta\kappa}\left(\alpha-\phi\right) - \frac{x^2}{2} + \frac{3}{4\theta\kappa}\left(\theta-\alpha x\right)\right]\left(\alpha-\phi\right) + \frac{\theta\kappa}{2x} - \frac{1}{2}\left(\theta\kappa-\alpha x\right) - kx$
so that (3.3.46) becomes

$$\varphi' + A\varphi = 0. \tag{3.3.47}$$

The solution of (3.3.47) yields

$$\varphi = Ke^{-At_1} \tag{3.3.48}$$

3.4 Chapter summary

In this Chapter we focused on the solution of the partial differential equation for pricing future contracts for electricity given in equation (3.1.1). We determined the symmetries admitted by this partial differential equation and used them to construct the solutions in line with the Lie algorithm reviewed in Chapter Two. Our calculations showed that our model for pricing future contracts for electricity admits six Lie point symmetries and a solution symmetry. We determined the Lie symmetry algebra admitted by (3.1.3) and found that it may be decomposed into $\{sl(2,R) \oplus W_3\} \oplus_s \infty A_1$. We also constructed a transformation that map the symmetries admitted by our given equation into a constant coefficient heat equation. In addition we derived the adjoint representation group which was used in the construction of a one-dimensional optimal system for equation (3.1.3). Two invariant solutions out of the eleven optimal systems were calculated and we believe that these solutions will one day find practical applications in line with the reported thinking in Ibragimov (1995, Vol. 2, p.29).

Chapter 4

A Stochastic Volatility Model for Prices of Electricity Future Contracts

Theory without practice is pointless. Practice without theory is mindless.

4.1 Introduction

It is known that new events influence stock market prices either positively or negatively. For example, political disturbances (among others), especially in oil producing countries, affect commodity markets such as crude oil prices substantially. The electricity market as an incomplete market has peculiar characteristics: nonstorability and inventories cannot be held. Overtime-forced outages of generation plants or unexpected contingencies in transmission networks often result in short time fluctuations in prices. This first of the two Chapters that constitute Part Two of this Thesis is devoted to developing a model to estimate the parameters of stochastic volatility of the electricity market with reference to the Pennsylvania daily prices of Electricity Futures Contract. We propose a model of dynamic linear type incorporating switching regimes as this is a particular class of state-space models that allow many of the relevant inferences to be performed exactly using the Kalman filter. Although the Kalman filter was designed initially for tracking problems, it has recently been very successful in estimating parameters in a

wide range of applications including those in Mathematics of Finance. In the following Section we give without proof some theorems and definitions necessary for our theoretical framework followed by Section 4.3 in which the pricing and some stochastic volatility models are discussed. We introduce the concept of Kalman filter state-space or dynamic linear modeling and we show the equivalence of the Kalman filter state-space and the popular Heston/CIR dynamic models in Section 4.4. In Section 4.5 we discuss modified Kalman filter algorithms and filtering equations. Implementation of the algorithm is discussed in Section 4.6 and empirical results are presented in Section 4.7. Thereafter we conclude.

4.2 Some Elements of Stochastic Calculus

As in Subsection 1.4.1 of Chapter One, we assume that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the sample space, $\mathcal{F}$ the σ-algebra-generated process $\omega = (w_t : t \in \mathbb{R})$ and $\mathbb{P}$ the probability measure, $\mathbb{P}$: $\mathcal{F} \mapsto [0, 1]$. We use the fundamental Brownian motion w_t on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to represent our important stochastic engine for modelling the randomness in the financial market. Stochastic calculus is the calculus that has been developed to work with the stochastic process. We limit our discussion to stochastic processes known as Itô processes. Reference can be made to Øksendal (1995) for a more extensive and rigorous treatment of stochastic differential equations.

Definition 4.2.1 (Brownian Motion)**.** [1] The stochastic process $\omega = \{w_i : t \in \mathbb{R}\}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a Wiener or Brownian motion process if the following properties hold almost surely.

(i) $P(W_0 = 0)$=1;

[1] In Financial Mathematics Brownian motion (also called the Wiener process), is a particular type of stochastic process that is often incorporated into models of financial asset to model uncertainty in the market (see for example, Neftci (1996) and or Hull (1989) for a good introduction). The (Financial Mathematics) Brownian motion process is actually a mathematical model of a phenomenon first reported by the Scottish botanist Robert Brown in 1827. He observed under the lens of the microscope that pollen grains suspended in water behave in a random manner.

(ii) For each $n \geq 1$ and any $0 \leq t_0 < t_1 < \ldots < t_n$ the random variables $W_1 - W_0, W_2 - W_1, \ldots, W_n - W_{n-1}$ are independent;

(iii) For any t and $h > 0$(i.e. $t > h \geq 0$), the random variable $W_{t+h} - W_t$ is normally distributed with:

(a) $\mathbb{E}[W_{t+h} - W_t] = \mu h, \quad \mu$ a real positive constant

(b) $\mathbb{E}[W_{t+h} - W_t]^2 = \sigma^2 h, \quad \sigma^2$ a positive constant
where μ and σ^2 are the drift and variance parameters respectively; and

(iv) W_t is continuous in $t \geq 0$.

The process with $\mu = 0$ and $\sigma = 0$ is called normalized or standard Brownian motion process.

Lemma 4.2.2. *If W_t is Brownian motion process, then $Cov[W_{t+h}, W_t] = min(t + h, t)$.*

Proof. (see Sobczyk 1995, p63). □

Remark 4.2.1 This lemma is a very elementary application of independent increments and the mean-zero properties of the Brownian motion process. It also demonstrates how application of independent increments can rely on the zero-mean property.

Definition 4.2.1 [1-dimensional Itô Process] Let $(W_t)_{0 \leq t \leq T}$ be a Brownian motion. The Itô process (stochastic process) X_t on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is then given by

$$X_t = X_0 + \int_0^t \mu(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s \tag{4.2.1}$$

often written in a shorter form

$$dX_t = \mu(X_s, t)dt + \sigma(X_t, t)dWs \tag{4.2.2}$$

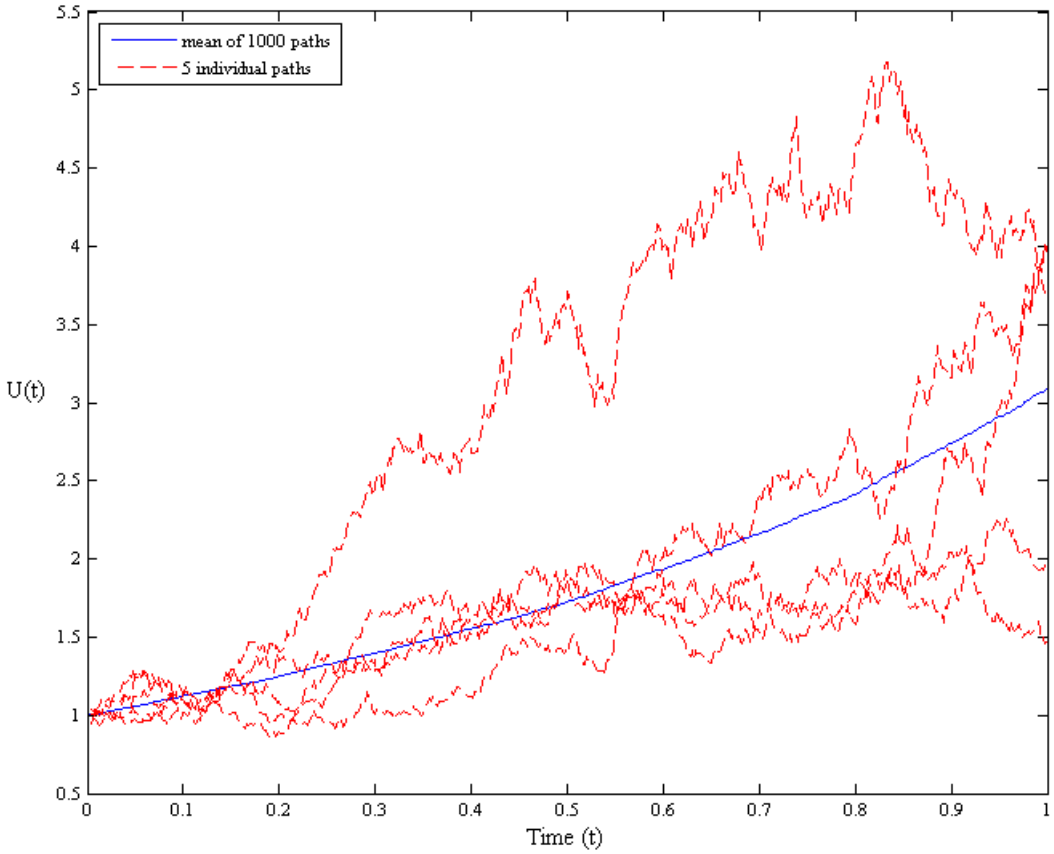


Figure 4.2.1: Simulation of sample Brownian path

such that the following conditions hold almost surely:

$$P\left[\int_0^t |\sigma(X_s, s)ds| < \infty, \forall t \geq 0\right] = 1$$

and

$$P\left[\int_0^t |\mu(X_s, s)ds| < \infty, \forall t \geq 0\right] = 1.$$

Equation (4.2.1) is generally known in financial economics as a stochastic differential equation (SDE) because with X_t already a stochastic process it is a differential equation with a noise term added. SDE (4.2.1) consists of two terms; the first term μdt defined as the drift term and the second term σdW_t which specifies the random part (the noise) of the process sometimes called the diffusion part. For the existence and uniqueness of the SDE given in (4.2.1) we need the following existence and uniqueness condition on μ and σ to be fulfilled.

Theorem 4.2.3 (Existence and uniqueness)**.** *The condition that guarantees the existence and uniqueness of the solution of SDE (4.2.1) is the growth condition, i.e., μ and σ satisfy*

$$|\mu(x,t)| + |\sigma(x,t)| \leq C(1+|x|), \quad x \in \mathbb{R}, t \in [0,T]$$

for some constant C, which guarantees global existence; and the Lipshitz condition

$$|\sigma(x,t) - \sigma(y,t)| + |\mu(x,t) - \mu(y,t)| \leq D|x-y|, \quad x,y \in \mathbb{R}, t \in [0,T]$$

for some constant D, which guarantees local uniqueness; where $\mathcal{F}_t$ is the filtration generated by $W = \{W_t : t \in \mathbb{R}\}$.

Proof. See Øksendal (2000).

□

The Itô formula is one fundamental result that enables the use of and solution of SDEs. It is the stochastic analogue to the chain rule in ordinary mathematical analysis.[2] It transforms the Brownian motion given a function $Y_t = f(X_t, t)$, where X_t is defined in (4.2.1), the dynamics of Y_t is then given by applying the second-order Taylor expansion.

Theorem 4.2.4 (Itô formula)**.** *Let X_t be a stochastic process given by the SDE (4.2.1) and let $u(x,t) \in C^{1,2}([0,\infty] \times \mathbb{R})$[3]. Then $Y_t = u(X_t, t)$ is an Itô process and*

$$dY_t = \frac{\partial u}{\partial t}((X_t,t))dt + \frac{\partial u}{\partial t}((X_t,t))dX_t + \frac{1}{2}\frac{\partial^2 u}{\partial^2 t}((X_t,t))(dX_t)^2$$

in which the following multiplication rules have been used

[2] Itô's Lemma (Itô, 1951) is to stochastic calculus what the Taylor expansion is to ordinary calculus. It is largely used to construct differential equations for a function (e.g., options) of stochastic variable(s) like a commodity price.

[3] i.e. u is twice continuously differentiable on $[0,\infty)$

$\cdot$	dt	dW_t
dt	0	0
dW_t	0	dt

Proof. Detailed proof in either Liptser and Shiryayev (1977) or Etheridge (2002).

□

Remark 4.2.2 The class of Itô processes does not include all processes for which Itô's formula works, but it is sufficiently broad to include the majority of applications in Financial Mathematics. The Brownian motion has two major drawbacks that do not allow it function properly in financial market setting. These drawbacks include (a) asset prices, S_t, are always positive ($S_t \in (0, \infty)$) and, since the price of an asset is a normal random variable, it can theoretically become negative ($X_t \in (-\infty, \infty)$), and (b) fluctuations in the price are proportional to the price of the asset. Instead we introduce a nonnegative functional of Brownian motion called *geometric Brownian motion*[4] defined as follows:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{4.2.3}$$

which is a short form of the following equation

$$S_t = S_0 + \int_0^t \mu S_z dz + \int_0^t \sigma S_z dW_z \tag{4.2.4}$$

We assume for now that the daily asset returns follow a log normal distribution and we denote this by $r_t = \ln\left(\frac{S_t}{S_0}\right)$. By applying Itôs formula we get the following expression

$$dr_t = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t$$

finding the primitive function

$$r_t = \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t$$

and finally ending up with, see for example, Hull (1997) and Shreve (2004),

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right\}, \tag{4.2.5}$$

[4]The underlying stock is assumed to follow geometric Brownian motion in the famous Black-Scholes model.

where dt is the change in time for which S_t changes to $S_t + dS_t$, dS_t/S_t is the return to investment in the asset and dW_t is a standard Wiener process (see Osborn, 1959 for discussion), $dW_t = \epsilon_t\sqrt{dt}$ and $\epsilon_t \sim N(0, dt)$. Here the drift parameter, μ is a measure of growth of the asset (i.e. the expected return per unit time) and σ is the measure of volatility of the asset as estimated by the standard deviation of the returns. Equation (4.2.5) is the solution of the stochastic differential equation (4.2.3). It then follows that the geometric Brownian motion (GBM) with initial value S_0 has the following log-normal distribution

$$p\left(S_t, t; S_0, t_0\right) = \frac{1}{\sqrt{2\sigma^2\tau}S_t}\exp\left\{-\frac{\left[\ln\left(\frac{S_t}{S_0}\right) - \left(\mu - \frac{1}{2}\sigma^2\right)\tau\right]^2}{2\sigma^2\tau}\right\}, \tag{4.2.6}$$

where $\tau = t - t_0$. Equation (4.2.6) is the probability density function (pdf) of the GBM, distributed as

$$R_t \sim N\left(\ln S_0 + \left(\mu - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right)$$

The expected value of the process (4.2.5) is given by

$$E[S_t] = S_0 \exp(\mu t) \tag{4.2.7}$$

Interested readers may consult Samuelson (1965), Tuckwell (1988), Ross (2000) and Marathe and Ryan (2005) for details.

4.3 Heston's Stochastic Volatility Model Revisited

In the literature a different way to improve the traditional financial models based upon Brownian motion is represented by stochastic volatility models. The main future of stochastic volatility processes is, clearly, the fact that their unconditional volatility changes stochastically over time. A number of different stochastic volatility models have been developed since 1987.[5] A stochastic volatility model can be seen as a discrete-time approximation to the Hull and White derivative pricing model (Hull and White, 1987) in which the stock price dynamics are governed

[5]The first application of stochastic volatility for derivative pricing is due to Hull and White (1987).

by unobservable state variables. In their model stock price follow a diffusion process

$$\frac{dS_t}{S_t} = \alpha dt + \sigma(t) dW_1 \tag{4.3.1}$$

and the logarithm of $\sigma(t)$ follows a diffusion process given by the Ornstein-Uhlenbeck (O-U) process:

$$d(\log \sigma) = \lambda \left(\xi - \ln \sigma\right) dt + \gamma dW_2, \tag{4.3.2}$$

where S_t is the stock price, $\sigma(t)$ the instantaneous variance of S_1 and W_1, W_2 are two Wiener processes. This model therefore generalizes the classical Black-Scholes option-pricing formula of Black and Scholes (1973) to allow for stochastic volatility. The general form of the diffusion process for the short rate, r_t , with real-world drift, μ_r , and volatility, σ_r , is governed by the stochastic differential equation of the form

$$dr_t = \mu_r(r_t, t)dt + \sigma_r(r_t, t)dW_t, \tag{4.3.3}$$

where the functional forms μ_r and σ_r determine the behaviour of the short rate and dW_t is a standard Brownian motion. Let us denote by ν_t the value at time t of an interest rate contingent claim with maturity T. As it derives from the single factor model assumption, only the short rate and the time to maturity $T - t$ will affect the price of the claim so that we can write

$$S_t(r_t) \equiv \nu(r_t, t) \equiv \nu(r_t, t, T) \tag{4.3.4}$$

This consists of the determination of the fair value $\nu(r_t, t)$ of the bond for $t < T$. With the application of Itô's lemma and using (4.3.3) and (4.3.4) yields the claim dynamics

$$d\nu_t = \left(\frac{\partial \nu_t}{\partial t} + \mu_r \frac{\partial \nu_t}{\partial r_t} + \frac{1}{2}\sigma_r^2 \frac{\partial^2 \nu_t}{\partial r_t^2}\right) dt + \left(\sigma_r \frac{\partial \nu_t}{\partial r_t}\right) dW_t \tag{4.3.5}$$

Dividing both sides by ν_t yields the instantaneous return on the contingent claim by applying the definition of the market price at risk, λ_t(which may be interpreted as the extra profit on the portfolio per unit risk (Kwot 1998), we obtain

$$\frac{\partial \nu_t}{\nu_t} = (r_t + \lambda_t \nu(t, T))dt + \nu(t, T)dW_t \tag{4.3.6}$$

with

$$\nu(t, T) = \frac{\sigma_r(r_t, t)}{\nu_t} \frac{\partial \nu_t}{\partial r_t} \tag{4.3.7}$$

Equating the drift in (4.3.6) with the drift in (4.3.5) we obtain a second-order partial differential equation (called the Feynman-Kac equation) that must be satisfied by any interest rate contingent claim in no-arbitrage one factor model

$$\frac{\partial \nu_t}{\partial t} + (\mu_r - \lambda_t \sigma_r)\frac{\partial \nu_t}{\partial r_t} + \frac{1}{2}\sigma_r^2 \frac{\partial^2 \nu_t}{\partial r_t^2} - r_t \nu_t = 0 \tag{4.3.8}$$

The term $\nu_r - \lambda_t \sigma_r$ in (4.38) is often called the risk adjusted drift. As we see hereafter, equation (4.3.8) will be the fundamental equation on which any interest-rate contingent claim price can be computed as the solution of such a partial differential equation subject to an appropriate boundary condition. With different configurations of μ_r and σ_r as inputs, different interest rate contingent claims will produce the same partial differential equation but with different boundary conditions.

Under the risk-neutral measure Q, the term structure of interest rates in the ordinary Vasicek (1977) model evolves according to the stochastic differential equation

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t^Q \tag{4.3.9}$$

where $\theta = \mu - \dfrac{\lambda\sigma}{\kappa}$ is the risk-neutral mean, κ and σ are positive constants. This defines a random walk around a trend with a mean reverting characteristic. In this formulation the diffusion process allows for the possibility of negative interest rates. To rectify this problem of negative interest rates in the Vasicek model, Cox, Ingersoll and Ross (CIR) in 1985 introduced a modification of (4.3.9) known as the square-root process

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t^Q, \quad \theta, \kappa, \sigma > 0. \tag{4.3.10}$$

This model corresponds to a continuous time first-order autoregressive process where the randomly moving interest rate reverts elastically to its long-term value, θ. This implies that interest rates (commodity prices) are determined by the supply and demand of individuals. Equations (4.3.9) and (4.3.10) are special cases of the general mean-reverting process

$$dr_t = \kappa(\theta - \beta r_t)dt + \sigma r_t^{\delta} dW_t^Q \tag{4.3.11}$$

where $0 \leq \delta \leq 1$ constant. Set $\delta = 0$ in (4.3.11) and we obtain the generalized Vasicek model while setting $\delta = 0.5$ yields the generalized CIR model.

Hull and White (1990) propose a generalization of Vasicek and CIR models to be arbitrage-free such that r_t is governed by

$$dr_t = \kappa(\theta_t - \beta_t r_t)dt + \sigma_t r_t^{\delta} dW_t^{Q}, \tag{4.3.12}$$

for some constant $\kappa \geq 0$.

Consider the basic stochastic volatility model proposed by Heston (1993) that assumes a system of SDEs under the objective probability measure $\mathbb{P}$,

$$dr_t = \mu S_t dt + S_t \sqrt{\nu_t} dW_t^{S} \tag{4.3.13}$$

and

$$dr_t = \kappa[\theta - r_t]dt + \xi \sqrt{r_t} dW_t^{\nu} \tag{4.3.14}$$

where ν_t the instantaneous stochastic variance of the equity spot price S_t is represented by a CIR process defined by (4.3.14). The parameters in these equations are defined as follows:

μ is the deterministic drift (a.k.a the deterministic rate of return) of the asset, S_t

$$i.e., \quad \mu = \frac{S_t - S_{t-1}}{S_{T-1}},$$

θ is the long run average value of the stochastic variance of (a.k.a long vol);

$$i.e. \quad as \quad t \longrightarrow \infty, \quad \mathbb{E}(\nu_t) \longrightarrow \theta,$$

κ is the rate at which returns to θ, (a.k.a. the mean reversion rate of the volatility);

$$i.e., \quad \kappa \approx \mathbb{E}\left(\frac{\theta - \nu_t}{\theta}\right) = 1 - \frac{1}{\theta}\mathbb{E}(\nu_t)$$

is the volatility of volatility[6] (a.k.a vol of vol) which, as the name suggests, determines the variance of

$$i.e., \xi \approx \sqrt{Var(\nu_t)},$$

ρ is the correlation between the two Weiner processes dW_t^S and dW_t^{ν} ;

$$i.e., \quad \rho = Corr(dW_S^Q, dW_S^{\nu}), \rho \in [-1, 1].$$

[6]If the parameters obey the Feller condition (Albrecher *et al* 2007), then the process, ν_t , is strictly positive, i.e., $2\kappa\theta \geq \sigma^2$.

The stochastic behaviour of S_t assumed in (4.3.13) shows that the stock price follows a geometric Browning motion with stochastic specification of the volatility term inspired (Kellerhals, 2004) by the modification of the standard Heston model for pricing electricity forwards. The modification of (4.3.13) and (4.3.14) are based on the idea of Ross (1997) and using the Girsanov transformations $dW_t^{S,Q} = dW_t^{S,Q} + \lambda^* \sqrt{\nu_t} dt$ under the martingale measure, $\mathbb{Q}$, and $X_t^* = \ln S_t$, where $\lambda^* = \lambda - 0.5$ and X_t^* constitute the transformed spot price. These transformations above results in the following system of SDEs

$$dX_t^* = [\mu - \lambda\nu_t]dt + \sqrt{\nu_t} dW_t^S \tag{4.3.15}$$

and

$$d\nu_t = [\kappa(\theta - \nu_t) - \lambda_\nu \nu_t]dt + \xi\sqrt{\nu_t} dW_t^\nu \tag{4.3.16}$$

where $\lambda_\nu \nu_t$ is the market price of risk and λ_ν a constant. Motivated by SDEs (4.3.15) and (4.3.16), Kellerhals suggests a corresponding partial differential equation for the pricing of electricity forward contract under the equivalent martingale measure $\mathbb{Q}$ of the form

$$\frac{1}{2}\sigma^2 \nu_t \frac{\partial^2 y(t,\nu_t)}{\partial \nu_t^2} + (\kappa\theta - (\kappa + \lambda_\nu)\nu_t)\frac{\partial y(t,\nu_t)}{\partial \nu_t} + \frac{\partial^2 y(t,\nu_t)}{\partial t} = k_1 \nu_t y(t,\nu_t) \tag{4.3.17}$$

with boundary condition for the value at maturity Kellerhals (2004, p192)

$$y(T,\nu_t) = \exp(k_1 \nu_T) \tag{4.3.18}$$

where $k_1 = \lambda - \dfrac{\kappa + \lambda_\nu}{\sigma} - \dfrac{1}{2}(1 - \rho^2)$.

The model (4.3.17) is at the centre of this Thesis. We have already performed a symmetry analysis of a particular case of the model in Chapter Three and in the following Sections we outline methods of parameter estimation and their implementation.

Remark 4.3.1 Heston (1993)chooses the market price of volatility risk to be proportional to volatility, i.e., $\Lambda(S,\nu,t) = \kappa\sqrt{\nu}$ or $\Lambda(S,\nu,t)\xi\sqrt{\nu} = \kappa\xi\nu$. Let $\lambda = \kappa\xi$, so that the coefficient of $\partial u/\partial\nu$ in (4.3.17) becomes $[\kappa(\theta - \nu) - \lambda\nu]$. There are analytical advantages in this choice of market price of volatility risk. The drift term of the specified process (4.3.17) is an affine function of the state variable itself and its affinity makes the model easier to solve. For the

different exponential affine structure models proposed especially for the Cox-Ingersoll-Ross (CIR) model, see for example Cox *et al* (1985) and Heston (1993) while their implementations using Kalman filtering[7] , see for instance Chen and Scott (2003), Chetterjee (2003), Do (2008), Geyer and Pichler (1999) and Babs and Nowman (1999).

4.4 The Kalman State-Space Model or Dynamic Linear Model(DLM)

4.4.1 The Basics

Kalman (1960)and Kalman and Bucy (1961) introduced a very general model that incorporates many special cases of interest in time series data analysis, called the State-Space or Dynamic Linear Model (DLM) defined in its basic form as;

$$X_t = \alpha + \Phi X_{t-1} + \omega_t \tag{4.4.1}$$

and

$$Y_t = A_t X_t + \nu_t, \tag{4.4.2}$$

where equation (4.4.1) is the state equation which through a $p \times p$ transition matrix Φ determines the rule for the generation of X_{ti} from the past states $X_{(t-1),j}$, $j = 1, \ldots, p$ for $i = 1, \ldots, p$ and the time points $t = 1, \ldots, n$ of the original vector of interest called the state vector X_t, assumed to be directly unobservable. The ω_t are $p \times 1$ identically and independently distributed (i.i.d.) zero mean Gaussian vectors with covariance matrix Q. Also, α is a $p \times 1$ vector of constants such that, if $\mathbb{E}(X_t) = \mu$, then $\alpha = (I - \Phi)\mu$. Equation (4.4.2) is the observable equation which through a $q \times p$ measurement or observation matrix A_t determines Y_t as the directly observable linearly transformed version of X_t with added noise ν_t assumed to be a Gaussian white noise with a $q \times q$ covariance matrix, R. The noise processes $\{\omega_t\}$ and $\{\nu_t\}$ may or may

[7]The Kalman filter is an iterative procedure (as soon becomes clearer below) that forecasts the state variable one period into the future by a linear projection and then updates this forecast when the observation on the variable Y_t becomes available.

not be correlated as

$$Cov(w_t, \nu_t) = \begin{cases} \mathbb{E}(w_t \nu_t') = S & \text{at time } t \\ 0 & \text{otherwise.} \end{cases}$$

State-space models are based on the idea that the time series Y_t is an incomplete and noisy function of some underlying unobservable process $X_t, t = 1, ..., n$, called the state process. More generally, we might think of X_t as an auxiliary random process which facilitates the task of specifying the probability law of the time series: the observable process Y_t depends on the latent state process X_t, which has a simpler, Markovian dynamics, and we can reasonably assume that the observation Y_t only depends on the state of the system at the time the measurement is taken, X_t. Figure 4.4.1 represents the sketch of the dependences among variables that we are assuming.

$$\begin{array}{ccccccccccc} X_0 & \longrightarrow & X_1 & \longrightarrow & X_2 & \longrightarrow \cdots \longrightarrow & X_{t-1} & \longrightarrow & X_t & \longrightarrow & X_{t+1} \longrightarrow \cdots \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ & & Y_1 & & Y_2 & & Y_{t-1} & & Y_t & & Y_{t+1} \cdots \end{array}$$

Figure 4.4.1; State space model dependence structure

Formally, the nature of the assumptions of a state space model may be stated as:

(a) $X_t, t = 1, ..., n$ is a Markov chain; that is, X_t depends on the past values $X_0, X_1, ..., X_{t-1}$ only through X_t. Thus, the probability law of the process $X_t, t = 1, ..., n$ is specified by assigning the initial density $p_0(X_0)$ of X_0 and the transition densities $p(X_t|X_{t-1})$ of X_t conditionally on X_{t-1}.

(b) Conditionally on $X_t, t = 1, ..., n$, the Y_t are independent and Y_t depends on X_t only. It follows that, for any $n \geq 1, (Y_1, ..., Y_n)|X_1, ..., X_n$ have joint conditional density $\prod_{t=1}^{n} f(y_t|X_t)$.

Just as there are variations in stochastic volatility models in the literature, there are also several modifications of the general Kalman filtering method that have been developed for dealing with stochastic volatility models. We propose in the next Subsection some modifications to enhance parameter estimation of our model through the maximum likelihood estimation.

4.4.2 Gaussian-mixture: A Primer

A special case of a regime switching model is the independent and identically distributed mixture of two normal distributions in two states. Let the state (regime) that an unobservable process is in at time t be denoted as X_t, where there are $N = 2$, say, possible regimes ($X_t = 0, 1$). When the unobserved process is at state j, i.e., $X_t = j$, the observed sample y_t is presumed to have been drawn from a $N(\mu_j, \sigma_j^2)$ distribution. Hence the density of y_t conditional on the state variable X_t taking on the value j is given by

$$f(y_t|X_t = j; \theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}\right\}, j = 0, 1 \tag{4.4.3}$$

where θ is a vector of population parameters such that $\theta \equiv (\mu_0, \mu_1, \sigma_0^2, \sigma_1^2)'$. The unobservable regime $\{X_t\}$ is presumed to have been generated by some probability distribution for which the unconditional probability that $\{X_t\}$ takes on the value j is π_j:

$$\pi_j = P\{X_t = j; \theta\}, \quad \text{for} \quad j = 0, 1. \tag{4.4.4}$$

The probabilities π_o and π_1 are also included in θ so that θ is now given by the parameter vector $\theta \equiv (\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, \pi_0, \pi_1)'$. Combining (4.4.3) and (4.4.4) we have the joint probability density function (pdf) of y_t and X_t given by

$$P(y_t, X_t = j; \theta) = f(y_t|X_t = j; \theta) = P\{X_t = j; \theta\} \tag{4.4.5}$$

which is

$$P(y_t, X_t = j; \theta) = \frac{\pi_j}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}\right\} \tag{4.4.6}$$

We find the unconditional density of y_t by summing over all possible values that state variable can take:

$$\begin{aligned} f(y_t; \theta) &= \sum_{j=1}^{2} P(y_t, X_t = j; \theta) \\ &= \frac{\pi_1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(y_t - \mu_1)^2}{2\sigma_1^2}\right\} + \frac{\pi_2}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{(y_t - \mu_2)^2}{2\sigma_2^2}\right\} \end{aligned} \tag{4.4.7}$$

Since the regime X_t is unobserved, the expression by equation (4.4.7) is the relevant density describing the data, y_t , actually observed. If the state variable X_t is independent and identically

distributed across different time points, t, then the log likelihood for the observed data can be calculated using (4.4.7) as

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} log f(y_t; \theta) \tag{4.4.8}$$

where T is the time horizon. The maximum likelihood estimate of θ is obtained by maximizing (4.4.8) subject to the restrictions $\pi_0 + \pi_1 = 1$ and $\pi_j \geq 0$ for $j = 0, 1$. The probability density function of the form of equation (4.4.7) can be used to represent a broad class of different densities. Figure 4.4.2(top) gives an example of a Gaussian mixture for $N = 2$. However, a mixture of two Gaussian variables need not have the bimodal appearance as in Figure 4.4.2(top), but can produce a unimodal density allowing skew or kurtosis different from that of a single Gaussian variable as in Figure 4.4.2(bottom).

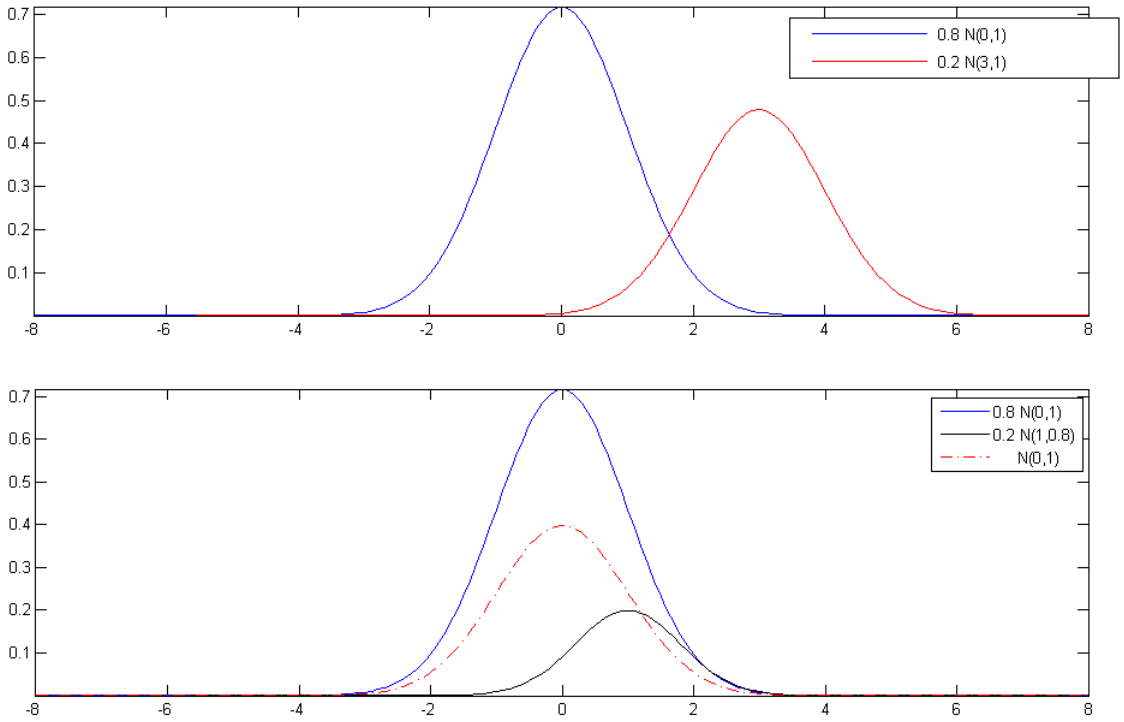


Figure 4.4.1: Gaussian mixture of two normal distributions with varying means and variances: bimodal (top) and unimodal bottom

This basic idea given in this Subsection serves as a primer to Section 4.5, where we incorporate the expectation maximization (EM) algorithm of Dempster *et al* (1977) which is basically a two-phase iterative algorithm whereby inference about an unobserved regime is computed

(expectation) and the maximum likelihood estimates of the parameters are calculated (maximization) for the Gaussian mixtures. Hamilton (1994) illustrates which regime is more likely to have been responsible for producing the observation y_t as.

$$P\{X_t = j|y_t;\theta\} = \frac{P(X_t = j;\theta)}{f(y_i;\theta)} = \frac{\pi_j f(y_t|X_t = j;\theta)}{f(y_i;\theta)} \tag{4.4.9}$$

Hamilton gives the maximum likelihood estimates of the parameter set θ as

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T} y_t P\{X_t = j|y_t;\hat{\theta}\}}{\sum_{t=1}^{T} P\{X_t = j|y_t;\hat{\theta}\}} \tag{4.4.10}$$

and

$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^{T} y_t P\{X_t = j|(y_t - \mu_j)^2;\hat{\theta}\}}{\sum_{t=1}^{T} P\{X_t = j|y_t;\hat{\theta}\}}, \tag{4.4.11}$$

$$\hat{\pi}_j = \frac{1}{T}\sum_{t=1}^{T} P\{X_t = j|y_t;\hat{\theta}\}. \tag{4.4.12}$$

In the discrete time log-normal stochastic volatility models the approach advocated by Harvey *et al* (1994) has been influential. Their approach was to remove the predictable part of the returns. So we think of $Y = M$ again and work with $\log R_t^2 = \nu_t + \log \varepsilon_t^2$. If the volatility has short memory then this form of the model can be handled using the Kalman filter while long memory models are often dealt with in the frequency domain. Either way this delivers a Gaussian quasi-likelihood which can be used to estimate the parameters of the model. The linearised model is non-Gaussian[8] due to the long left hand tail[9] of $\log \varepsilon_t^2$ which generates outliers when ε_t is small.

4.4.3 A Gaussian-mixture/AR Model

Recall that in(4.4.2), the measurement matrix A_t converts the unobservable measurement into data vectors Y_t so that (4.4.2) can be assumed to possess a Markov behavior with switching regimes (Kim, 1994, So *et al* 1998). We assume that the log-volatilities follow an autoregressive

[8]By taking logarithms of squared returns, we obtain a linear albeit non-Gaussian state space model. Because $\log y_t^2$ is not truly Gaussian, the Kalman filter yields minimum mean square linear estimators (MMSLE) of X_t of future observations rather than minimum mean square estimators (MMSE).

[9]Implications of tails are discussed in Chapter Five.

(AR(1)) process with a Markov switching mean so that the underlying process is observed via conditionally independent and normally distributed daily returns, say.

Suppose, instead of the observation stochastic process in the Heston/CIR model, we let$\{S_t\}$ be a price process and $r_t = \triangle \ln S_t$ be defined as the log-return of a commodity at time t. Under the assumption of efficient markets, the log-returns have null conditional mean: $E(r_{t+1}|r_1, ..., r_t) = 0$. Chan *et al* (1992) show that many of the specific stochastic differential equations used in the literature can be written as a special case of (4.3.12)

$$dr_t = (\theta_t - \beta_t r_t)dt + \sigma_t r_t^{\delta} dW_t^Q. \tag{4.4.13}$$

If we let $\theta_t = \theta$ and $\beta_t = \beta$ and allow the volatility parameter to be time-varying in line with Ball and Torous (1999), then a simple discretization of (4.4.13) leads to

$$\Delta r_t = \theta + \beta r_{t-1} + \sigma_t r_{t-1}^{\delta} \varepsilon_t \tag{4.4.14}$$

where $\Delta r_t = r_t - r_{t-1}$ and ε_t is a standard normal variable. This model allows log-volatility to evolve stochastically as a simple AR(1) process

$$\log \sigma_t^2 = \phi_0 + \phi_1 \log \sigma_{t-1}^2 + \omega_t \tag{4.4.15}$$

where the disturbance term $\omega(\omega_t \sim iid(0, \sigma_\omega^2))$ which makes process (4.4.15) stochastic in the variance is itself subject to random shocks. By using the residual in (4.4.15) to write the system in a state-space form and then applying the Kalman filter recursively, one builds up the log-likelihood function. The transformation is employed on the residual (Harvey, *et al* 1994, Mahieu and Schotman, 1998 and So, *et al* 1998 defined their residual variously in the neighbourhood of $r_t = \vartheta \ \exp\{\nu_t/2\}\varepsilon_t$ through ARCH and GARCH models):

$$R_t = \sigma_t r_{t-1}^{\delta} \varepsilon_t = \Delta r_t - \theta - \beta r_{t-1} \tag{4.4.16}$$

If we take the log of the square of the residual in (4.4.16) we obtain

$$\log R_t^2 = \log \sigma_t^2 + 2\delta \log r_{t-1} + \log \varepsilon_t^2 \tag{4.4.17}$$

If we let $Y_t = \log R_t^2$ which is observable given the observed returns, $\nu_t = \log \sigma_t^2$ is a state variable (i.e., log-volatility) and $\gamma_t = \log \varepsilon_t^2$ in (4.4.17) then we have the system rewritten into

a state-space form as

$$Y_t = \nu_t + 2\delta \log r_{t-1} + \gamma_t \tag{4.4.18}$$

and

$$\nu_t = \phi_0 + \phi_1 \nu_{t-1} + \omega_t \tag{4.4.19}$$

NB: Equation (4.4.18) is a general set-up that applies to both diffusion and non-diffusion models; since our approach is non-diffusion, we set $\delta = 0$ in (4.4.18) to obtain

$$Y_t = \alpha + \nu_t + \gamma_t \tag{4.4.20}$$

where

$Y_t = \log R_t^2$ is the log of the squared standardized returns R_t of the asset S_t at time t,

$\alpha =$ deterministic component,

$\nu_t = \log(R_t^2/\varepsilon_t^2)$ represents an autoregressive model of order one (AR(1)) in (4.4.17),

$\gamma = U_t Z_{t0} + (1 - U_t) Z_{t1}$

U_t is the Markov regime state variable which is an i.i.d. Bernoulli process such that $Pr\{U_t = 0\} = \pi_0, Pr\{U_t = 1\} = \pi_1$ with $(\pi_0 + \pi_1 = 1)$, and

Z_{tj} are two i.i.d. Gaussian processes such that $Z_{tj} \sim N(\mu_j, \sigma_j^2)$, $j = 0, 1$ with $\mu_0 = 0$ so that the mixture γ_t of the two Gaussian distributions forms a white noise. The idea of writing the observation equation in the form (4.4.20) above is to make room for simplicity and a more general approach in the sense of allowing the dynamics of the observation error, γ_t, to depend upon parameters that are to be fitted. Suppose further that, instead of the CIR process representation for volatility, ν_t follows a first-order autoregressive (AR (1)) process defined by

$$\nu_t = \phi_0 + \phi_1 \nu_{t-1} + \omega_t \tag{4.4.21}$$

or equivalently

$$\nu_t = \phi_0(1 - \phi_1) + \phi_1 \nu_{t-1} + \sigma\omega_t, \tag{4.4.22}$$

where ϕ_0, ϕ_1, σ are parameters and the white Gaussian noise $\omega_t \sim N(0, \sigma^2)$ in (4.4.21) or $\omega_t \sim N(0, 1)$ in (4.4.22). This assumption is necessary in constructing regime-switching regression models or Kalman filtering algorithms with Markov switching-regimes, see for example, Goldfeld and Quandt (1973)[10], Hamilton (1989, 2005), and Shumway and Stoffer (1991, 2009).

[10]Markov-switching regressions were introduced in econometrics in 1973 by Stephen M. Goldfeld and Richard

4.4.4 Analogy between the Gaussian-mixture/AR model and Heston/CIR model

Let the vectors of parameters for the Heston/CIR model and the Gaussian-mixture/AR model be denoted respectively by

$\psi = (\mu, \kappa, \theta, \xi, \sigma_s^2, \sigma_\nu^2)$ Basic Heston/CIR model

and

$\Theta = (\alpha, \phi_0, \phi_1, \sigma, \mu_0^2, \mu_1, \sigma_1^2, \pi_1)$ Gaussian-mixture/AR model.

The constant α in the Gaussian-mixture /AR model represents the deterministic component of Y_t just as μ represents the deterministic drift (a.k.a the deterministic rate of return) of the price of asset, S_t, in the Heston/CIR model.

$\Rightarrow \mu \equiv \alpha$.

θ is the long run average value of the variance of ν_t, i.e., as $t \to \infty, \mathbb{E}(\nu_t) \to \theta$.

If we let θ be the mean of ν_t as an AR (1) expressed in the Gaussian-mixture/AR model, then

$$\Rightarrow \theta \equiv \begin{cases} (1-\phi_1)\phi_0 + \phi_1 \nu_{t-1} & \text{conditional on } \nu_{t-1}, \\ \phi_0 & \text{unconditionally (in the long run).} \end{cases}$$

κ is the rate at which reverts to θ (a.k.a. volatility's mean reversion rate);

$$\Rightarrow \kappa \equiv \mathbb{E}(\frac{\theta - \nu_t}{\theta}) = 1 - \frac{1}{\phi}\mathbb{E}(\nu_t)$$

$$= \begin{cases} \frac{\phi_1(\phi_0+1)}{\phi_0}\nu_{t-1} & \text{conditional on } \nu_{t-1}, \\ 0 & \text{unconditionally.} \end{cases}$$

ξ is the volatility of volatility which, as the name suggests, determines the variance of ν_t

$$\xi \equiv \sqrt{Var(\nu_t)} = \begin{cases} \sigma & \text{conditional on } \nu_{t-1}, \\ \frac{\sigma}{\sqrt{1-\phi_1}} & \text{unconditionally.} \end{cases}$$

ρ is the correlation between the two Weiner processes dW_t^s and dW_t^ν comparable to the correlation between the two Gaussian processes γ_t, the observation noise, and ω_t, the state noise, in which case

$$\rho \equiv Corr(\gamma_t, \omega_t) = \begin{cases} 0, & \text{when the state and observation noises are uncorrelated,} \\ g(\alpha, \mu_0, \sigma_o^2, \mu_1, \sigma_1^2, \pi_1), & \text{say, when they are correlated at time, } t, \end{cases}$$

E. Quandt.

where $g(.)$ stands for some function of the parameters in its argument.

4.5 Applying the Kalman State-Space methodology to Stochastic Volatility modeling

4.5.1 The Kalman filter

A Kalman filter is simply an *optimal recursive data-processing algorithm.* It combines all available measurement data plus prior knowledge about the system to produce an estimate of the desired variables in such a manner that the error is minimized statistically. One of the many ways the filter obtains optimal estimates of desired quantities from data provided by a noisy environment is the Bayesian viewpoint. In the Bayesian principle we want the filter to propagate the *conditional probability density* of the desired quantity, conditioned on knowledge of the actual data coming from the measuring devices. The key notion here is that given the data $\mathbf{Y}_t = (Y_t, Y_{t-1}, ..., Y_1)$, inference about the state of nature, X_t, can be performed through a direct application of Bayes' theorem:
$Pr\{\text{State of nature}|\text{Data}\} \propto Pr\{\text{Data}|\text{State of nature}\} \times Pr\{\text{State of nature}\}$

$$Pr(X_t|Y_t) \propto Pr(Y_t|X_t, Y_{t-1}) \times Pr(X_t|Y_{t-1}). \tag{4.5.1}$$

Expression (4.5.1) denotes the *posterior distribution* for X at time t, whereas the first and second expressions on the right hand side denote the *likelihood* and the *prior distribution* for X, respectively. Given the realization of the state variables at time t and $t-1(S_t = j$ and $S_{t-1} = i$ where $i, j = 0$ or $1)$ and using the notation X_t^{t-1} to denote the variable conditional on the realized states j and i, the Kalman filter can now be represented in what follows (see also, Hamilton, 1989, Kim, 1994).
Let $X_t^s = \mathbb{E}(X_t|Y_s)$, where $Y_s = \{y_1, ..., y_s\}$ represents the conditional expectation of a commodity price at time t given observations up to and including time s, so that
$P_{t_1,t_2}^s = \mathbb{E}\{(X_{t_1} - X_{t_1}^s)(X_{t_2} - X_{t_2}^s)'\} = P_t^s$ when $t_1 = t_2 = t$.

Let the initial conditions[11] be

$$X_0^0 = \mu \quad \text{and} \quad P_0^0 = \Sigma_0 \quad \text{for} \quad t = 1, \ldots, n, \tag{4.5.2}$$

while

$$X_n^n \quad \text{and} \quad P_n^n \quad \text{for} \quad t > n, \tag{4.5.3}$$

are initial conditions for the accomplishment of (4.5.6) and (4.5.7). The Kalman gain and smooth gain matrix are defined respectively as

$$K_t = P_t^{t-1} A_t' [A_t P_t^{t-1} A_t^{'} + R]^{-1} \tag{4.5.4}$$

$$J_{t-1} = P_t^{t-1} \Phi^{'} [P_t^{t-1}]^{-1}. \tag{4.5.5}$$

Property 4.5.1 (Shumway and Stoffer, 2009) *The Kalman Filter:*
For the state-space model specified in (4.4.1) and (4.4.2) with the initial conditions in (4.5.3) for t = 1,...,n
when $s < t$,

$$X_t^{t-1} = \Phi X_{t-1}^{t-1} \tag{4.5.6}$$

$$P_t^{t-1} = \Phi X_{t-1}^{t-1} \Phi' + Q \tag{4.5.7}$$

when $s = t$,

$$X_t^t = X_t^{t-1} + K_t (Y_t - A_t X_t^{t-1}) \tag{4.5.8}$$

$$P_t^t = [I - K_t A_t] P_t^{t-1} \tag{4.5.9}$$

where K_t *is the Kalman gain defined in (4.5.4) above.*

Remark 4.5.1 When $t > n$, prediction is accomplished using initial conditions X_n^n and P_n^n and (4.5.6) and (4.5.7).

Property 4.5.2 *The Kalman Smoother* (obtaining estimators X_t^n for X_t based on the entire

[11] The initialization of the covariance matrix $\sum_0$ can be arbitrary, as long as it is nonzero, as the filter eventually converges and "forget" initialization errors (Jazwinski, 1972, Anderson and Moore, 1979 and Kailath *et al* 2000).

data sample $Y_1, ..., Y_n, t \leq n$). *For the state-space model specified in (4.4.1) and (4.4.2) with the initial conditions in (4.5.3) obtained from Property 4.5.1, and for* $t = n, n-1, ..., 1$,

$$X_n^{t-1} = X_{t-1}^{t-1} + J_{t-1}(X_t^n - X_t^{t-1}) \tag{4.5.10}$$

$$P_n^{t-1} = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}' \tag{4.5.11}$$

where J_{t-1} *is the smoother gain matrix defined in (4.5.5) above.*

The recursion can be understood in two stages: prediction and correction, described visually in Figure 4.5.1 below.

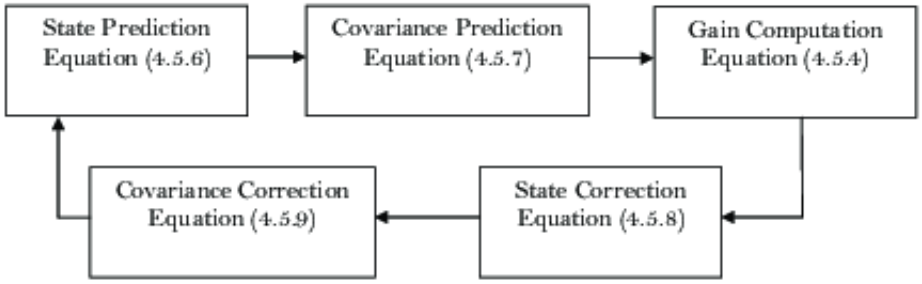


Figure 4.5.1: Prediction and correction stages of the Kalman-filter recursion

4.5.2 Modified Kalman filter: DLMs with switching

In modeling change in an evolving time series we assume discontinuous changes in the dynamics of some underlying model such as

(i) Changes occurring over time in error covariances,

(ii) Assigning mixture distributions to the observation error, ν_t, and

(iii) Allowing switches in the design matrix - in the classical regression case.

To incorporate a reasonable switching structure for the measurement matrix into the DLM that is compatible with practical situations, we assume that there are m possible state configurations in a nonstationary independent process defined by the time-varying probabilities

$$\pi_j(t) = \Pr(A_t = M_j), \qquad j = 1, ..., m \text{ and } t = 1, ..., n \tag{4.5.12}$$

such that important information about the current state of the measurement process is given by the filtered probabilities of being in state j, defined as the conditional probabilities

$$\pi_j\left(t\,|t\right) = \Pr\left(A_t = M_j\,|Y_t\right), \qquad j = 1,\ldots,m \text{ and } t = 1,\ldots,n \tag{4.5.13}$$

which also vary as a function of time. This gives the estimates of the probability of being in state j given the data to time t. The modified predictors X_t^{t-1} are given by

$$\mathbb{E}(X_t\,|Y_{t-1}) = X_t^{t-1} = \Phi X_{t-1}^{t-1} \tag{4.5.14}$$

with associated error variance-covariance matrix given by

$$P_t^{t-1} = \Phi P_{t-1}^{t-1}\Phi' + \mathrm{Q} \tag{4.5.15}$$

The modified filters X_t^t are given by

$$\mathbb{E}(X_t\,|Y_t) = X_t^t = X_t^{t-1} + \sum_{j=1}^{m} \pi_j(t\,|t) K_{tj}\varepsilon_{tj}, \tag{4.5.16}$$

where the innovation (or residual), i.e., error in predicting Y_t from the point $t-1$ is thus

$$\varepsilon_{tj} = Y_t - M_j X_t^{t-1} \tag{4.5.17}$$

and the associated error variance-covariance matrix is given by

$$P_t^t = \sum_{j=1}^{m} \pi_j(t\,|t)(I - K_{tj}M_j)P_t^{t-1}, \tag{4.5.18}$$

where

$$K_{tj} = P_t^{t-1}M_j'\left(\Sigma_{tj}^{-1}\right)' + \mathrm{Q} \tag{4.5.19}$$

and the innovation value,

$$\Sigma_{tj} = M_j P_t^{t-1} M_j' + R. \tag{4.5.20}$$

Also, the modified filters $\pi_j\left(t \mid t\right)$ are given by

$$\pi_j(t\,|t) = \frac{\pi_j(t) f_j(t\,|t-1)}{\sum\limits_{k=1}^{m} \pi_k(t) f_k(t\,|t-1)}, \tag{4.5.21}$$

where the $f_i\,(t \mid t-1)$ denote the conditional density of Y_t given the past $Y_t, Y_{t-1}, ..., Y_1$ and for $A_t = M_j$ for $j = 1, ..., n$, and we assume the distribution $\pi_j(t)$ for $j = 1, ..., m$ has been specified before observing $Y_t, Y_{t-1}, ..., Y_1$.

If we have no reason to prefer one state over another at time t, the choice of uniform priors, $\pi_j(t) = m^{-1}, j = 1, ..., m$, suffices. Smoothness can be introduced by letting

$$\pi_j(t) = \sum_{i=1}^{\mathrm{m}} \pi_i(t-1\,|t-1)\pi_{ij}, \tag{4.5.22}$$

where the nonnegative weights π_{ij} are chosen so that $\sum_{i=1}^{m} \pi_{ij} = 1$. Although $f_j\,(t \mid t-1)$ can be expressed in an explicit form, its evaluation can be highly computationally intensive and as such a remedy is to approximate it using the closest (in the sense of Kullback-Leibler distance) normal distribution. In this case the approximation leads to a choice of normal distribution with the same mean and variance associated with $f_j\,(t \mid t-1)$; that is, we approximate $f_j\,(t \mid t-1)$ by a normal with mean $M_j X_t^{t-1}$ and variance $\Sigma_{tj} M_j P_t^{t-1} M_j' + R$ as in the innovation value in (4.5.20) above.

4.5.3 The Maximum Likelihood Estimation Procedure

The joint density of the observed data is given by

$$f(\mathbf{y}_1, \cdots, \mathbf{y}_n) = \prod_{t=1}^{n} f(\mathbf{y}_t\,|Y_{t-1}\,) = \prod_{t=1}^{n} \sum_{j=1}^{m} \Pr(A_t = M_j\,|Y_{t-1}\,) f(\mathbf{y}_t\,|A_t = M_j, Y_{t-1}) \tag{4.5.23}$$

and so the log-likelihood can be written as

$$\ln \mathcal{L}_Y(\Theta) = \sum_{t=1}^{\mathrm{n}} \ln \Big(\sum\nolimits_{j=1}^{m} \pi_j(t) f_j(t\,|t-1)\Big), \tag{4.5.24}$$

where $\Theta = \{\mu_0, \Phi, Q, R\}$ is the vector of parameters containing respectively the elements of the initial mean, the transition matrix, the state and observation covariance matrices. We consider maximizing (4.5.24) directly as a function of the parameters using a Newton method, or we may consider applying the EM algorithm to the complete data likelihood. Thus the parameters in the model are $\Theta = (\alpha, \phi_0, \phi_1, \sigma, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2, \pi_1)$ and they can be estimated by the method of maximum likelihood on the basis of the following likelihood

$$\ln \mathcal{L}_Y(\Theta) = \sum_{t=1}^{n} \ln \left(\sum_{j=0}^{1} \pi_j(t) f_j(t\,|t-1)\right), \tag{4.5.25}$$

where the density $f_j\,(t \mid t-1)$ is approximated by the normal density with mean $\nu_t^{t-1} + \mu_{\mathrm{j}}$ and variance σ_j^2. The quasi-maximum likelihood estimates of the model can be obtained by maximizing the log-likelihood function with respect to the unknown parameters.

4.5.4 The filtering Equations

Equations (4.5.4−4.5.11) are related to (4.5.12−4.5.21) in the model by the following filtering equations for ease of programming in R:

$$\nu_{t+1}^t = \phi_0 + \phi_1 \nu_t^{t-1} + \sum_{j=0}^{1} \pi_{tj} k_{tj} \varepsilon_{tj} \tag{4.5.26}$$

$$P_{t+1}^t = \phi_1^2 P_t^{t-1} + \sigma_\omega^2 - \sum_{j=0}^{1} \pi_{tj} k_{tj}^2 \Sigma_{tj} \tag{4.5.27}$$

$$\varepsilon_{t0} = y_t - \alpha - \nu_t^{t-1} - \mu_0 \tag{4.5.28}$$

$$\varepsilon_{t1} = y_t - \alpha - \nu_t^{t-1} - \mu_1 \tag{4.5.29}$$

$$\Sigma_{t0} = P_t^{t-1} + \sigma_0^2 \tag{4.5.30}$$

$$\Sigma_{t1} = P_t^{t-1} + \sigma_1^2 \tag{4.5.31}$$

$$k_{t0} = \frac{\phi_1 P_t^{t-1}}{\Sigma_{t0}} \tag{4.5.32}$$

$$k_{t1} = \frac{\phi_1 P_t^{t-1}}{\Sigma_{t1}}. \tag{4.5.33}$$

4.6 Estimation of Parameters

As indicated in Sections 4.4 and 4.5 the Kalman-filtering procedure allows us to estimate the state variables over time given particular assumptions about the process for which all of the previous probabilistic results assumed that the parameters of the process were known. The Kalman filtering paradigm also allows one to calculate efficiently the likelihood of a set of observations given a particular set of parameters (see, e.g., Harvey 1989, Chapter 3.4, for details). By varying the parameters and rerunning the Kalman filtering for each set of parameters, we

can identify the set of parameters that maximizes this likelihood function. The iterative procedure was performed by the Stochastic Volatility function (SVfilter) in R software programming until an appropriate convergence criterion is satisfied. In our model there are seven parameters to be estimated $\Theta = (\alpha, \phi_0, \phi_1, \sigma, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2, \pi_1)$ plus the terms in the covariance matrix for the errors of measurement (P_t^t). We used the optimization function optm in R to request the numerical optimization, a Broyden-Fletcher-Goldfarb-Shanno (SFGS) method and the associated Hessian matrix to determine estimates of the parameters and standard errors for these estimates respectively. To be sure that our routine for the estimation of the maximum likelihood reaches global (rather than local) maximum, we reran the optimization problem from a variety of initial values of the (especially $\phi_0 = 0.1, 0.2, ..., 0.9$ and recall that $\phi_0 + \phi_1 = 1$) parameters. In all cases we started the Kalman filter with a prior mean ($\alpha = \sum_t^N y_t/N$) and covariance matrix ($P_0^0 \neq 0$) based respectively on the observed means and covariance in the data. Although the likelihood scores vary somewhat, the estimated state variables and parameters did not appear to be very sensitive to the assumed initial mean and covariance. Some results of the estimation are plotted for visual interpretion.

4.7 Empirical Results

To conduct an empirical implementation of the proposed model we use the historical time series data of Pennsylvania Daily Electricity Forward Contract[12] from January 1, 2002 to October 10, 2010. Consider the daily log-return series as discussed earlier in Section 4.3. The series show signs of positive autocorrelation in the squared returns. This characteristic is commonly referred to as volatility clustering, a common feature of financial returns data that usually induce excess kurtosis. Panels (b) and (c) of Figure 4.7.1 show the daily log-returns of the prices of Pennsylvania Electricity Futures Contract. It can be seen that the price volatility is not constant over time and in addition, the volatility is lowest at the second quarter of the entire time horizon (Figure 4.7.1 (c) and (d)). As a consequence, a normal distribution is not capable of describing the return series adequately. As we see in Chapter Five both histogram

[12]Http://www.eia.doe.gov/cneaf/electricity/wholesale/wholesale

displays and goodness-of-fit tests of daily returns clearly expose the inadequacy of the normal distribution to describe the system and correctly estimate the probability of both low returns around zero and extremely high absolute returns. One possibility to overcome the shortcomings of the normal distribution is to employ a mixture of two (see Section 4.2.2) or more normal distributions. Mixture distributions are useful in the context of overdispersed or multimodal data that may be caused by unobserved heterogeneity in the data.

It appeared from Figure 4.7.1 (c) that quiet periods, characterized by relatively small returns, alternate with relatively volatile periods, where price changes are rather large. This can be confirmed by the examination of the first-order autocorrelation function (ACF) of returns and squared returns in Figure 4.7.2 ((a) and (b)). While the autocorrelations of the return series only show minor activity (Figure 4.7.2 (a)), the autocorrelation function of squared returns show significant correlations up to an extended lag length (Figure 4.7.2 (b)).

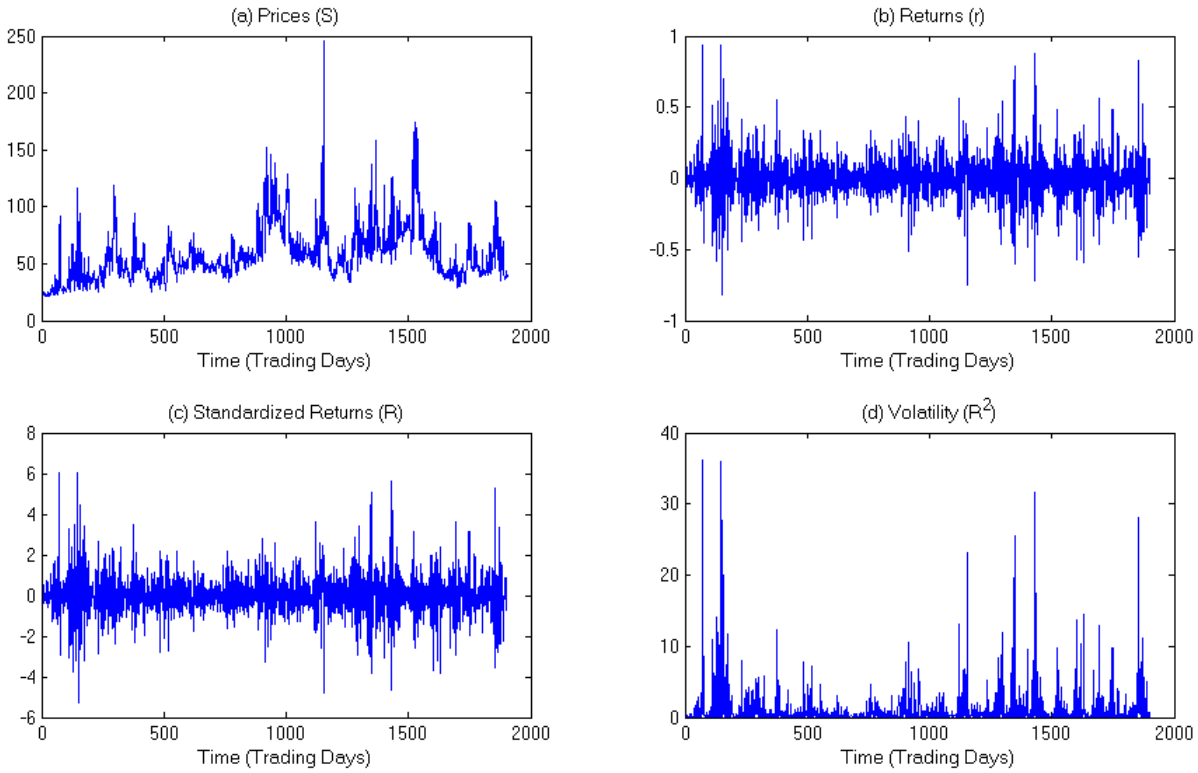


Figure 4.7.1: Plots of prices of Pennsylvania Electricity futures Contracts for 1900 trading days

Table 4.7.1: Values of estimates of parameters and their standard errors for various combinations of π_0 and π_1

Prob	π_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	π_1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Para	ϕ_0	.1445	.1531	.1606	.1582	.1481	.1402	.1332	.1269	.1234
		(.1645)	(.1806)	(.1881)	(.1815)	(.1718)	(.1630)	(.1555)	(.1447)	(.1294)
	ϕ_1	.9363	.9294	.9256	.9268	.9296	.9324	.9350	.9387	.9437
		(.0263)	(.0280)	(.0286)	(.0272)	(.0257)	(.0247)	(.0240)	(.0231)	(.0221)
	σ	.3544	.3834	.3914	.3867	.3769	.3649	.3511	.3303	.2965
		(.0843)	(.0839)	(.0835)	(.0803)	(.0771)	(.0747)	(.0728)	(.0705)	(.0675)
	α	-2.5810	-2.5810	-2.5691	-2.6317	-2.7138	-2.7588	-2.8401	-2.9475	-3.1198
		(2.3588)	(2.3889)	(2.3665)	(2.3367)	(2.3076)	(2.2872)	(2.2698)	(2.392)	(2.1470)
	σ_0^2	.2829	.5269	.7028	.8632	1.0088	1.1428	1.2717	1.4103	1.5941
		(.2513)	(.1295)	(.1028)	(.0875)	(.0769)	(.0697)	(.0652)	(.0628)	(.0595)
	μ_1	-1.7896	-1.9139	-2.0693	-2.2560	-2.4807	-2.7713	-3.1899	-3.8652	-5.1415
		(.1934)	(.1700)	(.1708)	(.1815)	(.1981)	(.2204)	(.2525)	(.3030)	(.3641)
	σ_1^2	2.4744	2.5427	2.6085	2.6726	2.7379	2.8046	2.8601	2.8608	2.6600
		(.0672)	(.0745)	(.0837)	(.0957)	(.1112)	(.1315)	(.1601)	(.2062)	(.2659)
Criteria	IV	1537.47	1474.39	1420.39	1378.82	1350.42	1336.13	1338.65	1365.04	1438.29
	LLH	1304.31	1280.40	1263.51	1251.22	1242.16	1235.77	1232.20	1232.59	1240.46
	Noi	32	37	39	22	28	42	26	41	53

NB: Prob = probability; Para = parameter with its standard error in parenthesis; Criteria: IV = initial value; LLH = log likelihood value and Noi = number of iterations before convergence.

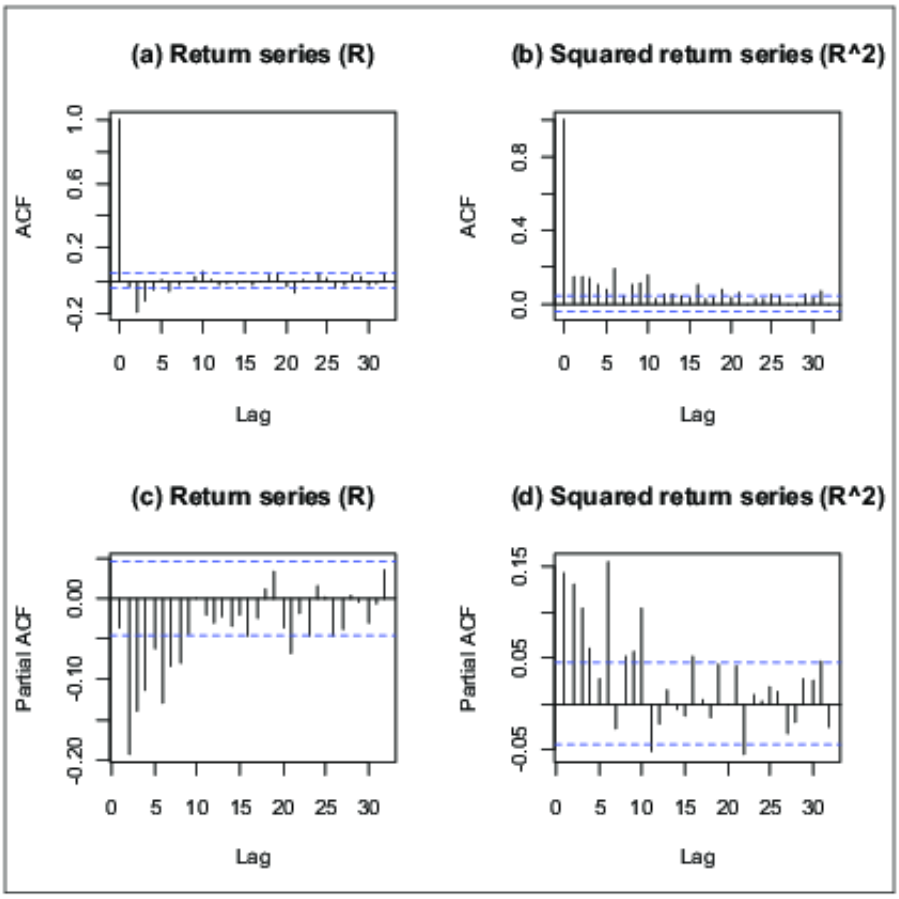


Figure 4.7.2: Autocorrelation and partial autocorrelation functions for the return series ((a) and (c)) and squarded series ((a) and (d))

The main steps we performed towards the estimation of the parameters Θ were the following. Firstly we chose an initial guess for Θ, (see Table 4.7.2). The starting values of time-varying probabilities π_0 and π_1 as stated in Section 4.6 (see Table 4.7.1) and the combination that generates the highest value of the log-likelihood is chosen. This approach assures optimal value of estimates of the parameters. The log-likelihood[13] is also initialized to zero. At each iteration the value of the log-likelihood function was compared against the values in the previous iteration and, if the difference between the current and the old value is positive and smaller than a specified quantity (we choose 0.00001), the iterative procedure terminates.[14] The results

[13]The log-likelihood is actually the "strength of evidence" about the likelihood that a given estimate is optimal relative to others.

[14]The choice of the maximum value of the log-likelihood function is informed by the MLE principle that gives

are given in Table 4.7.2. Estimates came from the combination $(\pi_0, \pi_1) = (0.1, 0.9)$ with the highest log-likelihood function value of 1304.3068 in 32 iterations.

Table 4.7.2: Estimation Results Penn Fit

Parameter	Initial Values	Estimates(standard error)
ϕ_0	0	0.1445(0.1645)
ϕ_1	0.95	0.9363(0.0263)
σ	0.20	0.3544(0.0843)
α	57.98	-2.5810(2.3588)
σ_0^2	1.00	0.2849(0.2513)
μ_1	-3.00	-1.7896(0.1934)
σ_1^2	2.00	2.4744(0.0672)

Figure 4.7.1 shows the observed prices, S_t, log returns, r_t, standardized returns, R_t, and the squared standardized returns, R_t^2, which we refer to in this Chapter as a measure of observed volatility. On closer examination of Figure 4.7.1 (c and d) one observes volatility clustering in the process. The influence of initial values of π_0 and π_1 on the parameter estimation is shown in Figure 4.7.3 where the three state parameters (Figure 4.7.3(a)) seem to be invariant while the other two variances (σ_0^2 and σ_1^2) change monotonically with increasing values of π_0. The optimal value of the log-likelihood function is shown to be a function of the combinations of the initial values of π_0 and π_1 (Figure 4.7.3 (b)).

The first four iterations (see, for instance, Figure 4.7.4) confirm the fact that the filter eventually forgets initialization errors and then slowly converges to the optimal values of the parameters. The coefficient of mean-reversion is estimated by $\phi_1 = 0.9363$. This value translates to a *half-life* of price of electricity of nine months ($= -ln(0.5)/\phi_1$) which is a significant rate of mean-reversion. The key property of the mean reversion is its half-life which is the time taken for the price to revert to half of its long-run level from the current level if no random shocks arrive.

the optimal values of the parameters of the population that have most likely generated that sample. We are also guided by the fact that the estimators of the maximum likelihood function are consistent and asymptotically efficient. It can be shown that each iteration on this algorithm increases the value of the likelihood function.

Hence the half-life of nine months means that on the average shocks to the price of electricity takes about nine months to decay to half their deviation from the long run level estimated as $\phi_0 = 0.1445$. Related to the coefficient of mean reversion is the parameter σ (= 0.3544 or 35.44%) which measures how volatile the price of electricity fluctuates around its long-run mean.

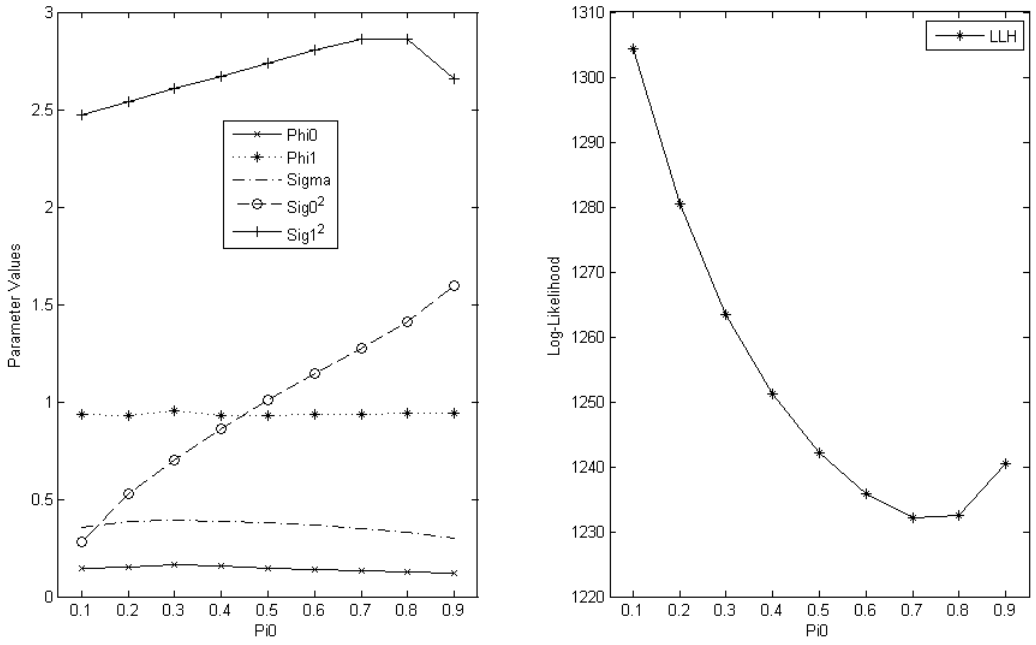


Figure 4.7.3: Influence of initial values of π_0 and π_1 on the parameters and the log-likelihood function

Figures 4.7.5 and 4.7.6 depict this 35% high volatility and the nine months half-life. The estimates of σ_0^2 and σ_1^2 as 0.2829 and 2.4744 respectively indicate that major daily shocks occur frequently with sizeable effect on volatility of the system at the two states (or regimes). This is understandable and reflects the probability of the system of being in state $j = 1(\pi_1 = 0.9)$.

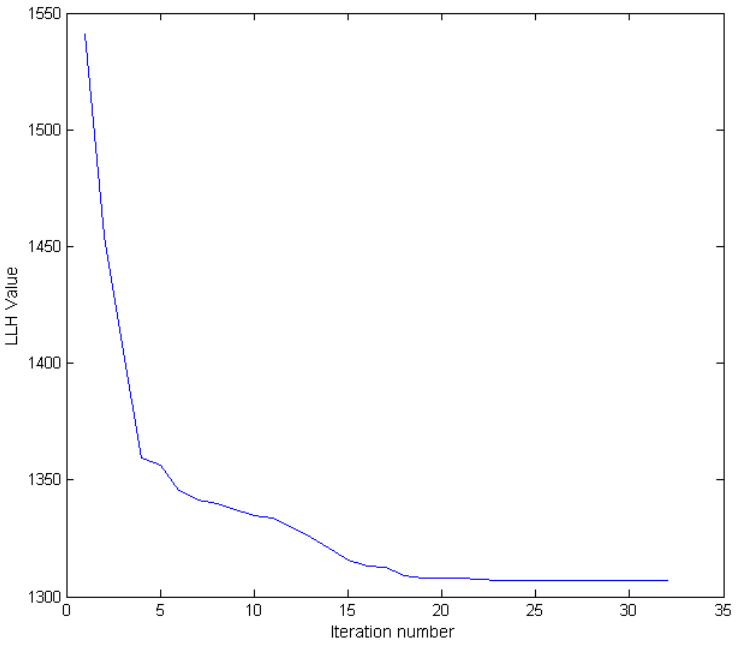


Figure 4.7.4: The convergence sequence of the algorithm

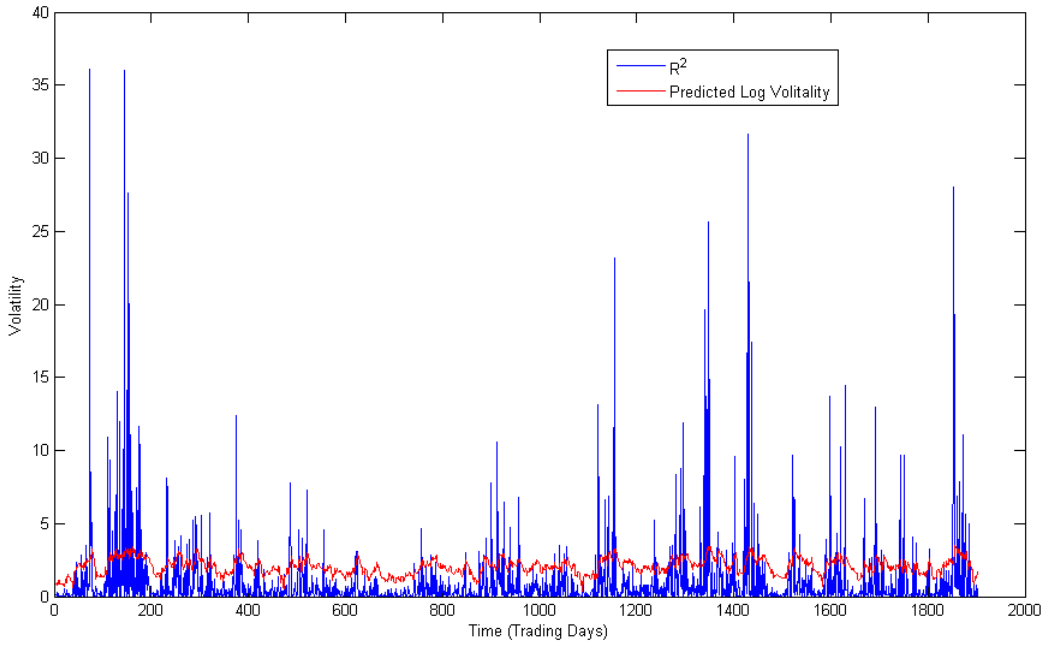


Figure 4.7.5: Observed volatility and prediced (filtered) log volatility compared

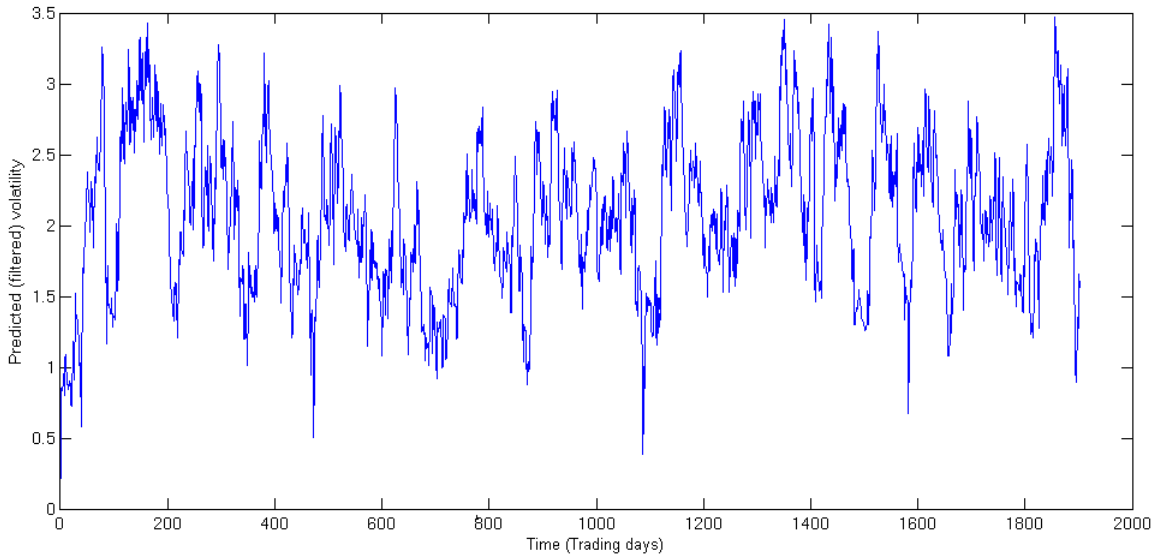


Figure 4.7.6: Predicted volatility

4.8 Chapter Summary

We derived a dynamic linear model in Section 4.4 based on the general Heston's (1997) type diffusion stochastic differential equation underlying energy futures prices. This model is flexible in the sense that we were able to construct, through the model observation equation, a Markov switching regime so that the modified Kalman filter is implemented. Empirical results show that volatility appeared to be very high for the daily data which exhibited volatility clustering and followed by mean reversion with half-life of nine months. This result is similar to those obtained in Krichene (2008) for crude oil pieces using GARCH(1,1) and Kellerhals (2004) for California electricity futures and spot prices using affine structure models. Our preliminary study of this dataset as visualized in Figure 4.7.2 show an autocorrelation of lag one authenticating our modeling the process state variable as a first order autoregressive AR(1) process.

Chapter 5

Selection of a Model for the Process Describing Energy Prices

5.1 Introduction

The values of the rate of mean-reversion and volatility parameters obtained in Chapter Four show that there are frequent large fluctuations in prices of Pennsylvania futures contracts. The observed volatility clustering indicates the presence of heavy tails in the returns series. In this second and concluding chapter of Part Two of this Thesis we continue in the modeling of the energy price generating process by the examination of the properties of the process. We propose an appropriate probability distribution for the process using empirical results from goodness-of-fit tests. We identify two families of probability density functions to study the characteristics of these series: the Generalized Hyperbolic (GH) distributions for the return series, and the extreme value distributions for the volatility series. Each of these families has members which we discuss shortly. We briefly review what motivates our further search for the characteristics of these series in the next Section while we present the concomitant variables of interest in Section 5.3. The candidate probability distributions for the two series are defined in Sections 5.4 and 5.5 respectively for return and volatility series. The normal (Gaussian) and the lognormal distributions however are not discussed further as they have been presented in Chapter One.

Strategies for selection of candidate models are presented in Section 5.6 while implementation and discussion of empirical results are in Sections 5.7 and 5.8 respectively. Included in these two sections is a comparative study of the characteristics of three basic energy datasets; electricity, crude oil and natural gas. Section 5.9 summarizes the Chapter findings.

5.2 Motivation

The large price fluctuations frequently observed in energy markets lead to nonnormal deviations from the long-term mean towards which the prices revert. Schwartz (1997) introduced the mean-reversion process which has become a popular class of stochastic models in recent literature on commodity price. Given prior information on the behaviour of stocks and that the log-returns of assets are frequently heavy tailed thereby violating the normal hypothesis implied by geometric Brownian motion (see, for example, Fama (1965), Mandelbrot (1963)), it becomes necessary to call for generalizations in modeling large changes in futures prices.The combination of features of normal and stable distributions especially those of the Levy processes of hyperbolic type offer more flexibility in modeling financial time series data. In addition economic analysis of risks in commodity markets depends upon accurate estimation of the probability of tail quantiles. Barndorff-Nielsen (1994) found a good fit in GH distributions to Danish stock returns. Hyperbolic distributions, a family member of the GH which also have exponentially decreasing tails,were independently suggested as distributions of German stock returns represented in the stock index DAX by Eberlein and Keller (1995) and Küchler *et al* (1994). The logarithm of the density of a Hyperbolic distribution is an hyperbola. Extreme value theory can provide a promising estimation of the tail part of risk. In this Chapter we intend to apply such extreme value models as the Weibull distribution to estimate the tail risk of our data sets. From the early 1990s application of extreme value theory in modeling financial extremes has become more and more popular, especially in measuring Value at Risk (VaR) on the tails of the Profit & Loss (P& L) distribution (Chen and Chen, 2002). Malevergne and Sornette (2004) found the so called "modified" Weibull distribution useful for financial purposes and specifically for portfolio and risk management. This is because the distribution offers a flex-

ible parametric representation of the distribution of returns on assets either in a conditional or unconditional framework. Also in this class is the Inverse Gaussian (IG) distribution for modeling nonnegative random variables. These distributions are compared especially against the Gaussian model. There are some well-known empirical facts about log returns on stocks (Granger, 2005):

(i) Log returns are reasonably approximated by uncorrelated identically distributed random variables (independent in the Gaussian case).

(ii) The empirical distribution is leptokurtic and heavier-tailed compared to the normal distribution.

(iii) Although there is no significant serial correlation in stock returns, there is serial correlation in squared-log returns.

Empirical and theoretical investigations of (i) above have a long history. Fama (1963, 1965), Mandelbrot (1963) and Mandelbrot and Taylor (1967) proposed Pareto-stable distributions to explain the excess kurtosis in stock returns while Mittnik and Rachev (1993) give an overview and comparison of alternative distributions in modeling stock returns. Clewlow and Strickland (2000), Eydeland and Wolyniec (2003) and Pilipovic (1998) incorporated an additional jump-noise term into the stochastic differential equation in defining Schwartz' dynamics which unfortunately makes statistical fitting more cumbersome as the noise comes in a multiplicative manner. A deseasonalized spot price model as the exponential of a non-Gaussian Ornstein-Uhlenbeck process, as suggested by Benth and Salityte-Benth (2004) has the advantage when fitting the model to data. This alternate definition of the stochastic dynamics of the spot price suits the objectives of this Chapter.

5.3 Variables of Interest

The basic quantity under investigation and for individual product with index i is the commodity price process $(S_t)_{t\in[0,\infty]}$. The time, t, runs over the trading days (weeks) of the spot/futures

prices to time T. For each product we calculate the returns defined by the logarithmic differences,

$$r_i(t) = \ln S_i(t) - \ln S_i(t-1) = \ln\left(\frac{S_i(t)}{S_i(t-1)}\right) \tag{5.3.1}$$

where $(S_t)_{t\geq 0}$ is the price process (over time, t, where $(0 \leq t \leq T)$ of the Levy process $L_t = \ln(S_t)$ through a geometric Levy process

$$S_t = S_0 \exp(L_t). \tag{5.3.2}$$

Because the products under study are priced under different market environments and recorded in different units, we standardize $r_i(t)$ for ease of comparison of the distributional forms. If we let $\overline{r_i}$ and υ_i be the sample mean and standard deviation of the returns series then define the standardized process as

$$R_i(t) = \frac{r_i(t) - \overline{r_i}}{\upsilon_i} \tag{5.3.3}$$

where $-\infty \leq R_i(t) \leq \infty$. While we work with $R_i(t)$ as in Amaral *et al* (2000) and Barndorff-Nielsen and Prause (2001), some other researchers such as Wang *et al* (2007) suggest standardized volatility series as

$$G_i(t) = \frac{|r_i(t) - \overline{r_i}|}{\upsilon_i} \qquad \text{for} \quad G_i(t) \geq 0, \tag{5.3.4}$$

where υ_i in (5.3.4) is defined as

$$\upsilon_i = \frac{1}{N}\sum_{t=1}^{N} |r_i(t) - r_i(t-1)| \tag{5.3.5}$$

Mandelbrot (1963), Müller *et al* (1990) and Guillaume *et al* (1997) stressed that the modeling of high frequency data in finance is to analyze volatility on different time scales. They then defined volatility measure, υ_i, as the average of absolute logarithmic price change as in (5.3.5) instead of the standard deviation of the dataset. Given a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, a Levy process $L = L_t, t > 0$ is an infinitely divisible continuous-time stochastic process, $L_t : \Omega \to \mathbb{R}$, with stationary and independent increments. Levy processes are more versatile than Gaussian-driven processes as they can model skewness, excess Kurtosis and even Jumps. Let $X_1, X_2, ..., X_n$ be n statistically independent observations of a random variable $X(t)$, here X represents our $R(t) \in \mathbb{R}$ for the standardized return series or $G(t) = |R(t)| \in \mathbb{R}^+$ for volatility measure.

5.4 The Generalized Hyperbolic Distribution Family

We start with an exposition of the univariate Generalized Hyperbolic (GH) Distribution introduced in the literature by Ole Barndorff-Neilsen in 1977 while modeling particle size from a diamond mine (see, e.g., Barndorff-Neilsen, 1977) and the subclasses which are relevant for application in this Thesis. The distribution is well applied in economics particularly in the fields of modeling financial markets and risk management due to its semi-heavy tails.

5.4.1 The Generalized Hyperbolic Distribution

A random variable X is said to follow a Generalized Hyperbolic (GH) distribution if its probability density function is given by

$$f_{GH}(x;\alpha,\beta,\delta,\lambda,\mu) = \frac{(\gamma/\delta)^{\gamma}}{\sqrt{2\pi}K_{\lambda}(\delta\gamma)} \frac{K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2+(x-\mu)^2})}{(\sqrt{\delta^2+(x-\mu)^2)/\alpha})^{\frac{1}{2}-\lambda}} \beta(x-\mu) \tag{5.4.1}$$

where $\gamma = \sqrt{\alpha^2-\beta^2}, \mu, \lambda, \alpha, \beta, \delta \in R$ and $\mu, \beta$and δ are location, asymmetry and scale parameters respectively while K_{λ} is the modified Bessel function of the third kind with index λ. $\delta \geq 0$ and $0 \leq |\beta| < \alpha$. The mean and variance of this distribution are respectively given by

$$\mathbb{E}[X] = \mu + \frac{\delta\beta K_{\lambda+1}(\delta\gamma)}{\gamma K_{\lambda}(\delta\gamma)} \tag{5.4.2}$$

and

$$\mathbb{V}[X] = \frac{\delta K_{\lambda+1}(\delta\gamma)}{\gamma K_{\gamma}(\delta\gamma)} + \left(\frac{\beta\delta}{\gamma}\right)^2 \left(\frac{K_{\lambda+2}(\delta\gamma)}{K_{\lambda}(\delta\gamma)} - \frac{k^2_{\lambda+1}(\delta\gamma)}{K^2_{\lambda}(\delta\gamma)}\right) \tag{5.4.3}$$

Special cases of the generalized hyperbolic distribution (see, e.g., Jørgensen (1982), Barndorff-Neilsen and Stelzer (2004)) are

(i) When $\lambda = -\frac{1}{2}$, the GH specializes to the Normal Inverse Gaussian (NIG) and

(ii) When $\lambda = 1$, the GH becomes the Hyperbolic distribution.

Definition 5.4.1 (Modified Bessel Function of the Third Kind with Index λ)**.** [1] The integral representation of the modified Bessel function of the third kind with index λ can be found in

[1] In addition we have an explicit from of the Bessel fuction, $K_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}}\exp(-x)$

Barndorff-Neilsen *et al* (1982) and Abramowitz and Stegun (1972):

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1} exp\left\{-\frac{x}{2}\left(y - y^{-1}\right)\right\} \mathrm{d}y, \quad x > 0. \tag{5.4.4}$$

The substitution $y = x\sqrt{\chi/\psi}$ can be used to obtain the following relation which allows one to bring the GH (5.4.1) into a closed-form expression

$$\int_0^\infty y^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{y} + y\psi\right)\right\} \mathrm{d}y = 2\left(\frac{\chi}{\psi}\right)^{\lambda/2} K_\lambda\left(\sqrt{\psi\chi}\right). \tag{5.4.5}$$

Asymptotic relations for small arguments x can be used for calculating the densities of special cases of the GH density as follows

$$K_\lambda(x) \sim \Gamma(\lambda) 2^{\lambda-1} x^{-\lambda} \quad \text{as } x \downarrow 0 \quad \text{and } \lambda > 0 \tag{5.4.6}$$

and

$$K_\lambda(x) \sim \Gamma(\lambda) 2^{\lambda-1} x^{-\lambda} \quad \text{as } x \downarrow 0 \quad \text{and } \lambda < 0 \tag{5.4.7}$$

The asymptotic relation for large arguments x is given in footnote 3.

5.4.2 The Normal Inverse Gaussian

A random variable X follows a Normal Inverse Gaussian (NIG) distribution with parameter vector $(\alpha, \beta, \mu, \delta)$ if its probability density function is

$$f_{NIG}(x : \alpha, \beta, \mu, \delta) = \frac{\alpha\delta \exp\left[p\left(x\right)\right]}{\pi q\left(x\right)} K_1\left[\alpha q\left(x\right)\right] \tag{5.4.8}$$

where $p(x) = \delta\sqrt{\left(\alpha^2 - \beta^2\right)} + \beta\left(x - \mu\right)$, $q(x) = \sqrt{\left(x-\mu\right)^2 + \delta^2}$ and K_1 is the modified Bessel function[2] of the third kind with order one (see e.g., Abramowitz and Stegun 1972). Here $\mu \in \mathbb{R}$ is a location density, $\beta \in \mathbb{R}$ is the skewness parameter and, if $\beta < 0$, the NIG is negatively skewed; $\alpha \geq |\beta|$ measures the heaviness of the tails (shape of the distribution) and finally $\delta > 0$ is the scale parameter. The NIG is a very flexible member of the family of distributions enjoying the convolution property as shown in Kalemanova and Werner (2006):

[2]Specifically $K_1(x) = \frac{x}{4}\int_0^\infty \exp\left\{t + \frac{x^2}{4t}\right\} t^{-2} dt, \quad x \in \mathbb{R}.$

Property 5.4.1

The NIG is a mixture of normal and inverse Gaussian distributions. Let

$$X|\,Y = \begin{cases} \quad y \sim N(\mu + \beta y, y) \\ Y \sim IG\left(\delta\gamma, \gamma^2\right) \text{ with } \gamma := \sqrt{\alpha^2 - \beta^2} \end{cases} \tag{5.4.9}$$

then $X \sim NIG\left(\alpha, \beta, \mu, \delta\right)$ is what is denoted by the density function

$$f_{NIG}\left(\alpha, \beta, \mu, \delta\right) = \int_0^\infty f_N\left(x; \mu + \beta y, y\right).f_{IG}\left(y; \delta\gamma, \gamma^2\right) dy. \tag{5.4.10}$$

Property 5.4.2

The NIG distribution is closed under convolution. In fact it is the only member of the family of general hyperbolic distributions to have the property that for independent random variables, $X \sim NIG\left(\alpha, \beta, \mu_X, \delta_X\right)$ and $Y \sim NIG\left(\alpha, \beta, \mu_Y, \delta_Y\right)$, their sum is NIG distributed, that is,

$$X + Y \sim NIG\left(\alpha, \beta, \mu_X, \delta_X\right) * NIG\left(\alpha, \beta, \mu_Y, \delta_Y\right) = NIG\left(\alpha, \beta, \mu_X + \mu_Y, \delta_X + \delta_Y\right) \tag{5.4.11}$$

The mean, variance, skewness and kurtosis of this random variable X are, respectively,

$$\mathbb{E}\left[X\right] = \mu + \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}}, \tag{5.4.12}$$

$$\mathbb{V}\left[X\right] = \frac{\delta\alpha^2}{\left(\alpha^2 - \beta^2\right)^{3/2}}, \tag{5.4.13}$$

$$\mathbb{S}\left[X\right] = \frac{3\left(\beta/\alpha\right)}{\left(\delta\sqrt{\alpha^2 - \beta^2}\right)^{1/2}} \tag{5.4.14}$$

and

$$\mathbb{K}\left[X\right] = \frac{3\left(1 + 4(\beta/\alpha)^2\right)}{\delta\sqrt{\alpha^2 - \beta^2}}. \tag{5.4.15}$$

However, moment estimators as starting values of the NIG distribution may be used. If $\bar{m}_i, i = 1, 2, 3, 4$, are the sample mean, variance, skewness and kurtosis respectively, then define

$$\hat{\gamma} = \frac{3}{\bar{m}_3\sqrt{3\bar{m}_4 - 5\bar{m}_3^2}}.$$

The moment estimators are then given by

$$\hat{\mu} = \bar{m}_1 - \hat{\beta}\hat{\delta}\Big/\hat{\gamma},$$

$$\hat{\beta} = \left(\bar{m}_3\bar{m}_2\hat{\gamma}^2\right)/3,$$

$$\hat{\delta} = \left(\bar{m}_2^2\hat{\gamma}^3\right)\Big/\left(\hat{\beta}^2 + \hat{\gamma}^2\right)$$

and

$$\hat{\alpha} = \left(\hat{\beta}^2 + \hat{\gamma}^2\right)^{1/2}.$$

These initial values can also be estimated by the method of moments (Bolviken and Benth 2000) from a given sample $x_1, x_2, ..., x_n$ for $X \sim NIG(\alpha, \beta, \mu, \delta)$ through the ratio
$(\mathbb{S}[X])^2/\mathbb{K}[X]$ $, \mathbb{K}[X] > 0$.

5.4.3 The Hyperbolic Distribution

The random variable X is said to have a Hyperbolic (HYP) distribution if its probability density function is given by

$$f_{HYP}(x; \alpha, \beta, \delta, \mu) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \exp\left\{-\alpha\left(u(x)\right) + \beta\left(x - \mu\right)\right\}, \tag{5.4.16}$$

where $u(x) = \sqrt{\left(\delta^2 + (x-\mu)^2\right)}$ and $-\infty \leq x \leq \infty$. The domain of variation of the parameters is $\mu \in R, \delta > 0$, and $0 \leq |\beta| < \alpha$. The first application of the hyperbolic distribution to finance is in Eberlein and Keller (1995). The alternative set of distributions for modeling skew and heavy-tailed data is the skew extension to the Student's $t-$distribution. Hansen (1994) was the first to propose a skew extension to the Student's $t-$distribution for modeling financial returns. There are several versions of this distribution, for details, see for example Fernandez and Steel (1998), Branco and Dey (2001), Jones and Faddy (2003) and Azzalini and Capitanio (2003). However, all these skew-type distributions have both tails behaving like polynomials which mean that they fit fat-tailed data well but deficient in handling substantial skewness. The probability density function derived by Aas and Haff (2006) as a limiting case of the GH distribution $\left(\lambda = -\frac{\nu}{2} \quad and \quad \alpha \to |\beta|\right)$ in (5.4.1) which they referred to as GH skew Student's

$t-$distribution. The main attraction of this distribution is that unlike any other member of the GH family, it has one tail determined by a polynomial and the other by exponential behaviour. In addition, it is almost as analytically tractable as the NIG distribution. Therefore, the skew Student's $t-$distribution has one heavy and one semi-heavy tail.

5.4.4 The Skewed Student's $t-$distribution

A random variable X is said to follow a GH skew Student's $t-$distribution (SSt) if its (Aas and Haff, 2006) probability density function is given by

$$f_{SSt}\left(x;\nu,\mu,\beta,\delta\right)=\begin{cases}\dfrac{2^{(1-\nu)/2}\delta^{\nu}\left|\beta\right|^{(\nu-1)/2}K_{(\nu+1)/2}\left(\left|\beta\right|u\left(x\right)\right)}{\Gamma\left(\dfrac{\nu}{2}\right)\sqrt{\pi}\left(u\left(x\right)\right)^{(\nu+1)/2}}\exp\left\{\beta\left(x-\mu\right)\right\} & \text{for} \quad \beta\neq 0,\\ \dfrac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{\delta\Gamma\left(\dfrac{\nu}{2}\right)\sqrt{\pi}}\left[1+\dfrac{\left(x-\mu\right)^{2}}{\delta^{2}}\right]^{-(\nu+1)/2} & \text{for } \beta=0,\end{cases} \tag{5.4.17}$$

where $u\left(x\right)=\sqrt{\left(\delta^{2}+\left(x-\mu\right)^{2}\right)}$. It can be recognized that the density in (5.4.17) is that of a non-central (scaled) Student's $t-$distribution with ν degrees of freedom when $\beta=0$. The mean and variance of a SSt distributed random variable X are respectively

$$\mathbb{E}\left[X\right]=\mu+\frac{\beta\delta^{2}}{\nu-2} \tag{5.4.18}$$

and

$$\mathbb{V}\left[X\right]=\frac{2\beta^{2}\delta^{4}}{\left(\nu-2\right)^{2}\left(\nu-4\right)}+\frac{\delta^{2}}{\nu-2}. \tag{5.4.19}$$

Another subclass of the GH distributions family is the Variance-Gamma distribution that we consider in the next Subsection. It is the normal variance-mean mixture where the mixing density is the gamma distribution. The tails of the distribution decrease more slowly than the normal distribution. It is therefore suitable to model phenomena where numerically large values are more probable than is the case for the normal distribution. The distribution was introduced in the finance literature by Madan and Seneta (1990) and has been successful applied in diverse fields such as modeling returns from financial assets and turbulent wind speeds.

5.4.5 The Variance-Gamma distribution

Let X be a continuous random variable. X is said to be distributed as the Variance-Gamma (VG) distribution if its probability density function is of the form

$$f_{VG}\left(x;\alpha,\mu,\lambda,\beta\right)=\frac{\left(\alpha^{2}-\beta^{2}\right)^{\lambda}\left|x-\mu\right|^{\lambda-1/2}K_{\lambda-1/2}\left(\alpha\left|x-\mu\right|\right)}{\sqrt{\pi}\Gamma\left(\lambda\right)\left(2\alpha\right)^{\lambda-1/2}}\exp\left(\beta\left(x-\mu\right)\right), \tag{5.4.20}$$

where $-\infty<x<\infty$, μ (location parameter), α , β (asymmetry parameter) are real and $\lambda>0$. Here, $\Gamma\left(.\right)$ denotes the Gamma function, and K_{λ}, the Bessel function of the third kind. The mean and variance of X are

$$\mathbb{E}\left[X\right]=\mu+\frac{2\beta\lambda}{\alpha^{2}-\beta^{2}} \tag{5.4.21}$$

and

$$\mathbb{V}\left[X\right]=\frac{2\lambda}{\left(\alpha^{2}-\beta^{2}\right)}\left(1+\frac{2\beta}{\alpha^{2}-\beta^{2}}\right). \tag{5.4.22}$$

The class of Variance-Gamma distributions is closed under convolution in the following sense that if X_1 and X_2 are independent random variables that are variance-gamma distributed with the same values of the parameters α and β, but possibly different values of the other parameters, λ_1,μ_1 and λ_2,μ_2 respectively, then X_1+X_2 is variance-gamma distributed with parameters $\alpha,\beta,\lambda_1+\lambda_2$ and $\mu_1+\mu_2$.

5.5 The Extreme Value distribution family

5.5.1 The Weibull Distribution

A random variable X is assumed to follow a three-parameter Weibull distribution if its probability density function is of the form

$$f_{WEI}\left(x;\nu,\alpha,\beta\right)=\begin{cases}\left(\frac{\beta}{\alpha}\right)\left(\frac{x-\nu}{\alpha}\right)^{\beta-1}\exp\left\{-\left[\frac{\left(x-\nu\right)}{\alpha}\right]^{\beta}\right\}, & x\geq\nu\\ 0, & \text{otherwise}\end{cases} \tag{5.5.1}$$

where $\alpha,\beta\in\mathbb{R}^{+}$ while $\nu\geq0$ (i.e., nonnegative). Here α is the scale parameter, β is the shape parameter (and provides information about the properties of incurred risk mode) and

$\nu = \min(x_1, x_2, ..., x_n)$. The cumulative density function of X is

$$F_{WEI}(x; \nu, \alpha, \beta) = 1 - \exp\left\{-\left[\frac{(x-\nu)}{\alpha}\right]^{\beta}\right\}, \qquad x > \nu. \tag{5.5.2}$$

If $\nu = 0$ in equations (5.5.1) and (5.5.2), we have a 2-parameter Weibull distribution. However, in most practical situations $\nu \neq 0$ so that the following transformation becomes necessary

$$W = \begin{cases} X - \nu, & \text{if} \quad \nu \quad \text{is} \quad \text{positive} \\ X + \nu, & \text{if} \quad \nu \quad \text{is} \quad \text{negative} \end{cases} \tag{5.5.3}$$

The new random variable[3] $W \sim Wei(\alpha, \beta)$ has a probability density function given by

$$f_W(w; \alpha, \beta) = \begin{cases} \left(\frac{\beta}{\alpha}\right)\left(\frac{w}{\alpha}\right)^{\beta-1} \exp\left\{-\left[\frac{(w)}{\alpha}\right]^{\beta}\right\}, & w \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{5.5.4}$$

The maximum likelihood estimate of the mean and variance of X are given in Johnson *et al* (2004) as follows

$$\mathbb{E}[X] = \mu_X = \nu + \alpha\Gamma\left(1 + \frac{1}{\beta}\right) \tag{5.5.5}$$

and

$$\mathbb{V}[X] = \sigma_X^2 = \alpha^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}\right)\Gamma\left(1 + \frac{1}{\beta}\right)\right], \tag{5.5.6}$$

with skewness and kurtosis coefficients respectively as

$$\mathbb{S}[X] = \frac{\Gamma\left(1 + \frac{3}{\beta}\right)\alpha^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3} \tag{5.5.7}$$

and

$$\mathbb{K}[X] = \frac{\Gamma\left(1 + \frac{4}{\beta}\right)\alpha^4 - 4S(W)\mu\sigma^3 - 6\mu^2\sigma^2 - \mu^4}{\sigma^4}. \tag{5.5.8}$$

Parameter estimation [4] *:* The three parameters of the Weibull distribution in (5.5.1) can be estimated using the relationships (Derman *et al* 1973) as follows

$$\nu = \min(x_1, x_2, ..., x_n), \tag{5.5.9}$$

[3]The values of α and β remain the same as in (5.5.1) and (5.5.2), but ν as a location parameter only "shifts" the entire distribution to the left or right along the real line to locate the mean of the distribution.

[4]The values of $z = 1/\beta$ can be obtained quite easily from the table of the gamma function $\Gamma(1+z)\Gamma(1+z)/\Gamma(1+2z)$ and $\Gamma(1+z)$ provided in Derman *et al* (1973).

$$\alpha = \frac{\mu_X - \nu}{\Gamma\left(1+\frac{1}{\beta}\right)} \tag{5.5.10}$$

and

$$\frac{(\mu_X - \nu)}{\sigma_X^2 + (\mu_X - \nu)^2} = \frac{\Gamma\left(1+\frac{1}{\beta}\right)\Gamma\left(1+\frac{1}{\beta}\right)}{\Gamma\left(1+\frac{2}{\beta}\right)}. \tag{5.5.11}$$

Prominent among the Weibull subfamily are the Double Weibull (DW), Compound Weibull (CW) and Modified Weibull (MW). Because of the accumulated evidence against stable Paretian distributions, Mittnik and Rachev (1993) suggests the double Weibull (DW) distribution for stock returns with density

$$f_{DW}(x;\alpha,\lambda) = \frac{\beta}{2\alpha}\left|\frac{x-\nu}{\alpha}\right|^{\beta-1}\exp\left\{-\left(\frac{x-\nu}{\alpha}\right)^{\beta}\right\} \quad \alpha>0, \beta>0, -\infty<x<\infty. \tag{5.5.12}$$

One of their arguments in favour of this distribution is that tails decrease exponentially. (They estimate α to be close to 1.).

5.5.2 The Generalized Inverse Gaussian Distribution

The probability density function (pdf) of a Generalized Inverse Gaussian (GIG) is given for a random variable $X, X \in \mathbb{R}^+$, as

$$f_{GIG}(x,\lambda,\beta,\gamma) = \begin{cases} \dfrac{(\gamma/\beta)^{\lambda/2}}{2K_\lambda\left(\sqrt{\beta\gamma}\right)}x^{\lambda-1}\exp\left\{\dfrac{1}{2}\left(\dfrac{\beta}{x}+\gamma x\right)\right\}, & x>0 \\ 0, & \text{otherwise.} \end{cases} \tag{5.5.13}$$

This pdf has domain of the variation of the parameters $\beta, \gamma \in R^+$and $\lambda \in R$ while K_λ is the modified Bessel function of the third kind with index λ. The parameters satisfy the following conditions

$$\begin{aligned} &\beta \geq 0, \gamma > 0, \text{ if } \lambda > 0, \\ &\beta > 0, \gamma > 0, \text{ if } \lambda = 0, \\ &\beta > 0, \gamma \geq 0, \text{ if } \lambda < 0. \end{aligned} \tag{5.5.14}$$

Jørgensen (1982) and Barndorff-Neilsen and Stelzer (2004) have shown that if $\beta, \gamma > 0$, then the rth moment of X can be computed using

$$\mathbb{E}[X^r] = \left(\frac{\beta}{\gamma}\right)^{r/2}\frac{K_{\lambda+r}\left(\sqrt{\beta\gamma}\right)}{K_\lambda\left(\sqrt{\beta\gamma}\right)} \tag{5.5.15}$$

especially when $r = \pm 1$ and 2, and

$$\mathbb{E}\left[\log\left(X\right)\right] = \left.\frac{d\mathbb{E}\left[X^r\right]}{dr}\right|_{r=0}. \tag{5.5.16}$$

Equation 5.5.13 needs to be evaluated numerically. Jørgensen (1982) investigated a class of GIG and observes as follows

(i) When $\lambda = -\frac{1}{2}$, then the GIG specializes to the two parameter inverse Gaussian(IG) distribution, which can be given the probabilistic interpretation as a distribution of the first hitting time to the level $\sqrt{\beta}$ of a Brownian motion with drift $\sqrt{\gamma}$ and unit diffusion coefficient (Rydberg 1999);

(ii) When $\lambda > 0$ and $\gamma \to 0$ as $\beta \to \infty$, GIG tends to the Gamma distribution;

(iii) When $\lambda < 0, \beta > 0$ as $\gamma \to 0$, GIG tends to the Inverse Gamma (IGam) distribution.

This distribution has a tail of the Pareto type.

5.5.3 The Inverse Gaussian Distribution

A nonnegative random variable X has an Inverse Gaussian (IG) distribution with probability density function of the form

$$f_{IG}(x;\alpha,\beta) = \begin{cases} \dfrac{\alpha}{\sqrt{2\pi\beta x^3}} \exp\left\{-\left[\dfrac{(\alpha-\beta x)^2}{2\beta}\right]\right\}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{5.5.17}$$

with a corresponding distribution function as

$$F_{IG}(x;\alpha,\beta) = \begin{cases} \dfrac{\alpha}{\sqrt{2\pi\beta}} \int_0^x z^{-3/2} \exp\left\{-\dfrac{(\alpha-\beta z)^2}{2\beta z}\right\} dz, & x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{5.5.18}$$

where $\alpha, \beta \in \mathbb{R}^+$ and β is the diffusion coefficient. The first four central moments of $X \sim IG\left(\alpha,\beta\right)$ have been shown in Johnson *et al* (1994) (with $\mu = \frac{\alpha}{\beta}, \lambda = \frac{\alpha^2}{\beta}$) to be

$$\mathbb{E}\left[X\right] = \frac{\alpha}{\beta}, \quad \mathbb{V}\left[X\right] = \frac{\alpha}{\beta^2}, \quad \mathbb{S}\left[X\right] = \frac{3}{\sqrt{\alpha}}, \quad \mathbb{K}\left[X\right] = 3 + \frac{15}{\alpha}. \tag{5.5.19}$$

The square of the coefficient of variation (CV) of X is equal to $\frac{1}{\alpha}$.

The IG has some attractive statistical and probabilistic properties in modeling nonnegative and positively skewed data. For example, it belongs to the exponential family, has the reproductive property and possesses similar inferential properties to that of the normal distribution, see for example, Mudholkar and Natarajan (2002) and Chhikara and Folks (1989). This distribution can produce stable estimates of parameters in the presence of outliers, and, in general, it is highly flexible because it allows for different degrees of kurtosis and asymmetry other than modality and bimodality.

5.5.4 The Gamma Distribution

A nonnegative random variable X has a Gamma distribution with probability density function of the form

$$f_{GAM}(x;\alpha,\beta) = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, & x>0,\ \beta>0 \\ 0, & \text{elsewhere} \end{cases} \tag{5.5.20}$$

here $\Gamma(.)$ is the Gamma function and the associated moments of X are

$$\mathbb{E}[X] = \frac{\alpha}{\beta}, \quad \mathbb{V}[X] = \frac{\alpha}{\beta^2} \quad \text{and} \quad \mathbb{E}[\log(x)] = \psi(\alpha) - \log(\beta). \tag{5.5.21}$$

5.5.5 The Inverse Gamma Distribution

A nonnegative random variable X has an Inverse Gamma (IGam) distribution with probability density function of the form

$$f_{IGam}(x;\alpha,\beta) = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left\{-\dfrac{\beta}{x}\right\}, & x>0,\ \beta>0 \\ 0, & \text{elsewhere} \end{cases} \tag{5.5.22}$$

If $X \sim Gam(\alpha,\beta)$, then $X^{-1} \sim InvGam(\alpha,\beta)$. The following moments are associated with X:

$$\mathbb{E}[X] = \frac{\beta}{\alpha-1} \qquad \text{if } \alpha>1, \tag{5.5.23}$$

$$\mathbb{V}[X] = \frac{\beta}{(\alpha-1)^2(\alpha-2)} \qquad \text{if } \alpha>2 \tag{5.5.24}$$

and

$$\mathbb{E}\left[\log\left(x\right)\right] = \log\left(\beta\right) - \psi\left(\alpha\right). \tag{5.5.25}$$

5.6 Procedures for Model Selection

In Sections 5.4 and 5.5 we described different competing probabilistic models from the family of generalized hyperbolic distributions and the family of distributions modeling non-negative random variables respectively. In this section we outline different methods for choosing the best fitting model to a given dataset. Suppose there are two families, say, $\mathcal{F} = \{f\left(x;\theta\right);\theta \in \mathbb{R}^p\}$ and $\mathcal{G} = \{g\left(x;\varphi\right);\varphi \in \mathbb{R}^q\}$, the problem is to choose the correct family for a given dataset $\{x_1, x_2, ..., x_n\}$. The methods we describe in the following Subsections are used for model discrimination in the next Section.

5.6.1 Maximum Likelihood Criterion

Suppose a random variable X has a density function $f\left(x;\theta_1,\theta_2,...,\theta_k\right)$ that depends on k parameters. Let $\hat{\theta}_i$ denote the maximum likelihood estimator (MLE) of θ_i for the likelihood function $\mathcal{L}\left(X,\theta\right) = \prod_{i=1}^{n} f\left(x_i;\theta_1,\theta_2,...,\theta_k\right)$. Similarly let $\hat{\theta}'_i$ denote the MLE of θ'_i from another density function with likelihood function $\mathcal{L}\left(X,\theta'\right)$. The maximum likelihood principle proposed in Cox (1962) is a maximum likelihood ratio test procedure

$$\mathcal{T}\left(\hat{\theta},\hat{\theta}'\right) = \sum_{i=1}^{n} \ln\left(\frac{f\left(x_i|\hat{\theta}\right)}{f\left(x_i|\hat{\theta}'\right)}\right), \tag{5.6.1}$$

where $\hat{\theta}$ and $\hat{\theta}'$ are maximum likelihood estimators of parameter vectors of competing models. Because the estimators provide the best explanation of the observed data, we choose the density $\mathcal{F}$ if $\mathcal{T} > 0$, otherwise choose $\mathcal{G}$. The solution $\mathcal{F}$ is sometimes called the Cox's statistic. Lu *et al* (2002) observed that the statistic $\ln\mathcal{T}$ should be asymptotically normally distributed when properly normalized. Other researchers, see for example, Voung (1989), Fearn and Nebenzah (1991), Martial *et al* (2001), Kundu and Manglick (2004) and Kundu *et al* (2005) studied the regularity conditions needed for the asymptotic distribution to hold. They used the likelihood

ratio test and extensive simulation study to determine the probability of correct selection for different sample sizes. Kundu and his team exploit the asymptotic property of $\mathcal{T}$ and determine the minimum sample size required for discriminating among different competing models.

5.6.2 Minimum Distance Criterion

It is natural to choose a particular model among competing models based on which of the models has a function *closest* to the empirical cumulative distribution function (ECDF) of a given dataset according to some distance measure between the two distribution functions.

Definition 5.6.1. Let $X_1, X_2, ..., X_n$ denote a random sample from a cumulative distribution function $F(.)$ and let $Y_1 \leq Y_2 \leq ... \leq Y_n$ denote the corresponding order statistics. The sample cumulative distribution function, denoted by $F_n(x)$ is defined by

$$F_n(x) = \frac{1}{n}(\text{number of } Y_j \leq x) \equiv \frac{1}{n}(\text{number of } X_i \leq x). \tag{5.6.2}$$

We then define the distance statistic as

$$\mathcal{D}_n = \sup_{-\infty<x<\infty} |F_n(x) - F(x)|, \tag{5.6.3}$$

where $\mathcal{D}_n$ is a random quantity that measures how far the empirical distribution function, $F_n(x)$, deviates from assumed distribution function $F(x) = \int_a^x f(y,\theta)\,dy$. Here $f(y,\theta)$ is the pdf of the order statistics. $\mathcal{D}_n$ is popularly called the Kolmogorov statistic and it is distribution-free in the sense that the critical values do not depend on the specific distribution being tested.

To implement this procedure, a candidate from each parametric family that has the smallest Kolmogorov-Sminorv (K-S) distance is identified and then the best fitted distributions compared. With a test of appropriate size, α , any hypothesis regarding the distributional form is rejected if the test statistic $\mathcal{D}_n$ is greater than the tabulated value, or, which is the same, if the $p-$value is lower than the significance level, α .

5.6.3 Akaike Information Criterion

Suppose X is a continuous random variable as defined in subsection 5.4.1 representing a model, say,

$$X = h(t, q) + \varepsilon \tag{5.6.4}$$

where h is a mathematical model such as a partial differential equation, probability density function, etc; ε is a random error term that is independent and identically distributed with a probability distribution such as the normal. In 1973 Hirotugu Akaike proposed in a seminal paper, Akaike (1973), a criterion for selecting a model from candidate models with equal data sample sizes. This criterion, known as the Akaike Information Criterion (AIC), is generally regarded as the first and still continues to be the most widely known model selection criterion because of its utilization of the relationship between the maximum likelihood and the Kullback-Leibler information. The motivation is that:

(i) The actual model is unknown to the researcher;

(ii) The parameter vector θ in g, say, must be estimated from the empirical data y generated from $f(x)$ which is a realization of X having specified dimension and structure;

(iii) For a parametric candidate model of interest, the likelihood function reflects the conformity of the model to the observed data. Thus selecting the fitted model that maximizes the empirical likelihood invariably leads to choosing the most complex model in the candidate collection;

(iv) The estimator $\hat{\theta}(y)$ of θ is a random variable and therefore the information $I\left(f, g\left(. \left| \hat{\theta}(y) \right.\right)\right)$ is also a random variable. The selection target is

$$\min_{g \in G} E_y \left[I\left(f, g\left(. \left| \hat{\theta}(y) \right.\right)\right)\right]. \tag{5.6.5}$$

In (5.6.5) G is a collection of admissible models (in terms of probability density functions), $\hat{\theta}$ MLE based on model g and data y, where y as stated earlier is the random sample from a density function $f(x)$. The criterion for selection is based on

$$\min_{g \in G} E_y E_x \left[\ln\left(g\left(x \left| \hat{\theta}(y) \right.\right)\right)\right]. \tag{5.6.6}$$

Hence an appropriately unbiased estimate of $E_y E_x \left[\ln\left(g\left(x\left|\hat{\theta}\left(y\right)\right.\right)\right)\right]$ for large sample and good model is

$$\ln\left(\mathcal{L}\left(\hat{\theta}\left|y\right.\right)\right) - k \tag{5.6.7}$$

where $\mathcal{L}$ is the likelihood function, $\hat{\theta}$ the MLE of θ and k the number of estimated parameters (including the variance) in the model. Here *good* model refers to the model $g(y)$ that is close to f in the sense of having the minimum AIC value.

Based on the AIC value we evaluate the following:

(i) The loss of information when a fitted model is used rather than the best approximating model is given by the AIC differences

$$\Delta_i = AIC_i - AIC_{\min}, \tag{5.6.8}$$

where AIC_{min} is AIC value for the best model in the set.

(ii) The likelihood of a model being useful in making inference concerning the relative strength of evidence for each of the models in the set is given by

$$\mathcal{L}\left(g_i\left|y\right.\right) \propto \exp\left(-\frac{1}{2}\Delta_i\right). \tag{5.6.9}$$

(iii) The *Akaike weight of evidence* in favour of model i being the best approximating model in the set is

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^{R}\exp\left(-\frac{1}{2}\Delta_r\right)} \tag{5.6.10}$$

where R is the total number of models in the set. Readers interested in AIC are referred to Akaike (1974), Çetin and Erar (2002), Bozdogan (1987, 2000) and Burnham and Anderson (2002) for details.

5.6.4 The Normality Hypothesis

The Shapiro-Wilk test, see for example, Shapiro and Wilk (1965) is one of the most powerful normality tests in statistics literature applied for testing log return series. Normality is tested

by matching two alternative variance estimates: a non-parametric estimator obtained by a linear combination of ordered sample and the usual parametric estimator. The weights, a_i , are available in many statistical tables. The Shapiro-Wilk statistic, $\mathcal{W}$, is defined by

$$\mathcal{W} = \frac{\left(\sum_{i=1}^{n} a_i \left(x_{(i)}\right)\right)^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}, \tag{5.6.11}$$

where the $x_{(i)}$ are the ordered sample values ($x_{(1)}$ is the smallest) and the a_i are constants generated from the means, variances and covariances of the order statistics of a sample of size n from a normal distribution (see Pearson and Hartley 1972, Table 15). It gives the value of the statistic $\mathcal{W}$ and the corresponding $p-$value which is compared to a specified significant level, α . The normality hypothesis is rejected if $p < \alpha$ implying that $\mathcal{W}$ lies in the critical region. We discuss possible skewness in a model because it is fundamental to mainstream financial modeling, portfolio investment decisions, and in many statistical testing procedures relating to asset returns. Skewness is defined as follows

$$\gamma_1 = \frac{\mu^3}{\mu_2^3}. \tag{5.6.12}$$

where $\mu^3 = \mathbb{E}(x_i - \mu)^3, \mathbb{E}$ is the expectation operator, μ is the mean of random return variable x_i and $\mu_2 \equiv \sigma^2$ is the variance. For a *normal distribution*, $\gamma_1 = 0$; otherwise, the distribution is asymmetric. Skewness is positive when the right hand tail is heavier and negative when the left hand tail is heavier.

There are at least four different alternate approaches to perform a significance test for skewness. The first alternative is to assume that the dataset is i.i.d. normally distributed and then apply the standard test for skewness (see Alles and Kling 1994). Second alternative is to adjust the level of significance to take into account the observed autocorrelation (if any) (see Alles and Kling 1994). The third way is to filter the autocorrelation out from the data and then apply the standard tests for skewness. The forth alternative is to test a wide range of distributions and consider the kurtosis and skewness together (see Badrinath and Chatterjee 1988, 1991). However, Töyli (2002) reports that there are relatively few studies considering the skewness in stock market data and most of the results are contradictory (see also Kon 1984, Fielitz and Rozell 1983).

An undisputable exception from the classical asset returns' normality assumption is that empirical return distributions indicate substantial excess kurtosis. A large positive value for kurtosis indicates that the tails of the distribution are longer (heavier) than those of a normal distribution, while a negative value indicates shorter tails (becoming like those of a box-shaped uniform distribution). It was Mandelbrot and Fama that first reported this fundamental deviation from normality (Mandelbrot 1963, Fama 1965). The kurtosis is defined as

$$\gamma_2 = \frac{\mu^4}{\mu_2^2} - 3 = k - 3 \tag{5.6.13}$$

where $\mu^4 = \mathbb{E}(x_i - \mu)^4, \mathbb{E}$ is the expectation operator,x_i and $\mu_2 \equiv \sigma^2$ is the variance. For a normal distribution, the value of k is three. When the $\gamma_1 > 0$ in (5.1.2) the distribution is referred to as *leptokurtic* and called *platykurtic* if $\gamma_1 < 0$.

5.7 Implementation, Simulation and Application

We study two energy datasets[5] in addition to the electricity dataset introduced in Chapter four. These three datasets are the U.S. Daily Electricity Prices for Pennsylvania State (PJMW) from January 01, 2002 to October 28, 2010 corresponding to 1,900 observations, the Weekly Nigeria Bonny Light (Crude Oil) Spot Price FOB (US Dollars per Barrel) from January 03, 1997 to November 05, 2010 corresponding to 721 observations and the Daily Natural Gas Futures Contract 1 (US Dollars per Million BTU) from January 13, 1994 to November 09, 2010 corresponding to 4214 daily observations. The following acronyms are used for the datasets: Penn03 and Penn04 for Electricity Prices standardized log returns $R(t) \in \mathbb{R}$, and volatility measure $G(t) \in \mathbf{R}^+$, respectively. Similarly we denote Crude Oil Prices by Bonny03 and Bonny04 and Natural Gas by NatGas03 and NatGas04. Implementation of these models are based on the R packages[6] "fBasics", "SuppDists", "MASS", "HyperbolicDist", etc, for the GIG class of distributions including the Weibull and Lognormal distributions. The R package "ghyp" was used for the class of GH distributions and its special cases which include the Variance-

[5]Source: Energy Information Administration (EIA)[Intercontinental Exchange(ICE)].

[6]http://www.r-project.org/; The HyperbolicDist Package by David Scott; The ghyp Package by Wolfgang Breymann and David Luethi; and The fBasics Package by Diethelm Wuertz.

Gamma (VG) and Hyperbolic Skewed Student's $t-$ (SSt) distributions. All these packages are available from the Comprehensive R Archive Network (CRAN). With special statistical functions in these packages we implement density, cumulative distribution functions, quantiles and random seed generation. Other functions implement simulation for maximum likelihood estimates (mle) (especially the Nelder and Mead algorithm) of parameters and tests making descriptive statistics of data and comparative study of various classes of models possible.

5.8 Empirical Results

In this section we present the estimation results of each subclass of distributions considered in Sections 5.4 and 5.5 using discussions of different criteria of section 5.6 above. The results in Table 5.8.1 indicate that NIG is the best model given the set of six candidate models ($R = 6$) for Penn03. It satisfied the selection conditions of section 5.6 having the least AIC value of 5003.911 with Akaike weight of evidence of 0.7239 (or 72%) for being the best fitting model. Although the second best GH has the highest LLH value of -2497.952, the estimates of model's five parameters as against four for NIG (with respect to equation (5.6.7)) raised its AIC value to 5005.903. The weight of evidence in its favour is 0.2674 (or 27%). Only NIG and GH out of the six competitors took up 99% weight of evidence for fitting the Penn03 dataset. To discriminate between NIG and GH or to what extent NIG is better than GH we resort to evidence ratio (ER) $w_{NIG}/w_{GH} = 0.7239/0.2674 = 2.71$ which shows that NIG is about three times better than the GH in fitting the dataset among other candidate models. Use of the rule of the thumb given in Burnham and Anderson (2002), a $\Delta_i < 2$ suggest substantial evidence for model i, values within the interval $3 \leq \Delta_i \leq 7$ indicate that the model has considerably less support whereas a $\Delta_i > 10$ indicates that the model is very unlikely to fit the data well. With this rule, GAUSS has no support. The value of Δ_{GAUSS} is 388.88 and with weight of evidence, w_{GAUSS}, of 0.0000 (or 0%).

From Table 5.8.2, SSt, NIG, HYP and VG have $\Delta_i < 2$ with $32\%, 28\%, 20\%$ and 13% weights of evidence respectively. The evidence ratio ER indicates that SSt is only 1.17 more likely to be better model relative to NIG and so any of the them (SSt and NIG) is a good fit with SSt

Table 5.8.1: Penn03 Model fits: $\bar{\alpha}$ parametrization

	Model					
Parameter	GH	HYP	NIG	SSt	VG	GAUSS
λ	-0.4639	1.0000	-0.5000	-	1.0270	-
$\bar{\alpha}$	0.5221	0.1590	0.5197	-	-	-
μ	-0.0568	-0.0717	-0.0563	-0.0451	-0.0693	-0.0000
ν	-	-	-	2.9926	-	-
σ	0.9980	0.9645	0.9988	1.0988	0.9692	1.0002
γ	0.0567	0.0717	0.0562	0.04989	0.0692	-
LLH	-2497.952	-2504.560	-2497.955	-2502.537	-2505.855	-2694.396
AIC	5005.903	5017.120	5003.911	5013.073	5019.710	5392.791
Δ_i	1.992	13.209	0.0000	9.164	15.799	388.880
w_i	0.2674	0.0010	0.7239	0.0074	0.0003	0.0000

NB: Tables (5.8.1)−(5.8.3) are results from $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$ parametrization while Tables (5.8.6)−(5.8.9) are from $(\lambda, \alpha, \mu, \Delta, \delta, \beta)$ parametrization.

being superior for the Bonny03 dataset. It has the highest LLH and has the least AIC value.

The model is, however, about two or three times more likely to be the best model than the HYP and VG respectively. GH has considerably less support by the data and the Gaussian is ruled out of contention as being very unlikely. Similarly, results in Table 5.8.3 indicate that SSt is the only model with $\Delta_i < 2$ and weight of evidence approximately 75% of all the weights of the contending models. GH is, however, in the far second position with $\Delta_i = 2.08(> 2)$ and with Akaike weight of 26%. SSt therefore has ER = 2.8291 about three times more likely to be the best model than the GH given the candidate models under consideration for NatGas02 dataset.

The models fitted for the volatility measure datasets are all two parameter models. From Table 5.8.4, the Weibull distribution has the highest log-likelihood value of -1166.496 and the least AIC value of 2336.992 with 100% weight of evidence against five other candidate models in the Penn03 dataset. This shows that the Penn03 dataset has heavy tail with the Weibull

Table 5.8.2: Bonny03 Model fits: $\bar{\alpha}$ parametrization

	Model					
Parameter	GH	HYP	NIG	SSt	VG	GAUSS
λ	0.2552	1.0	-0.5	2.9789	2.1248	-
$\bar{\alpha}$	1.6827	1.3753	1.6638	-	-	-
μ	0.3034	0.3250	0.2972	0.2519	0.3469	0.0000
ν	-	-	-	5.9578	-	-
σ	0.9739	0.9649	0.9687	0.9763	0.9622	1.0000
γ	-0.2967	-0.3252	-0.2972	-0.2526	-0.3469	-
LLH	-992.2024	-992.3098	-991.9710	-991.8174	-992.7348	-1021.1339
AIC	1994.405	1992.620	1991.942	1991.635	1993.470	2046.268
Δ_i	2.770	0.985	0.307	0.000	1.835	54.630
w_i	0.0803	0.1960	0.2750	0.3207	0.1281	0.0000

Table 5.8.3: NatGas04 Model fit: $\bar{\alpha}$ parametrization

	Model					
Parameter	GH	HYP	NIG	SSt	VG	GAUSS
λ	-1.8945	1.0000	-0.5000	-	1.4825	-
$\bar{\alpha}$	0.3397	0.6827	0.9053	-	-	-
μ	-0.0513	-0.0477	-0.0479	-0.0484	-0.0416	0.0000
ν	-	-	-	4.0538	-	-
σ	0.9925	0.9682	0.9822	1.0032	0.9686	1.0000
γ	0.0494	0.0472	0.0478	0.0485	0.0415	-
LLH	-5652.951	-5674.559	-5660.107	-5652.910	-5682.775	-5976.288
AIC	11315.90	11357.12	11328.21	11313.82	11373.55	11956.58
Δ_i	2.08	43.30	14.39	0.00	59.73	642.76
w_i	0.2610	0.0000	0.0006	0.7384	0.0000	0.0000

Table 5.8.4: Model fits for Penn03 (Standard errors are in parenthesis)

	Normal	LN	Weibull	Gamma	IGam	IG
Shape	0.7140	-0.0357	0.9972	1.3116	1.9931	0.8974
	(0.0173)	(0.0228)	(0.0178)	(0.0383)	(0.0600)	(0.0172)
Scale	0.7520	0.9928	0.6791	0.9029	12.9675	1.2849
	(0.0122)	(0.0161)	(0.0164)	(0.0320)	(0.4439)	(0.0417)
Skew	-0.0041	9.3145	2.0572	1.8168	7.7504	2.4599
Kurt	-0.0028	183.2755	5.6531	4.7672	93.9212	10.5244
LLH	-2153.225	-2613.029	-1166.496	-2568.031	-3464.596	-1382.205
AIC	4310.449	5230.058	2336.992	5140.062	6925.191	2768.411
w_i	0	0	100	0	0	0

distribution fitting it much better than the other models under reference (see Figure 5.8.1). Similar inference is deduced from Table 5.8.5 for Bonny03 dataset where the weight of evidence is strongly in support for Weibull distribution as the best model in this dataset and the Gamma distribution is the best model for NatGas03 dataset (Table 5.8.6) given our criteria for selection.

Visual inspection of Figure 5.8.1 shows that tail performances of GH and NIG are outstanding (being closest to the data points) with respect to other distributions. The slight edge NIG has over GH here is that while NIG is fitted with four parameters, the GH is fitted with five. The SSt overestimated in both tails while HYP and VG underestimated the tails. The Gaussian (Normal) is far from fitting the data. An interesting observation from these plots is that the tail plots show almost linear behaviour especially in the left tail. The implication of this linear tail plot is that Penn03 dataset is drawn from a power-law distribution. In Figure 5.8.2 we observe that the five models are close in fitting the data especially in the left tail better than in the right tail. This is why selection among them is difficult, as we saw in their numerical values (Tables 5.8.2 & 5.8.8) and in addition, GAUSS has no support. However, SSt leads in the left tail but neck-neck with GH in the right tail. Similarly, SSt is superior in fitting NatGas dataset in Figure 5.8.3. A close observation of the shape of the tail plots show polynomial behaviour.

Table 5.8.5: Model fits for Bonny03 (Standard errors are in parenthesis)

	Normal	LN	Weibull	Gamma	IGam	IG
Shape	0.7452	1.3248	1.1613	1.6900	2.1110	1.3248
	(0.0249)	(0.0450)	(0.0338)	(0.0504)	(0.0639)	(0.0450)
Scale	0.6684	1.5976	0.7745	1.2482	6.0166	1.5976
	(0.0176)	(0.0842)	(0.0262)	(0.0432)	(0.2053)	(0.0842)
Skew	-0.1020	4.4738	1.5216	1.7007	6.6733	2.1042
Kurt	0.0086	32.6414	3.0268	5.6003	68.5634	5.40726
LLH	-731.583	-1024.527	-485.7715	-2339.715	-2208.712	-834.2527
AIC	1467.166	2053.054	975.543	4683.430	4413.424	1672.505
w_i	0	0	100	0	0	0

Table 5.8.6: Model fits for NatGas04 (Standard errors are in parenthesis)

	Normal	LN	Weibull	Gamma	IGam	IG
Shape	0.7106	-0.0187	1.1464	1.4023	6.0218	1.0325
	(0.0106)	(0.0155)	(0.0137)	(0.0412)	(0.1383)	(0.0131)
Scale	0.6908	1.0039	0.7885	0.9967	2.0214	1.5164
	(0.0075)	(0.0109)	(0.0112)	(0.0112)	(0.0409)	(0.0330)
Skew	0.0270	4.9991	1.9265	1.6450	25.8757	0
Kurt	0.327	44.0049	6.9472	4.3060	910.7773	0
LLH	-4419.797	-5915.751	-2944.366	-2486.86	-4565.691	-3604.112
AIC	8843.593	11835.50	5892.732	4977.72	9127.382	7212.225
w_i	0	0	0	100	0	0

Table 5.8.7: Penn03 parameter estimation and Model selection(alpha parametrization)

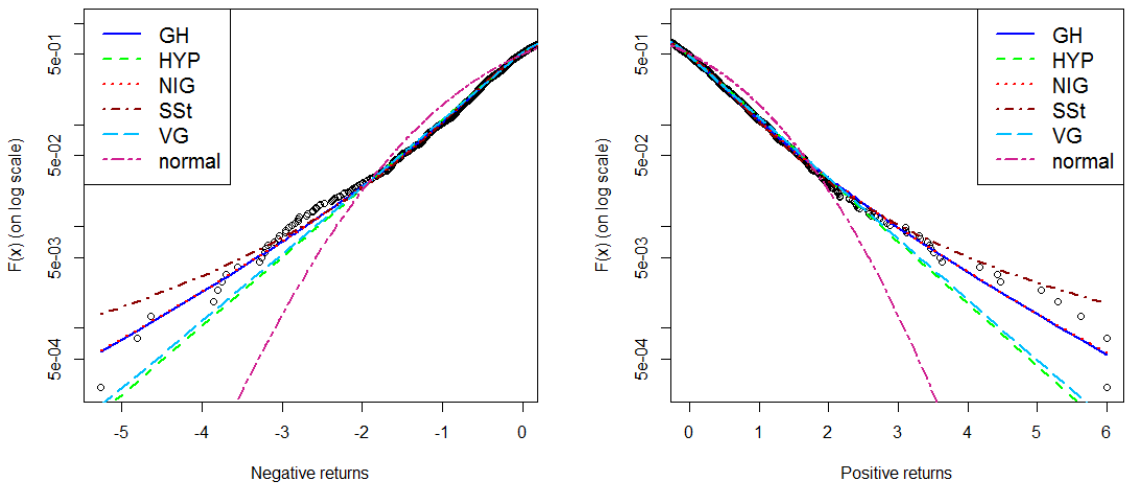
	Model					
Parameter	GH	HYP	NIG	SSt	VG	GAUSS
α	0.7418	1.4871	0.7239	-	-	-
β	0.0570	0.0771	0.0566	-	-	-
δ	0.7061	0.1072	0.7199	-	-	-
μ	-0.05674	-0.0717	-0.0564	-0.0451	-0.0693	0.0000
σ	-	-	-	1.0988	0.9692	1.0003
γ	-	-	-	0.0412	0.0692	
λ	-0.4631	-	-	2.9926	1.0275	-
LLH	-2497.952	-2504.560	-2497.624	-2502.537	-2505.855	-2694.396

Table 5.8.8: Bonny03 parameter estimation and Model selection(alpha parametrization)

	Model					
Parameter	GH	HYP	NIG	SSt	VG	GAUSS
α	0.6472	1.8404	1.3686	-	-	-
β	-0.2789	-0.3492	-0.3170	-	-	-
δ	1.7764	0.7600	1.2493	-	-	-
μ	0.2642	0.3249	0.2975	0.2519	0.3469	0.0000
σ	-	-	-	0.9763	0.9622	1.0000
γ	-	-	-	-0.2526	-0.369	-
λ	-2.3810	-	-	5.9578	2.1248	-
LLH	-991.742	-992.310	-991.971	-991.817	-992.735	-1021.134

Table 5.8.9: NatGas03 parameter estimation and Model selection(alpha parametrization)

	Model					
Parameter	GH	HYP	NIG	SSt	VG	GAUSS
α	0.1490	1.6102	0.9699	-	-	-
β	0.0490	0.0507	0.0494	-	-	-
δ	1.4182	0.4255	0.9344	-	-	-
μ	-0.0489	-0.0477	-0.0478	-0.0484	-0.0416	-0.0002
σ	-	-	-	1.0032	0.9686	1.0001
γ	-	-	-	0.0485	0.0485	-
λ	-1.9717	-	-	4.0538	1.4825	-
LLH	-5652.849	-5674.558	-5660.107	-5652.910	-5682.775	-5976.288



h!

Figure 5.8.1: Tail plots for Penn03: Left tail (left panel) and Right tail (right panel)

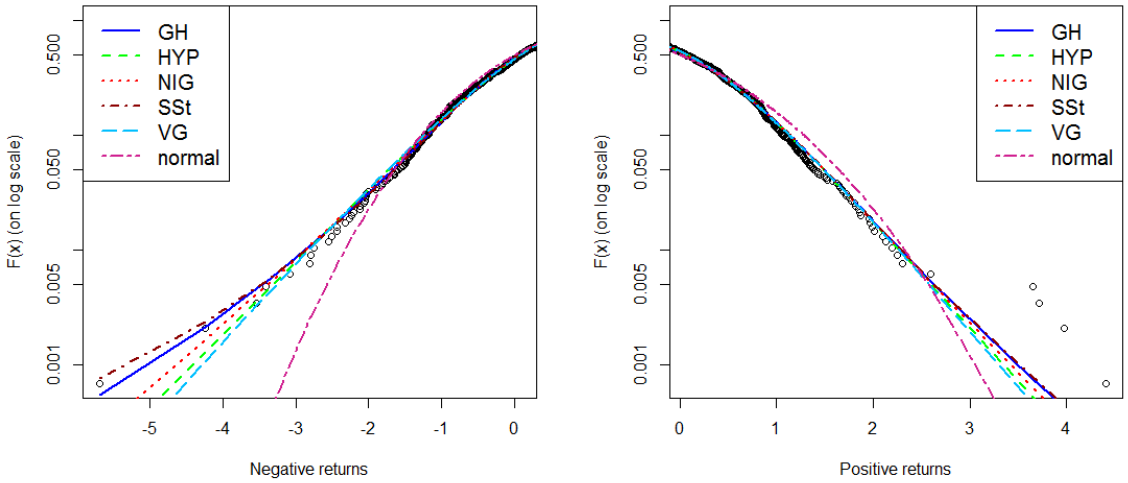


Figure 5.8.2: Tail plots for Bonny03: Left tail (left panel) and Right tail (right panel)

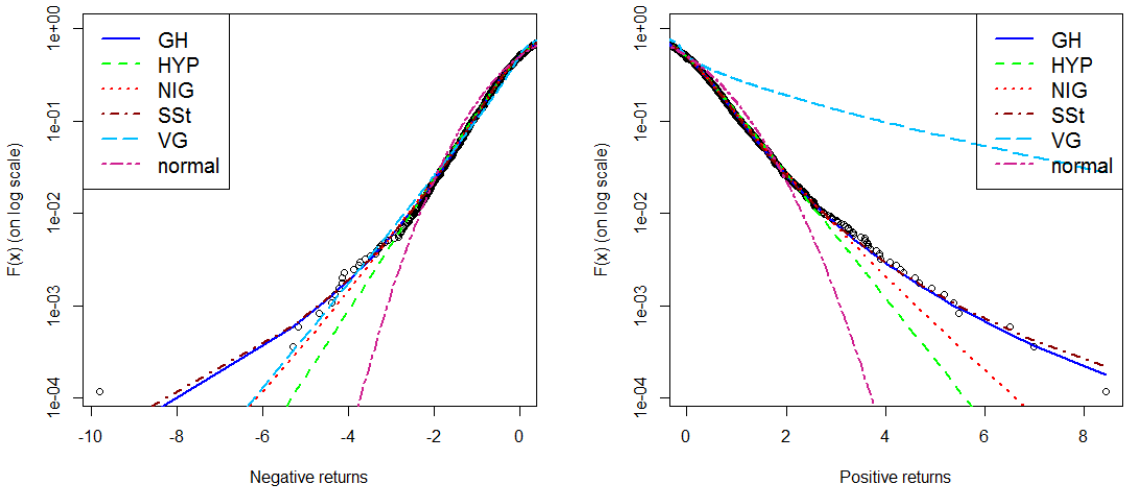


Figure 5.8.3: Tail plots for NatGas03: Left tail (left panel) and Right tail (right panel)

5.9 Chapter Summary

In order to analyse prices of energy futures in this Chapter we assumed that the log-return series of the prices are driven by Levy process of the generalized hyperbolic (GH) type. We compared five members of the GH family (the generalized hyperbolic (GH), hyperbolic (HYP), normal inverse Gaussian (NIG), variance gamma (VG) and hyperbolic skewed Student t (SSt) distributions) along with the normal distribution as the benchmark. This comparison was performed for the three datasets in the energy sector namely; electricity futures prices, crude oil prices and natural gas prices. We present in Table 5.9.1 a summary of the outcome when these distributions were fitted to the datasets. Using Akaike information criteria (AIC) and the log likelihood (LLH) criteria Table 5.9.1 shows that NIG and GH controls 99% weight of evidence for being best among the six candidate probability distribution functions in the family with NIG being exceptional for Pennsylvania dataset. The fit in Bonny shows a "kin" contest in which the best two had only 60% weight of evidence followed closely by HYP (20%) and VG (13%). Although SSt performed well, any of SSt, NIG and HYP is good enough to fit the Bonny crude oil dataset. The SSt is outstanding in fitting Natural Gas dataset and is recommended accordingly. The result for Bonny is similar to that of Krichene (2008) who fitted the NIG to his crude oil dataset (2000−2007) segmented into two, see for instance, summary results of parameter estimates in Table 5.9.2. Krichene shows that "NIG process fits closely oil price returns" during the period of investigation.

Table 5.9.1: Performance table for best distributions fitted to each series

	Pennsylvania Electricity	Bonny Crude Oil	Natural Gas
Distributions selected	NIG (72%)	SSt (32%)	SSt (74%)
	GH (27%)	NIG (28%)	GH (26%)

Enlarging the scope of models to five in the GH family increases our degrees of freedom and makes our choice superior. The NIG distribution has two tails that behave differently, but they are both semiheavy. One would therefore expect NIG to model skewness rather well, but only in cases where the tails are not too heavy. The SSt, on the other hand, is a distribution that is

Table 5.9.2: Our result for Bonny NIG compared to Krichene's (2008)

Normal Inverse Gaussian (NIG) Parameters					
	Lambda	Alpha.bar	Mu	Sigma	Beta
Our result	-0.5	1.6638	0.2972	0.9687	-0.2972
Krichene 2000m1-2003m4	-0.5	1.46	0.08	2.22	-0.08
Krichene 2003m5-2005m10	-0.5	2.68	0.29	1.69	-0.17

good in modeling skewness and heavy-tailed data. This explains why these two distributions dominated others in fitting the three datasets.

In modeling the volatility process for the three datasets we compared five probability density functions in the extreme value distribution family (the Weibull, lognormal, gamma, inverse gamma and the inverse Gaussian distributions) along with the normal distribution. With similar argument as presented above, the two-parameter Weibull density function is recommended for volatility in Pennsylvania electricity futures prices and Bonny light crude oil while the gamma density function is recommended for natural gas dataset. These results show the inability of the Gaussian process to fit high frequency data as underscored by Mandelbrot (1963) and Fama (1965) in which both authors proposed stable distributions for modeling skewness and kurtosis. The high kurtosis in the electricity returns series of Chapter Four is hereby addressed. The main attraction to the GH distributions is that they are constructed as mixtures of variance-mean normal distributions with time varying stochastic variance.

Chapter 6

Conclusion

> *We have not succeeded in answering all our problems. The answers we have found only serve to raise a whole lot of new questions. In some ways we feel we are as confused as ever, but we believe we are confused on a higher level and about more important things.* - B. Øksendal (2000)

This Thesis is in two parts: mathematical and statistical– with each part comprising two chapters. In Part One we discussed the theory and analysis of partial differential equations using the Lie symmetry technique to analyse an evolution partial differential equation arising from financial mathematics, see for instance, equation (3.1.1) of Chapter Three. The second part concerns applications to real life problems where calibrations and statistical goodness-of-fit tests were performed.

A formula (proposition 2.3.1) for the nth prolongation of a generator Γ with k independent and p dependent variables of an nth-order partial differential equation is proposed and we claim to have extended the result (equation (2.3.6)) derived by Mahomed and Leach in 1990. The basic problem in the modeling of physical and other phenomena is to find solutions of differential equations. Many methods of solution of differential equations use a change of variables that transforms a given differential equation into another equation with known properties. We constructed a transformation that maps symmetries of our PDE invertibly into the heat equation which is a well studied equation with appealing characteristics. As a result

of the symmetry analysis performed we also show that our given partial differential equation admits a finite number of Lie point symmetries characterized by the six-dimensional algebra isomorphic to $\{sl(2,R) \oplus W_3\} \oplus_s \infty A_1$, with one solution symmetry where the subalgebra is of the Heisenberg-Weyl type. Two general solutions calculated from the twelve optimal systems of invariant solutions are given in equations (3.3.40) and (3.3.49). It is our thinking that these equations will one day be found useful for practical applications.

The complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with natural filtration $\{\mathcal{F}_k\}_1^\infty$ and Levy proceses L_t were assumed. We propose a dynamic linear model (DLM) with switching regimes for modeling the stochastic volatility of log return series for prices of electricity contracts. A modified Kalman filter algorithm was introduced to fit the regime-switching Markov model and estimation of the parameters using quasi-maximum likelihood method were performed. Results displayed in Table 4.7.1 are comparable to results obtained by Kellerhals (2004) using affine structure models for spot and futures prices, and Krichene (2008) for crude oil prices using GARCH(1,1) models. It will be of interest to compare our model with models used by Kellerhals and Krichene using our dataset.

Two observations were immediate. The first is that both small and large changes come clustered, i.e., there are periods of low and high volatility. The second is that, from time to time, we observe rather large changes which may be hard to reconcile with the standard distributional assumption in statistics and econometrics, that is, normality. From empirical results the dataset exhibited volatility clustering followed by mean reversion with half-life of nine months. This informed our use of Gaussian mixtures in the model. The mixing of Gaussian distributions is well suited for financial modeling, as it allows for the construction of very flexible distributions. This fact is demonstrated in Chapter Five, where the normal-mean-variance mixture, on which the generalized hyperbolic distribution (GH) of Barndorff-Nielsen (1977) is based, generally exhibits heavier tails than the Gaussian distribution. We used this to great advantage. Interestingly we derived our DLM based on this idea and the generalization of the Vasicek and CIR models (or for some authors, the general Heston model) governed by the stochastic differential equation (4.3.12). The adequacy of our model was authenticated by the preliminary study of the dataset that displayed evidence of first-order autocorrelation showing that the state variable

is a first-order autoregressive AR(1) process.

The major concern in Chapter Five was the identification of the probability distribution of the process that generated the dataset. From each of the three datasets (Daily Pennsylvania Electricity Future Contract, Weekly Bonny Crude oil Spot prices and Daily Natural Gas Prices) we generated and studied two concomitant variables: log return series and volatility series. Five probability density functions of the generalized hyperbolic family (Generalized Hyperbolic (GH), Hyperbolic (HYP), Normal Inverse Gaussian (NIG), Variance-Gamma (VG), and Skew Student-t (SSt)) and five from the extreme value family (Weibull, Gamma, Lognormal, Inverse Gaussian and Inverse gamma) were fitted to the datasets and compared with the Normal distribution as the benchmark.

We established that energy return series is fat-tailed and with significant kurtosis. The normal distribution showed very poor fit in both series. Using the Akaike Information (AIC) and the Log-likelihood (LLH) criteria, we conclude that NIG (which is a mixture of the normal and the inverse Gaussian distributions) is best suited to fit and for prediction of prices for Pennsylvania electricity future contracts. This model performed well in fitting the crude oil dataset but ranked second only to SSt. The SSt (which also has a convolution property) dominated other five candidate models (74% dominance) in the natural gas dataset. With this result we posit that SSt is good for fitting oil and gas datasets while NIG is the choice for electricity series, see Table 6.0.1. These results are not surprising. The SSt has one heavy and one semi-heavy tail, i.e., one tail determined by a polynomial and the other by an exponential behaviour. The normal inverse Gaussian process L_t is a Levy process where increments in L_t are distributed according to the NIG distribution. Another appeal of the NIG distribution is that it is characterised by the first four moments(mean,variance, skewness and kurtosis). These are the moments we care about for inference in real life applications including risk management and derivative pricing. In fitting stochastic volatility series, the Weibull distribution performed wonderfully well in both the electricity and crude oil datasets while the gamma distribution is recommended for natural gas volatility series. The gamma process can be expressed as a limiting case of the generalized inverse Gaussian(GIG) process with $\lambda = -0.5$ (similar to the NIG). The Weibull on the other hand is popular in the analysis of lifetime data.

Table 6.0.1: Performance summary of different models

	Recommended Models	
Dataset	Return Series	Volatility Series
Electricity	NIG (72%) GH (27%)	Weibull
Crude Oil	SSt (32%) NIG (28%)	Weibull
Natural Gas	SSt (74%) GH (26%)	Gamma

Contrary to the assumption of "all things being equal" there are very high probability of higher or lower energy prices than previously expected over time. This suggests that compared to the normal distribution, the actual probability distribution of return and volatility series are fat-tailed, implying that the probability of large differences in prices of energy contracts is much higher than would be implied by time-invariant unconditional Gaussian distribution.

Bibliography

[1] Aas, K. and Haff, I. H. (2006) The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics*, **4** (2), 275–309.

[2] Abraham-Shrauner, B. and Guo, A. (1992) Hidden symmetries associated with the projective group of nonlinear first-order ordinary differential equations. *Journal of Physics, A: Mathematical and General*, **25**, 5597–5608.

[3] Abraham-Shrauner, B. and Guo, A. (1993) Hidden and nonlocal symmetries of nonlinear differential equations. In *Modern Group Analysis: Advanced Analytical and Computational Methods in Mathematical Physics.* Ibragimov, N. H., Torrissi, M. and Valenti, A., editors, p. 1–5. Kluwer Academics, Dordrencht.

[4] Abraham-Shrauner, B. and Guo, A. (1994) Hidden symmetries of differential equations. *Contemporary Mathematics,* **160**, 1–13.

[5] Abraham-Shrauner, B. and Govinder, K. S. (2006) Provenance of Type II hidden symmetries from nonlinear partial differential equations. *Journal of Nonlinear Mathematical Physics,* **13**, 612–622.

[6] Abraham-Shrauner, B., Govinder, K, S. and Leach, P. G. L. (1995) Integration of second-order equations not possessing point symmetries. *Physics Letters A*, **203**, 169–174.

[7] Abramowitz, M. and Stegun, I. A. (1972) *Handbook of Mathematical Functions.* Dover Publications, New York.

[8] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, eds., 2nd International Symposium on Information Theory. Akademia Kiado, Budapast, 267–281.

[9] Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* **19**, 716–723.

[10] Albrecher, H. , Mayer, P., Schoutens, W. and Tistaert, J. (2007) The little Heston trap. *WILMOT,* 83–92.

[11] Alles, L. A. and Kling, J. L. (1994) Regularities in the variation of skewness in asset returns. *The Journal of Financial Research,* **17**, 427–438.

[12] Amaral, L.A.N., Plerou, V., Gopikrishnan, P., Meyer, M. and Stanley, H. E. (2000) The distribution of returns of stock prices. *International Journal of Theoretical and Applied Finance,* **3** (3), 365–369.

[13] Anco, S. C. and Bluman, B. W. (1996) Derivation of conservation laws from nonlocal symmetries of differential equations. *Journal of Mathematical Physics,* **37**, 2361–2375.

[14] Anderson, B. O. D. and Moore, J. B. (1979) *Optimal filtering.* Prentice Hall Inc., New Jersey.

[15] Andriopoulos, K. (2008) *Classification of Differential Equations by Group Theoretical Methods.* PhD Thesis, School of Mathematical Sciences, University of KwaZulu-Natal, Durban, Republic of South Africa.

[16] Arrigo, D. J., Broadbridge, P. and Hill, J. M. (1993) Nonclassical symmetries and the methods of Bluman-Cole and Clarkson-Kruskal. *Journal of Mathematical Physics*, **34**, 4692–4703.

[17] Arrigo, D. J., Hill, J. M. and Broadbridge, P. (1994) Nonclassical symmetry reductions of the diffusion equation with nonlinear source. *IMA Journal of Applied Mathematics*, **52**, 1–24.

[18] Atkinson, A. C. (1987) *Plots, Transformations, and Regression.* Clarendon Press, London.

[19] Azzalini, A. and Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B*, **65**, 579–602.

[20] Babs, S. H. and Nowman, K. B. (1999) Kalman filtering of generalized Vasicek term structure models. *Journal of Financial and Quantitative Analysis*, **34**(1), 115–130.

[21] Badrinath, S. G. and Chatterjee, S. (1988) On measuring skewness and elongation in common stock return distributions: the case of the market index. *Journal of Business*, **61**(4), 451–472.

[22] Badrinath, S. G. and Chatterjee, S. (1991) A data-analytic look at skewness and elongation in common-stock-return distributions. *Journal of Business and Economic Statistics*, **9**(2), 223–233.

[23] Ball, C. A. and Torous, W. N. (1999) The stochastic volatility of short-term interest rates: some international evidence. *Journal of Finance*, **54**(6), 2339–2359.

[24] Barndorff-Nielsen, O. E. (1977) Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London, A*, **353**, 401–419.

[25] Barndorff-Nielsen, O. E (1994) Gaussian-inverse Gaussian process and the modeling of stock returns. Paper presented at the second Workshop on *Stochastics and Finance* 1994 in Berlin.

[26] Barndorff-Nielsen, O. E. and Prause, K. (2001) Apparent scaling. *Finance and Stochastics.* **5**, 103–113.

[27] Barndorff-Nielsen, O. E., Kent, J. and Sorensen, M. (1982) Normal variance-mean mixtures and z-distributions. *International Statistical Review*, **50**, 145–159.

[28] Barndorff-Nielsen, O. E and Stelzer, R. (2005) Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Lévy processes. *Scandinavian Journal of Statistics*, **32**, 617–637.

[29] Benth, F. E. and Saltyte-Benth, J. (2004) The normal inverse Gaussian distribution and spot price modeling in energy markets. *International Journal of Theoretical and Applied Finance*, **7**(2), 177–192.

[30] Birkhoff, G. D. (1933) *Aesthetic Measure*. Harvard University Press, Cambridge.

[31] Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 631–659.

[32] Bluman, G. W. (1980) On the transformation of diffusion processes into the Wiener process. *SIAM Journal of Applied Mathematics*, **39**, 238–247.

[33] Bluman, G. W. (1983) On mapping linear partial differential equations to constant coefficient equations. *SIAM Journal of Applied Mathematics*, **43**(6), 1259–1273.

[34] Bluman, G. W. and Anco, S. C. (2002) *Symmetry and Integration Methods for Differential Equations.* Applied Mathematical Sciences no. 154. Springer-Verlag, New York.

[35] Bluman, G. W. and Cole, J. D. (1974) *Similarity Methods for Differential Equations*. Springer, New York.

[36] Bluman, G. W. and Kumei, S. (1989) *Symmetries and Differential Equations*. Applied Mathematical Sciences no. 81, Springer-Verlag, New York.

[37] Bølviken, E. and Benth, F. E. (2000) Quantification of risk in Norwegian stocks via the normal inverse Gaussian distribution. In Proceedings of the AFIR 2000 Colloquium, Troms, Norway, 87–98.

[38] Bozdogan, H. (1987) Model selection and Akaike information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, **52**, 345–370.

[39] Bozdogan, H. (2000) Akaikes information criterion and recent developments in the information complexity. *Journal of Mathematical Psychology*, **44**, 62–91.

[40] Branco, M. D. and Dey, D. K. (2001) A general class of multivariate skew-elliptical distributions.*Journal of Multivariate Analysis*, **79**, 99–113.

[41] Breymann, W. and Luethi, D. (2008) The ghyp: a package on generalized hyperbolic distributions.

[42] Burnham, K. P. and Anderson, D. R. (2002) *Model Selection and Inference: a Practical Information-Theoretic Approach*. Second Edition, Springer-Verlag, New York.

[43] Cantwell, B. J. (2002) *Introduction to Symmetry Analysis*. Cambridge University Press, Cambridge.

[44] Çetin, M. C. and Erar, A. (2002) Variable selection with Akaike information criterion: a comparative study. *Hacettepe Journal of Mathematics Statistics*. **31**, 89–97.

[45] Champagne, B., Hereman, W. and Winternitz, P. (1991) The computer calculation of Lie point symmetries of large systems of differential equations. *Computer Physics Communications*, **66**, 319–340.

[46] Chan, K. C., Karolyi, A., Longstaff, F. and Sanders, A. (1992) An empirical comparison of alternative models of the short term interest rate. *Journal of Finance*, **47**, 1209–1227.

[47] Chen, R. R. and Scott, L. (2003) Multi-factor Cox-Ingersoll-Ross models of term structure: estimation and tests from a Kalman filter model. *Journal of Real Estate Finance and Economics*, **27**(2), 143–172.

[48] Chetterjee, R. (2003) *Application of the Kalman filter for estimating continuous time term structure models: evidence from the UK and Germany*. Working paper, Department of Economics, University of Glasgow.

[49] Chhikara, R. S. and Folks, J. L. (1989) *The Inverse Gaussian Distribution*. Marcel Dekker, New York.

[50] Chou, K. S and Li, G. X. (2001) A note on optimal system for the heat equation. *Journal of Mathematical Analysis and Applications*, **261**, 741–751.

[51] Clarkson, P. A. (1989) New similarity solutions of the modified Boussinesq equation. *Journal of Physics A: Mathematical and General*, **22**, 2355–2367.

[52] Clarkson, P. A. (1989) New similarity solutions and Painlevé analysis for the symmetric regularized long wave and the modified Benjamin-Bona-Mahoney equations. *Journal of Physics A: Mathematical and General*, **22**, 3821–3848.

[53] Clarkson, P. A. (1995) Nonclassical symmetry reductions of the Boussinesq equation. *Chaos, Solitons and Fractals*, **5**, 2261–2301.

[54] Clarkson, P. A. and Kruskal, M. D. (1989) New similarity solutions of the Boussinesq equation, *Journal of Mathematical Physics* **30**, 2201–2213.

[55] Clewlow, L. and Strickland, C. (2000) *Energy Derivatives: Pricing and Risk Management*. Lacima Publications, London.

[56] Cont, R. (2001) Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, **1**, 223–236.

[57] Cook, R. D. and Weisberg, S. (1982) *Residuals and Influence in Regression*. Chapman and Hall, London.

[58] Cox, D. R. (1962) Further results on tests of separate families of hypothesis. *Journal of the Royal Statistical Society, Series B*, **24**, 406–424.

[59] Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985) The theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.

[60] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via EM algorithm (with discussions). *Journal of the Royal Statistical Society, Series B*, **35**, 1–38.

[61] Derman, C., Gleser, J. and Olkin, I. (1973) *A Guide to Probability Theory and Applications.* Holt, Rinehart and Winston Inc., New York.

[62] Dimas, S. and Tsoubelis, D. (2005) SYM: A new symmetry-finding package for Mathematica. In: N. H. Ibragimov, C. Sophocleous and P. A. Damianou (eds), *Group Analysis of Differential Equations*, p. 64–70. University of Cyprus, Nicosia.

[63] Dimas, S. and Tsoubelis, D. (2006) A new Mathematica-based program for solving overdetermined systems of PDEs. 8th International Mathematical Symposium (Avignon, France).

[64] Dimas, S., Andriopoulos, K., Tsoubelis, D. and Leach, P.G.L. (2009) Compete specification of some partial differential equations that arise in financial mathematics. *Journal of Nonlinear Mathematical Physics*, **16** (Supplement), 73–92.

[65] Dineen, S. (2005) *Probability Theory in Finance: a Mathematical Guide to Black-Scholes Formula.* Graduate Studies in Mathematics, Volume **70**. American Mathematical Society, Providence.

[66] Do, B. (2008) *Estimating stochastic volatility option pricing models with Kalman filtering.* Working Paper. Department of Accounting and Finance, Monash University, Austrialia.

[67] Drapper, N. R. and Smith, H. (2005) *Applied Regression Analysis.* Fourth edition, John Wiley, New York.

[68] Dresner, L. (1999) *Applications of Lie Theory of Ordinary and Partial Differential Equations.* IOP Publishing Ltd, Philadelphia.

[69] Duertz, D. (2007) The *f*Basics Package

[70] Eberlin, E. and Keller, U. (1995) Hyperbolic distributions in finance. *Bernoulli*, **1**(3), 281–299.

[71] Etheridge, A. (2002) *A Course in Financial Calculus.* Cambridge University Press, Cambridge.

[72] Eydeland, A. and Wolyniec, K. (2003) *Energy and Power Risk Management.* John Wiley and Sons, New York.

[73] Fama, E. (1963) Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, **36**, 420–429.

[74] Fama, E. (1965) The behaviour of stock market prices. *Journal of Business.* **38**, 34–105.

[75] Fernandez, C. and Steel, M. (1998) On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359–371.

[76] Fearn, D. H. and Nebenzahl, E. (1991) On the maximum likelihood ratio method of deciding between Weibull and gamma distributions. *Communications in Statistics, Theory and Methods*, **20**(2), 579–593.

[77] Feller, W. (1957) *An Introduction to Probability Theory and its Applications.* Volume I, 2nd edition. John Wiley and Sons, Inc., New York.

[78] Fielitz, B. and Rozell, J. (1983) Stable distributions and mixture of distributions hypotheses for common stock returns. *Journal of the American Statistical Association*, **69**, 108–113.

[79] Fushchych, W. I. and Kornyak, V. V. (2001) Computer algebra application for determining Lie and Lie-Backlund symmetries of differential equations. *W. I. Fushchych Scientific Works*, **3**, 497–505.

[80] Gazizov, R. K. and Ibragimov, N. H. (1998) Lie symmetry analysis of differential equations in finance. *Nonlinear Dynamics*, **17**, 387–407.

[81] Geyer, A. L. J. and Pichler, S. (1999) A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure of interest rates. *Journal of Financial Research*, **22**(1), 107–130.

[82] Goard, J. (2000) New solutions to the bond-pricing equation via Lie's classical method. *Mathematical and Computer Modeling,* **32**, 299–313.

[83] Goldfield, S. M. and Quandt, R. E. (1973) A Markov model for switching regressions. *Journal of Econometrics*, **1**, 3–16.

[84] Govinder, K. S. and Leach, P.G.L. (1996) The nature and use of symmetries of ordinary differential equations. *South African Journal of Science*, **92**, 23–28.

[85] Guillaume, D. M., Dacorogna, M. M., Davé , R. D., Müler, U. A., Oslen, R. B., and Pictet, O. V. (1997) From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, **1**, 95–129.

[86] Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357–384.

[87] Hamilton, J. D. (1994) *Time Series Analysis*. Princeton University Press, Princeton.

[88] Hamilton, J. D. (2005) *Regime-switching models*. Prepared for Palgrave Dictionary of Economics. Department of Economics, University of California, San Diego.

[89] Hansen, B. (1994) Autoregressive conditional density estimation. *International Economic Review*, **35**, 705–730.

[90] Harvey, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

[91] Harvey, A. C., Ruiz, E. and Shephard, N. (1994) Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247–264.

[92] Head, A. K. (1993) LIE, a PC program for Lie analysis of differential equations. *Computer Physics Communications.* **71**, 241–248.

[93] Head, A.K. (1996) LIE, a PC program for Lie analysis of differential equations (updated version). *Computer Physics Communications.* **76**, 311–313.

[94] Hereman, W. (1994) Review of symbolic software for the computation of Lie symmetries of differential equations. *Euromathematics Bulletin*, **1**(1), 45–79.

[95] Hereman, W. (1996) Symbolic software for the Lie analysis. In: N. H. Ibragimov (ed), *CRC Handbook of Lie Group Analysis of Differential Equations Vol. 3: New Trends in Theoretical Development and Computational Methods*, p. 367–413. CRC Press, Boca Raton.

[96] Hereman, W. and Nuseir, A. (1997) Symbolic methods to construct exact solutions of nonlinear partial differential equations. *Mathematics and Computers in Simulation*, **43**, 13–27.

[97] Heston, S. L. (1993) A closed-form solution for options with stochastic volatility, with applications to bond and currency options. *Review of Financial Studies*, **6**, 327–343.

[98] Hull, J. (2002) *Options, Futures, and Other Derivatives.* 5th Edition, Prentice Hall, New Jersey.

[99] Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities. *Journal of Finance*, **42**, 281–300.

[100] Hull, J. and White, A. (1990) Pricing interest rate derivatives. *Review of financial Studies*, **3**, 573–592.

[101] Hurst, S. R., Platen, E., and Rachev, S. T. (1995) Option pricing for asset returns driven by subordinated processes. Working Paper, The Australian National University, Canberra.

[102] Hydon, P. E. (2000) *Symmetry Methods for Differential Equations: A Beginner's Guide.* Cambridge University Press, Cambridge.

[103] Ibragimov, N. H. (1995) Galilian principle in thermal diffusion. *Symmetri*, **2**(3), 56–62.

[104] Ibragimov, N. H. (1999) *Elementary Lie Group Analysis and Ordinary Differential Equations.* John Wiley, Chichester.

[105] Ibragimov, N. H. and Gazizov, R. K. (1998) Lie symmetry analysis of differential equations in finance. *Nonlinear Dynamics*, **17**, 387–407.

[106] Ibragimov, N. H., Torrisi, M., Valenti, A. (1991) Preliminary group classification of equations $u_{tt} = f(x, \nu_x) + g(x, \nu_x)$. *Journal of Mathematical Physics*, **32**(11), 2988–2995.

[107] Ibragimov, N. H. (2009) *Selected Works IV.* ALGA Publications, Karlskrona. Paper 6.

[108] Itô, K. (1951) On stochastic differential equations. *Memoirs of the American Mathematical Society*, **4**, 1–51.

[109] Jazwinski, A. H. (1972) *Stochastic Processes and Filtering Theory.* Academic Press, New York.

[110] Jørgensen, B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution.* Springer, New York.

[111] Johnson, N. L., Kotz, S. and Balakrishnan, N. (2004) *Continuous Univariate Distributions.* Vol.1., 2nd edition, Wiley and Sons, New York.

[112] Jones, M. C. and Faddy, M. J. (2003) A skewed extension of the $t-$distribution, with applications. *Journal of the Royal Statistical Society, Series B*, **65**(2), 159–174.

[113] Kailath, T., Sayed, A. H., and Hassibi, B. (2000) *Linear Estimation.* Prentice Hall, New Jersey.

[114] Kalemanova, A. and Werner, R. (2006) A short note on the efficient implementation of the normal inverse Gaussian distribution.
www.mathfinance.ma.tum.de/peppers/KalemanovaWerner-NoteOnImplementation.pdf

[115] Kalman, R. E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering: Transactions of the ASME*, **82**, 35–45.

[116] Kalman, R. E. and Bucy, R. (1961) New results in filtering and prediction theory. *Journal of Basic Engineering: Transactions of the ASME*, **83**, 95–108.

[117] Kellerhals, B. P. (2004) *Asset Pricing: Modeling and Estimation.* 2nd Edition, Springer-Verlag, Berlin.

[118] Kim, C.-J. (1994) Dynamic linear models with switching. *Journal of Econometrics*, **60**, 1–22.

[119] Kon, S. J. (1984) Models of stock returns–a comparison. *The Journal of Finance*, **39**(1), 147–165.

[120] Kovalev, V. F., Pustolalov, V. V. and Shirkov, D. V. (1998) Group analysis and renorm-group symmetries. $arXiv : hep - th/9706056v1 - 9June1997$.

[121] Krause, J. and Michel, L. (1990) Classifications of the Symmetries of Ordinary Differential Equations, Proceedings of the Summer School in *Group Theoritical Methods in Physics*, 4–9 June, V. V. Dodonov and V. I. Man'ko (ed). *Lecture Notes in Physics*, **382**, 251–262.

[122] Krichene, N. (2008) Crude oil prices: Trends and forecast. International Monetary Fund Working Paper (WP/08/133).

[123] Küchler, U., Neumann, K. Sørensen, M. and Stroller, A. (1994) Stock returns and hyperbolic distributions. Discussion paper 23, Sonderforschungsbereich 373, Humboldt-Universitt zu Berlin. Presented at the 2nd workshop on *Stochastics and Finance* in Berlin.

[124] Kundu, D. and Manglick, A (2004) Discriminating between the Weibull and log-normal distributions. *Naval Research Logistics*, **51**, 893–905.

[125] Kundu, D., Gupta, R. D. and Manglick, A (2005) Discriminating between the log-normal and the generalized exponential distributions. *Journal of Statistical Planning and Inference*, **127**, 213–227.

[126] Kwot, Y. K. (1998) *Mathematical Models of Financial Derivatives.* Springer-Varlag, Singapore.

[127] Leach, P. G. L. (2006) Mathematical modeling: Symmetry and singularity analyses. Lecture notes, School of Mathematical Sciences, University of KwaZulu-Natal, Durban.

[128] Leach, P. G. L. and Andriopoulos, K. (2005) Newtonian economics. *Proceedings of 10th International Conference in Modern group analysis.* 134–142.

[129] Leach, P. G. L. and Andriopouolos, K. (2006) A common theme in applied mathematics: An equation connecting applications in economics, medicine and physics. *South African Journal of Science*, **102**, Jan/Feb., 66–72.

[130] Leach, P. G. L., Andriopoulos, K. and Yannacopoulos, A. N. (2005) Nonlinear equations in financial mathematics: symmetries and linearization. Preprint, School of Mathematical sciences, University of KwaZulu-Natal, Durban.

[131] Lie, S. (1881) Über die integration durch bestimate integrale von einer Klasse linear particller differentialgleichungen. *Archiv für Mathematik*, **6**, 328–368 (translation in English by N. H. Ibragimov in *CRC Handbook of Lie group analysis of differential equations*, vol. 2, CRC Press, Boca Raton, 473–508; also Gesammelte Abhandlungen, III, 492–523.

[132] Liptser, R. S. and Shiryayev, A. N. (1977) *Statistics of Random Processes I. General Theory.* Springer-Verlag, New York.

[133] Lu, C., Danzer, R. and Fishcer, D.(2002) Fracture statistics of brittle materials: Weibull or normal distribution. *Physical Review E*, **65**, 067–102.

[134] Madan, D.B. and Seneta, E. (1990) The variance gamma (V.G.) model for share market returns. *Journal of Business*, **63**, 511–524.

[135] Mahieu, R. J. and Schotman, P. C. (1998) An empirical application of stochastic volatility models. *Journal of Applied Econometrics*, **13**, 333–360.

[136] Mahomed, F. M. and Leach, P. G. L. (1990) Symmetry Lie algebras of the nth–order ordinary differential equations. *Journal of Mathematical analysis and Applications*, **151**(1), 80–107.

[137] Malevergne, Y. and Sornette, D. (2004) Value-at-risk-efficient portfolios for a class of supper-and sub-exponentially decaying assets return distributions. *Quantitative Finance*, **4**, 17–36.

[138] Mandelbrot, B. (1963) The variation of certain speculative prices. *Journal of Business*, **36**, 394–419.

[139] Mandelbrot, B., Taylor, M. (1967) On the distribution of stock price differences. *Operations Research*, **15**, 1057–1062.

[140] Marathe, R. R. and Ryan, S. M. (2005) On the validity of the Brownian motion assumption. *The Engineering Economics*, **50**, 159–192.

[141] Mittnik, S. and Rachev, S. T. (1993) Modeling asset returns with alternative stable distributions. *Econometric Reviews*, **12**, 261–330.

[142] Mudholkar, G. and Natarajan, R. (2002) The inverse Gaussian models: analogoues of symmetry, skewness and kurtosis. *Annals of the Institute of Statistical Mathematics.* **54**, 138–154.

[143] Müler, U. A., Dacorogna, M. M., Oslen, R. B., Pictet, O. V., Swartz, M. and Morgenegg, C. (1990) Statistical study of foreign exchange rates, empirical evidence of a price change scaling law and intraday analysis. *Journal of Banking and Finance*, **14**, 1189–1208.

[144] Naicker, V. Andriopoulos, K. and Leach, P. G. L. (2005) Symmetry reduction of some nonlinear Hamilton-Jackobi-Bellman equations arising in the mathematics of finance. *Journal of Nonlinear Mathematical Physics*, **12**,264–274.

[145] Neftci, S. N. (2000) *An Introduction to the Mathematics of Financial Derivatives.* 2nd Edition, Academic Press, New York.

[146] Nucci, M. C. (1990) Interactive REDUCE programs for calculating classical, nonclassical and Lie-Bäcklund symmetries for differential equations. (preprint: Georgia Institute of Technology, Math 062090-051).

[147] Nucci, M. C. (1996) Interactive REDUCE programs for calculating Lie point, nonclassical, Lie-Bäcklund and approximate symmetries of differential equations: manual and floppy disc. In: N. H. Ibragimov (ed), *CRC Handbook of Lie Group Analysis of Differential Equations Vol. 3: New Trends in Theoretical Development and Computational Methods.* p. 415–481. CRC Press, Boca Raton.

[148] Øksendal, B. (2000) *Stochastic Differential Equations.* Springer-Verlag, Berlin.

[149] Olver, P. J. (1993) *Applications of Lie Groups to Differential Equations.* 2nd Edition, Springer-Verlag, Berlin.

[150] Olver, P. J. and Rosenau, P. (1987) Group-invariant solutions of differential equations. *SIAM Journal of Applied Mathematics*, **47**, 263–278.

[151] Osborne, M. F. M. (1972) Random nature of stock prices. *Journal of Economics and Business*, **6**, 220–233.

[152] Ovsyannikov, L. V. (1982) *Group Analysis of Differential Equations.* Academic Press, New York.

[153] Pennacchi, G. G. (1991) Identifying the dynamics of real investment rates and inflation: evidence using survey data. *Review of Financial Studies*, **4**, 53–86.

[154] Pilipovic, D. (1998) *Energy Risk: Valuing and Managing Energy Derivatives.* Cambridge University Press, Cambridge.

[155] Rege, S. and Menezes, A. G. (2009) Comparing the generalized hyperbolic to the normal inverse Gaussian distributions for the daily returns of the PSI20. Technical report. CEEA1A and Departmento de Economia e Gestão, Universidade dos Aores (accessed September, 2010).

[156] Romano, V. and Torrisi, M. (1999) Application of week equivalence transformations to a group analysis of a drift-diffusion model. *Journal of Physics A: Mathematical and General.* **32**, 7953–7963.

[157] Ross, S. (1997) Hedging long-run commitments: exercises in incomplete market pricing. *Economic Notes by Banca Monte die Paschi di Siena SpA*, **26**(2), 385–420.

[158] Ross, S. (2000) *Introduction to Probability Models.* 7th Edition, Harcourt Academic Press, New York.

[159] Rydberg, T. H. (1999) Generalized hyperbolic diffusion processes with application towards finance. *Mathematical Finance*, **9**, 183–201.

[160] Samuelson, P. A. (1965) Rational theory of warrant pricing. *Industrial Management Review*, **6**, 13–31.

[161] Schattschneider, D. (1990) *Visions of Symmetry: Notebooks, Periodic Drawings and Related Works of M. C. Escher.* W. H. Freeman and Co., New York.

[162] Schwartz, E. S. (1997) The stochastic behavior of commodity prices: implications for valuation and hedging. *Journal of Finance* , LII(3), 923973.

[163] Schwarz, F. A. (1982) A REDUCE package for determining Lie symmetries of ordinary and partial differential equations *Computer Physics Communications*, **27**, 179–186.

[164] Scott, D. (2007) The HyperbolicDist package

[165] Shapiro, S. S. and Wilks, M. B. (1965) An analysis of variance test for normality. *Biometrika*, **52**(3,4), 591–611.

[166] Sherring, J., Head, A. K. and Prince, G. E. (1997) Dimsym and LIE: symmetry determining packages. *Mathematical and Computational Modeling*, **25**, 153–164.

[167] Shirkov, D. V. and Kovalev, V. F. (2001) The Bogoliubov renormalization group and solution symmetries in mathematical physics. *Physical Reports*, **352**(4–6), 219–249.

[168] Shreve, S. E. (2004) *Stochastic Calculus for Finance II: Continuous-Time Models*). Springer-Verlag, New York.

[169] Shumway, R. H. and Stoffer, D. S. (1991) Dynamic linear models with switching. *Journal of American Statistical Association*, **86**(415), 763–769.

[170] Shumway, R. H. and Stoffer, D. S. (2009) *Time Series Analysis and its Applications with R Examples.* 3rd Edition, Springer-Verlag, New York.

[171] Sinkala, W. (2006) *Symmetry Analysis of Equations of Financial Mathematics.* PhD Thesis, School of Mathematical Sciences, University of KwaZulu-Natal, Durban, South Africa.

[172] Sinkala, W., Leach, P. G. L. and O'Hara, J. G. (2008a) An optimal system and group-invariant solutions of the CIR pricing equation. *Mathematical Methods in the Applied Sciences*, 665–678 (doi: 10.1002/mma.935).

[173] Sinkala, W., Leach, P. G. L. and O'Hara, J. G. (2008b) Invariance properties of a general bond-pricing equation. *Journal of Differential Equations*, **244**, 2820–2835.

[174] So, M., Lam, K. and Li, W. (1998) A stochastic volatility model with Markov switching. *Journal of Business and Economic Statistics*, **16**(2), 244–253.

[175] Sobczyk, K. (1995) *Stochastic Differential Equations with Applications to Physics and Engineering.* Kluwer Academy, Dordrecht.

[176] Sørensen, M. (2006) On the size distribution of sand. Working paper, Department of Applied Mathematics and Statistics, University of Copenhagen, Denmark.

[177] Stephani, H. (1989) *Differential Equations: Their Solutions Using Symmetries.* Cambridge University Press, New York.

[178] Töyli, J. (2002) *Essays on Asset Return Distributions.* D.Sc. Dissertation, Helsinki University of Technology, Helsinki, Finland.

[179] Tuchwell, H. C. (1988) *Elementary Applications of Probability Theory.* Chapman and Hall, London.

[180] Tuchwell, H. C. (1995) *EPPT-with an Introduction to Stochastic Differential Equations.* 3rd Edition, Chapman and Hall, London.

[181] Vasicek, O. (1977) An equilibrium characterization of the structure. *Journal of Financial Economics*, **5**, 177–188.

[182] Vuong, Q. H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**(2), 307–333.

[183] Wang, F., Weber, P., Yamasaki, K., Harlin, S. and Stanley, H.E. (2007) Statistical regularities in the return intervals of volatility. *European Physical Journal B*, **55**, 123–133.

[184] Weyl, H. (1952) *Symmetry.* Princeton University Press, New Jersey.

[185] Woof, C. and Hodgkinson, D. (1987) *muMATH; a Microcomputer Algebra System.* Academy Press, London.

[186] Yaglom, I. M. (1988) *Felix Klein and Sophus Lie: Evolution of the Idea of Symmetry in the Nineteenth Century.* Birkhäuser Verlag, Boston.