Lévy Option Pricing Models: Theory and Application

by

Kazuhisa Matsuda

A dissertation submitted to the Graduate Faculty in Economics
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy, The City University of New York

2006

UMI Number: 3213170

UMI Microform 3213170

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the
Graduate Faculty in Economics in satisfaction of the
dissertation requirements for the degree of Doctor of Philosophy.

April 4 2006	Salih N. Neftci
Date	Chair of Examining Committee
April 4 2006	Thom B. Thurston
Date	Executive Officer

Professor Salih N. Neftci

Professor Thom B. Thurston

Distinguished Professor Michael Grossman

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Lévy Option Pricing Models: Theory and Application

by

Kazuhisa Matsuda

Adviser: Professor Salih N. Neftci

The goal of this dissertation is to provide both the underlying theory and applications of Lévy option pricing models which have been developed recently. The dissertation is composed of five main parts. Part I covers the mathematical theory of Lévy processes which constitute a wide class of stochastic processes whose sample paths can be continuous, mostly continuous with occasional discontinuities, and purely discontinuous. Each Lévy process is defined and characterized using theorems such as the Lévy-Itô decomposition and the Lévy-Khinchin representation. They also are characterized in terms of their infinite divisibilities and the Lévy measures. Part 2 provides the introduction to Fourier transform and its application to the option pricing. Fourier transform option pricing method has become the mainstream pricing method because of its generality in the sense that it only requires the characteristic function of the log terminal asset price. Part 3 describes the mathematics of the Lévy option pricing models. The classic Black-Scholes (BS) model is characterized as the only continuous Lévy model. The

classic Merton jump-diffusion (MJD) model is characterized as the non-Gaussian Lévy model in which asset price dynamics is modeled by the jump diffusion process. The recently developed variance gamma (VG) model, normal inverse Gaussian (NIG) model, and Carr, Geman, Madan, and Yor (CGMY) model are characterized as pure jump Lévy models in which the asset price dynamics are modeled using pure jump Lévy processes. These pure jump Lévy models are calibrated to the S&P 500 futures options and we compare their pricing performance relative to classic models along with the implied dynamics of the log return density and the Lévy measure. Part 4 treats the regularization using the relative entropy in the Merton model to overcome the illposed problem of unregularized calibration with pure jump Lévy models. Part 5 studies the other approaches to recover the risk-neutral probability density function of terminal asset prices. One is the semiparametric volatility-interpolation method developed by Shimko which makes no assumptions regarding the underlying stochastic process of asset prices. The other is the mixture of two lognormals approach which is a parametric approach by Bahra.

To my mother and father

Acknowledgements

I would like to thank my dissertation supervisor Professor Salih N. Neftci for his support and encouragement. Professor Neftci inspired me to work on pure jump Lévy models and gave me numerous helpful advices. I am also grateful to Professor Thom B. Thurston and Professor Michael Grossman who always provided me with kind support and advice. The experience as a Ph.D. student here at the Graduate Center of CUNY was an exciting and enjoyable one. I am certain that many lessons I learned here will guide me to the right direction in my career and in my life.

Table of Contents

Abstract iii
Acknowledgements vi
Table of Contents vii
List of Tables xi
List of Figures xii

Chapter 1: Introduction to the Mathematics of Lévy Processes 1
1.1 Introduction 2
1.2 Function 2
1.2.1 Definition of Function 2
1.2.2 Left Limit and Right Limit of a Function 3
1.2.3 Right Continuous Function and Right Continuous with Left Limit (RCLL) Function 3
1.2.4 Continuous Function 5
1.2.5 Discontinuous Function 5
1.3 Stochastic Processes 6
1.3.1 Definition of Stochastic Processes 6
1.3.2 Putting Structure on Stochastic Processes 7
1.3.3 Processes with Independent and Stationary Increments: Imposing Structure on a Probability Measure 7
1.3.4 Martingale: Structure on Conditional Expectation 11
1.3.5 Example of Continuous Martingale: Standard Brownian Motion 13
1.3.6 Martingale Asset Pricing 15
1.3.7 Submartingales and Supermartingales 16
1.3.8 Markov Processes: Structure on Conditional Probability 17
1.3.9 Sample Path Properties of Stochastic Processes 23
1.4 Lévy Processes 28
1.4.1 Definition of Lévy Processes 28
1.4.2 Infinitely Divisible Random Variable and Distribution 30
1.4.3 Relationship between Lévy Processes and Infinitely Divisible Distributions 33
1.4.4 Lévy-Khinchin Representation 35
1.4.5 Lévy-Itô Decomposition of Sample Paths of Lévy Processes 41
1.4.6 Lévy Measure 44
1.4.7 Classification of Lévy Processes: In Terms of Gaussian or Not 47
1.4.8 Classification of Lévy Processes: In Terms of the Behavior of Lévy Measure 49
1.4.9 Classification of Lévy Processes: In Terms of the Total Variation of Lévy Process 50

1.4.10 Classification of Lévy Processes: In Terms of the Properties of Lévy Triplet by Sato (1999) 52
1.4.11 Classification of Lévy Processes: In Terms of the Sample Paths Properties of Lévy Processes 53
1.4.12 Lévy Processes as a Subclass of Markov Processes 55
1.4.13 Other Important Properties of Lévy Processes 60
1.5 Examples of Lévy Processes 61
1.5.1 Brownian Motion: The only Lévy Process with Continuous Sample Paths 61
1.5.2 Poisson Process 67
1.5.3 Compound Poisson Process 82

Chapter 2: Introduction to Option Pricing with Fourier Transform 89
2.1 Introduction 90
2.2 Prerequisite for Fourier Transform 90
2.2.1 Radian 91
2.2.2 Wavelength 91
2.2.3 Frequency, Angular Frequency, and Period of a Waveform 91
2.2.4 Sine and Cosine 93
2.2.5 Derivative and Integral of Sine and Cosine Function 93
2.2.6 Series Definition of Sine Function and Cosine Function 94
2.2.7 Euler's Formula 94
2.2.8 Sine Wave: Sinusoid 95
2.3 Fourier Transform (FT) 96
2.3.1 Definition of Fourier Transform 96
2.3.2 Examples of Fourier Transform 99
2.4 Properties of Fourier Transform 106
2.4.1 Dirac's Delta Function (Impulse Function) 106
2.4.3 Linearity of Fourier Transform 112
2.4.4 FT of Even and Odd Functions 113
2.4.5 Symmetry of Fourier Transform 114
2.4.6 Differentiation of Fourier Transform 116
2.4.7 Time Scaling of Fourier Transform 117
2.4.8 Time Shifting of Fourier Transform 118
2.4.9 Convolution: Time Convolution Theorem 119
2.4.10 Frequency-Convolution Theorem 121
2.4.11 Frequency Shifting: Modulation 122
2.4.12 Parseval's Relation 123
2.4.13 Summary of Important Properties of Fourier Transform 125
2.5 Existence of the Fourier Integral 125
2.6 Characteristic Function 126
2.6.1 Definition of a Characteristic Function 126
2.6.2 Properties of a Characteristic Function 128
2.6.3 Characteristic Exponent: Cumulant-Generating Function 129
2.6.4 Laplace Transform 130

2.6.5 Relationship with Moment Generating Function 131
2.6.6 Summary: How to Calculate Standardized Moments from Characteristic Function and Moment Generating function 134
2.6.7 Examples of Characteristic Functions 134
2.7 Option Pricing with Fourier Transform: Black-Scholes Example 134
2.7.1 Motivation 134
2.7.2 Derivation of Call Price with Fourier Transform: Carr and Madan (1999) 137
2.7.3 How to Choose Decay Rate Parameter α: Carr and Madan (1999)..... 140
2.7.4 Black-Scholes Model with Fourier Transform Pricing Method..... 142

Chapter 3: Calibration of Lévy Option Pricing Models: Application to S&P 500 Futures Option 143
3.1 Introduction 144
3.2 Pure Jump Lévy Models 148
3.2.1 Variance Gamma Model by Carr, Chang, and Madan (1998) 148
3.2.2 Normal Inverse Gaussian Model by Barndorff-Nielsen (1998) 153
3.2.3 CGMY Model by Carr, Geman, Madan, and Yor (2002) 158
3.3 Traditional Lévy Models 161
3.3.1 Black-Scholes Model (1973) 161
3.3.2 Merton Jump Diffusion Model (1976) 163
3.4 Calibration Methodology 168
3.4.1 Data Description 168
3.4.2 Introduction to Calibration 168
3.4.3 Pricing Methodology for Lévy Models 170
3.5 Calibration, Implied Dynamics, and In-Sample Performance of Lévy Models 171
3.5.1 Dynamics of Calibrated Parameters 171
3.5.2 Implied Dynamics of Log Return Probability Density Functions 174
3.5.3 Dynamics of Implied Lévy Density Functions of Log Returns 174
3.5.4 In-Sample Performance of Lévy Models 175
3.6 Out-of-Sample Pricing Performance of Lévy Models 176
3.7 Conclusion 180

Chapter 4: Parametric Regularized Calibration of Merton Jump-Diffusion Model with Relative Entropy 182
4.1 Introduction 183
4.2 Merton Jump Diffusion (MJD) Model (1976) 185
4.3 Calibration without Regularization 190
4.3.1 Calibration: An Inverse Problem 190
4.3.2 Calibration: An Illposed Problem 191
4.4 Regularized Calibration 192
4.4.1 Relative Entropy 192
4.4.2 Relative Entropy for Lévy Processes 194
4.4.3 Regularized Calibration 197

4.5 Empirical Example: Obtaining the Prior 200
4.5.1 Computation of the Statistical Prior with Time-Series Data 200
4.5.2 Computation of the Risk-Neutral Prior with Option Data..... 204
4.5.3 Comparison between the Statistical Prior and the Risk-Neutral Prior .. 205
4.6 Empirical Result of Regularized Calibration with Relative Entropy 206
4.6.2 Empirical Results..... 207
4.7 Conclusion 211

Chapter 5: Other Approaches to Obtain Option Price Implied Risk-Neutral Probability Density Function of Asset Prices..... 213
5.1 Introduction 214
5.2 Option Implied Method and Traditional Method to Estimate Risk..... 218
Neutral Density (RND) of Financial Asset Price 218
5.2.1 Volatility Interpolation (VI) Method of Shimko (1993) 218
5.2.2 Mixture of Two Lognormals Method (2LN) of Bahra (1997) 220
5.2.3 Traditional Method..... 223
5.3 Estimating the Dynamics of RNDs from S&P 500 Futures Options Data... 225
5.3.1 Data Set 225
5.3.2 Dynamics of RND Estimates..... 226
5.3.3 Instantaneous Profile and the dynamics of Moments..... 227
5.4 Goodness of Fit Test..... 228
5.5 Conclusion 230

Appendix 356

Bibliography 364

List of Tables

Table 1.1: Martingale Gambling Strategy 232
Table 1.2: One-to-one correspondence between an infinitely divisible distribution and a Lévy process 233
Table 1.3: Poisson process 234
Table 1.4: Compound Poisson process 235

Table 2.1: Relationship between frequency f and angular frequency ω 236
Table 2.2: Summary of Important Properties of Fourier Transform in Angular Frequency Domain ω Hz (radians/second) 237
Table 2.3: Summary of Important Properties of Fourier Transform in Frequency Domain f Hz (cycles/second) 238
Table 2.4: How to Calculate Standardized Moments from Characteristic Function and Moment Generating function 239
Table 2.5: Examples of Characteristic Functions 240

Table 3.1 Role of parameter Y 241
Table 3.2: Model Comparison 242
Table 3.3: Calibrated Parameters 243
Table 3.4: In-Sample Fit 244
Table 3.5: Out-of-sample absolute pricing errors per call option 245
Table 3.6: Orthogonality test results of out-of-sample dollar pricing errors 246

Table 4.1: Statistically estimated parameters of each model 247
Table 4.2: Annualized standardized moments of statistical log return probability density 248
Table 4.3: Risk-neutral prior parameters 249
Table 4.4: Annualized standardized moments of risk-neutral log return probability density 250
Table 4.5: Annualized standardized moments of statistical vs. risk-neutral prior log return probability density 251
Table 4.6: Calibration Results 252

Table 5.1: Moments of RND estimates with 9-month to maturity 253
Table 5.2: Moments of RND estimates with 6-month to maturity 254
Table 5.3: Moments of RND estimates with 3-month to maturity 255
Table 5.4: Moments of RND estimates with 1-month to maturity 256

List of Figures

Figure 1.1: Examples of a function f: a → f(a) 257
Figure 1.2: Plot of the probability density function of an exponential random variable with $\lambda = 0.01$ 258
Figure 1.3: Zero arrival of an event in [0, t] 259
Figure 1.4: One arrival of an event in [0, t] 260
Figure 1.5: k arrivals of an event in [0, t] 261
Figure 1.6: An event arrivals k times in the interval [0, t] 262
Figure 1.7: Simulated sample paths of Poisson processes with the intensity $\lambda = 5$ and $0 \leqslant t \leqslant 2$ 263
Figure 1.8: Simulated sample paths of a compound Poisson process with the intensity $\lambda = 5$ and $0 \leqslant t \leqslant 2$ 264

Figure 2.1: Illustration of wavelength λ 265
Figure 2.2: Plot of 5 Hz sine wave $g(t) = \sin(2\pi(5)t)$ 266
Figure 2.3: The definition of sine and cosine function with unit circle 267
Figure 2.4: Sine and cosine function $\sin(\theta)$ and $\cos(\theta)$ 268
Figure 2.5: Plot of a sine wave $g(t) = \sin(2\pi ft)$ with different fundamental frequency f_0 269
Figure 2.6: Plot of a 1 Hz sine wave $g(t) = a \sin(2\pi t)$ with different amplitudes a = 1/2, 1, and 2 270
Figure 2.7: Plot of a 1 Hz Sine Wave $g(t) = \sin(2\pi t + b)$ with Different Phase b = 0, $\pi/2$, and $-\pi/2$ 271
Figure 2.8: Plot of double-sided exponential function and its Fourier transforms. 272
Figure 2.9: Plot of rectangular pulse and its Fourier transforms 273
Figure 2.10: Plot of Dirac's delta function and its Fourier transforms 274
Figure 2.11: Plot of Gaussian function and Fourier transforms 275
Figure 2.12: Plot of a cosine wave and Fourier transforms 276
Figure 2.13: Plot of a sine wave and Fourier transforms 277
Figure 2.14: Plot of a function h(x) for n = 1, $n = 10^{1/2}$, and n = 10 278
Figure 2.15: Plot of even and odd functions 279
Figure 2.16: Plot of two Gaussian functions f and g and their convolution f*g 280
Figure 2.17: Relationship between a strike price K and a log-strike price $k \equiv \ln K$ 281
Figure 2.18: Plot of an exponential function $e^{0.5k}$ with respect to log-strike price $k \equiv \ln K$ 282

Figure 3.1: Plot of Lévy measure $l_{VG,X}(x; \theta,\sigma,\upsilon)$ of VG process 283
Figure 3.2: Plot of Lévy measure $l_{NIG,X}(x; \alpha,\beta,\delta)$ of NIG process..... 284
Figure 3.3: Plot of Lévy measure $l_{CGMY,X}(x; C, G, M, Y)$ of CGMY process 285
Figure 3.4: Plot of Lévy measure $l_{CGMY,X}(x; C, G, M, Y)$ of CGMY process 286
Figure 3.5: Plot of Lévy measure $l_{MJD,X}(x; \lambda, \mu, \delta)$ of MJD process with various values for λ 287
Figure 3.6: Dynamics of calibrated parameters: σ of Black-Scholes model..... 288
Figure 3.7: Dynamics of calibrated parameters: σ, λ, μ, δ of Merton jump-diffusion model 289
Figure 3.8: Dynamics of calibrated parameters: θ, σ, υ of Variance Gamma model 290
Figure 3.9: Dynamics of calibrated parameters: α, β, δ of Normal Inverse Gaussian model 291
Figure 3.10: Dynamics of calibrated parameters: C, G, M, Y of CGMY model ... 292
Figure 3.11: Black-Scholes model implied dynamics of log-return probability density..... 293
Figure 3.12: MJD model implied dynamics of log-return probability density..... 294
Figure 3.13: VG model implied dynamics of log-return probability density..... 295
Figure 3.14: NIG model implied dynamics of log-return probability density 296
Figure 3.15: Plot of implied log-return probability density 297
Figure 3.15: Plot of implied log-return probability density (Continued) 298
Figure 3.15: Plot of implied log-return probability density (Continued). 299
Figure 3.16: Dynamics of implied moments of log-return probability density..... 300
Figure 3.16: Dynamics of implied moments of log-return probability density (Continued)..... 301
Figure 3.17: Implied dynamics of Lévy density: MJD model..... 302
Figure 3.18: Implied dynamics of Lévy density: VG model..... 303
Figure 3.19: Implied dynamics of Lévy density: NIG model 304
Figure 3.20: Implied dynamics of Lévy density: CGMY model..... 305
Figure 3.21: Plot of implied Lévy density of log returns 306
Figure 3.21: Plot of implied Lévy density of log returns (Continued)..... 307
Figure 3.21: Plot of implied Lévy density of log returns (Continued)..... 308
Figure 3.22: Dynamics of in-sample dollar pricing errors: BS model 309
Figure 3.23: Dynamics of in-sample dollar pricing errors: MJD model 310
Figure 3.24: Dynamics of in-sample dollar pricing errors: VG model 311
Figure 3.25: Dynamics of in-sample dollar pricing errors: NIG model 312
Figure 3.26: Dynamics of in-sample dollar pricing errors: CGMY model 313
Figure 3.27: Dynamics of out-of-sample dollar pricing errors: BS model..... 314
Figure 3.28: Dynamics of out-of-sample dollar pricing errors: MJD model 315
Figure 3.29: Dynamics of out-of-sample dollar pricing errors: VG model..... 316
Figure 3.30: Dynamics of out-of-sample dollar pricing errors: NIG model 317
Figure 3.31: Dynamics of out-of-sample dollar pricing errors: CGMY model 318

Figure 4.1: 3D plot of the MJD error function 319
Figure 4.2: Strictly convex function $f(x) = x\ln x$..... 320

Figure 4.3: Relative entropy function for Gaussian case 321
Figure 4.4: Relative entropy function for MJD Process 322
Figure 4.5: Statistical daily log-return probability density 323
Figure 4.6: Log of statistical daily log-return probability density 324
Figure 4.7: Plot of statistically estimated MJD Lévy measure 325
Figure 4.8: Successively-updated risk-neutral prior parameters 326
Figure 4.9: Risk-neutral log-return probability density 327
Figure 4.10: Log of risk-neutral daily log-return probability density 328
Figure 4.11: Plot of risk-neutral MJD Lévy measure 329
Figure 4.12: Statistical vs. risk-neutral prior log-return probability densities 330
Figure 4.12: Statistical vs. risk-neutral prior log-return probability densities (Continued) 331
Figure 4.13: Plot of calibrated log-return probability density and Lévy measure on March 30, 2004 (245 days to maturity) 332
Figure 4.14: Plot of calibrated log-return probability density and Lévy measure on July 1, 2004 (180 days to maturity) 333
Figure 4.15: Plot of calibrated log-return probability density and Lévy measure on September 27, 2004 (120 days to maturity) 334
Figure 4.16: Plot of calibrated log-return probability density and Lévy measure on November 22, 2004 (80 days to maturity) 335
Figure 4.17: Plot of calibrated log-return probability density and Lévy measure on January 20, 2005 (40 days to maturity) 336
Figure 4.18: Plot of calibrated log-return probability density and Lévy measure on March 4, 2005 (10 days to maturity) 337

Figure 5.1: Implied volatility surface 338
Figure 5.2: Dynamics of risk-neutral densities obtained from volatility interpolation method 339
Figure 5.3: Dynamics of cumulative risk-neutral densities obtained from volatility interpolation method 340
Figure 5.4: Dynamics of risk-neutral densities obtained from mixture of two lognormals method 341
Figure 5.5: Dynamics of cumulative risk-neutral densities obtained from mixture of two lognormals method 342
Figure 5.6: Dynamics of risk-neutral densities obtained from the benchmark BS lognormal 343
Figure 5.7: Dynamics of cumulative risk-neutral densities obtained from the benchmark Black-Scholes lognormal 344
Figure 5.8: Risk-neutral density estimates with 9-month to maturity 345
Figure 5.9: Risk-neutral density estimates with 6-month to maturity 346
Figure 5.10: Risk-neutral density estimates with 3-month to maturity 347
Figure 5.11: Risk-neutral density estimates with 1-month to maturity 348
Figure 5.12: Dynamics of moments of risk-neutral density estimates 349
Figure 5.12: Dynamics of moments of risk-neutral density estimates (Continued) 350

Figure 5.13: Kolmogorov-Smirnov test of risk-neutral density estimates between volatility interpolation and benchmark BS 351
Figure 5.14: Kolmogorov-Smirnov test of risk-neutral density estimates between mixture of two lognormals and benchmark BS. 352
Figure 5.15: Kolmogorov-Smirnov test statistic of risk-neutral density estimates between volatility interpolation and mixture of two lognormals methods 353
Figure 5.16: Historical volatility and at-the-money implied volatility 354
Figure 5.17: Absolute deviation between historical volatility and at-the-money implied volatility 355

Chapter 1: Introduction to the Mathematics of Lévy Processes

1.1 Introduction

The goal of this chapter is to provide the foundations of the mathematics of Lévy processes for the readers with undergraduate knowledge of stochastic processes as simple as possible. The simplicity is a key because, for the beginners such as finance majors without the experience in stochastic processes, some available books on Lévy processes are not accessible. Lévy processes constitute a wide class of stochastic processes whose sample paths can be continuous, discontinuous, and purely discontinuous. Traditional examples of Lévy processes include a Brownian motion with drift (i.e. the only continuous Lévy processes), a Poisson process, a compound Poisson process, a jump diffusion process, and a Cauchy process. All of these are well studied and well applied stochastic processes. We define and characterize Lévy processes using theorems such as the Lévy-Itô decomposition and the Lévy-Khinchin representation and in terms of their infinite divisibilities and the Lévy measures ℓ.

1.2 Function

1.2.1 Definition of Function

Definition 1.1 Function A function $f : a \to f(a)$ or $f : A \to B$ on $\mathbb{R}$ uniquely maps a set of input values $a \in A$ to a set of output values $f(a) \in B$. The domain of a function is the set A on which a function is defined and the set of all actual outputs $f(a) \in B$ is called the range of a function.

A function is a many-to-one mapping (i.e. not one-to-many mapping). For example, a function $f(a) = a$ is a one-to-one mapping, $f(a) = -a^2$ is a two-to-one mapping except for $a = 0$, and $f(a) = \sin(2\pi a)$ is a many-to-one mapping as illustrated by Figure 1.1.

1.2.2 Left Limit and Right Limit of a Function

Definition 1.2 Left limit and Right limit of a function A function $f : a \to f(a)$ on $\mathbb{R}$ has a left limit $f(b-)$ at a point $a = b$ if $f(a)$ approaches $f(b-)$ when a approaches b from the below (the left-hand side):

$$\lim_{a \to b-} f(a) = f(b-) \quad (1\text{-}1).$$

A function $f : a \to f(a)$ on $\mathbb{R}$ has a right limit $f(b+)$ at a point $a = b$ if $f(a)$ approaches $f(b+)$ when a approaches b from the above (right-hand side):

$$\lim_{a \to b+} f(a) = f(b+) \quad (1\text{-}2).$$

1.2.3 Right Continuous Function and Right Continuous with Left Limit (RCLL) Function

Definition 1.3 Right continuous function A function f on $\mathbb{R}$ is said to be right continuous at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.

(2) Right limit of the function as a approaches b from the above (right hand side) exists, $\lim_{a \to b+} f(a) = f(b+)$.

(3) $f(b+) = f(b)$.

Definition 1.4 Right continuous with left limit (rcll) function A function f on $\mathbb{R}$ is said to be right continuous with left limit at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.

(2) Right limit of the function as a approaches b from the above (right hand side) exists, $\lim_{a \to b+} f(a) = f(b+)$. Left limit of the function as a approaches b from the below (left hand side) exists, i.e. $\lim_{a \to b-} f(a) = f(b-)$.

(3) $f(b+) = f(b)$.

The above definitions imply that a rcll function is right continuous, but the reverse is not true. In other words, a rcll function is more restrictive than a right continuous function because a rcll function needs left limit.

Consider a piecewise constant function defined as:

$$f(a) = \begin{cases} 0 & \text{if } a < 1 \\ 1 & \text{if } 1 \le a < 2 \\ 2 & \text{if } 2 \le a < 3 \end{cases} \qquad \text{(1-3).}$$

The right limit at a point $a = 1$ is equal to the actual value of the function at a point $a = 1$:

$$f(1+) = f(1) = 1,$$

this means f is right continuous at a point $a = 1$. But the left limit at a point $a = 1$ is not equal to the actual value of the function at a point $a = 1$:

$$f(1-) = 0 \neq f(1) = 1,$$

this means f is not left continuous at a point $a = 1$. Therefore, this function is right continuous with left limit. And the jump size is:

$$f(1+) - f(1-) = 1 - 0 = 1.$$

1.2.4 Continuous Function

Definition 1.5 Continuous function A function $f : a \to f(a)$ on $\mathbb{R}$ is said to be continuous at a point $a = b$ if it satisfies the following conditions:

(1) $f(b)$ is defined. In other words, a point b is in the domain of a function f.

(2) Right limit of the function as a approaches b from the above (right hand side) exists, $\lim_{a \to b+} f(a) = f(b+)$. Left limit of the function as a approaches b from the below (left hand side) exists, i.e. $\lim_{a \to b-} f(a) = f(b-)$.

(3) $f(b+) = f(b-) = f(b)$.

In other words, a continuous function is a left and right continuous function which in turn means that a continuous function is the most restrictive among rc, rcll, and continuous functions. All the functions in Figure 1.1 are continuous.

1.2.5 Discontinuous Function

Definition 1.6 Discontinuous function A function $f : a \to f(a)$ on $\mathbb{R}$ is said to be discontinuous at a point $a = b$ (called a point of discontinuity) if it fails to satisfy being a continuous function.

1.3 Stochastic Processes

1.3.1 Definition of Stochastic Processes

A stochastic process is a collection of random variables:

$$(X_{t\in[0,T]}),$$

where the index denotes time. Note that we are interested in the continuous time stochastic process where the time index takes any value in the interval $t\in[0,T]$. Discrete time stochastic process can be defined using a countable index set $t\in\mathbb{N}$:

$$(X_{t\in\mathbb{N}}).$$

A stochastic process is defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$ where Ω is an arbitrary set and $\mathbb{P}$ is a probability measure on $\mathcal{F}_{t\in[0,T]}$. $\mathcal{F}_{t\in[0,T]}$ is called a filtration which is an increasing family of σ-algebras of a subset of Ω which satisfy for $\forall 0\le s\le t$:

$$\mathcal{F}_s\subseteq\mathcal{F}_t.$$

Intuitively speaking, a filtration is an increasing information flow about $(X_{t\in[0,T]})$ as time progresses.

We can alternatively state that a real valued continuous time stochastic process is a random function:

$$X:[0,T]\times\Omega\to\mathbb{R} \qquad (1\text{-}4).$$

After the realization of the randomness ω, a sample path of $(X_{t\in[0,T]})$ is defined as:

$$X(\omega):t\to\mathbb{R} \text{ or } X(\omega):t\to X_t(\omega) \qquad (1\text{-}5).$$

A stochastic process $(X_{t\in[0,T]})$ is said to be nonanticipating with respect to the filtration $\mathcal{F}_t$ or $\mathcal{F}_t$-adapted if the value of X_t is revealed at time t for each $t \in [0,\ T]$.

1.3.2 Putting Structure on Stochastic Processes

The purpose of any mathematical modeling regardless of the field is to fit less complicated models to the highly complicated real world phenomena as accurate as possible. Mathematical models are less complicated in the sense that they make some simplifying assumptions or put some simplifying structures on the real world phenomena for the purpose of gaining tractability. There are some popular dependence structures put on stochastic processes which mathematicians have developed and used for years.

1.3.3 Processes with Independent and Stationary Increments: Imposing Structure on a Probability Measure

Before giving the definition of processes with independent and stationary increments, we must know the basics.

Definition 1.7 Conditional probability The conditional probability of an arbitrary event A given an event with positive probability B is:

$$\mathbb{P}(A\big|B) = \frac{\mathbb{P}(A \bigcap B)}{\mathbb{P}(B)} \qquad (1\text{-}6).$$

When $\mathbb{P}(B) = 0$, $\mathbb{P}(A\big|B)$ is undefined.

Definition 1.8 Statistical (Stochastic) independence Two arbitrary events A and B are said to be independent, if and only if:

$$\mathbb{P}(A \bigcap B) = \mathbb{P}(A)\mathbb{P}(B) \quad (1\text{-}7).$$

This definition of independence has two advantages. Firstly, it is symmetric in A and B. In other words, an event A's independence of an event B implies an event B's independence of an event A. Secondly, this definition holds even when an event B has zero probability, i.e. $\mathbb{P}(B) = 0$. When two arbitrary events A and B are independent, from the definition of a conditional probability:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A) \quad (1\text{-}8).$$

It is important to note that this is a result of statistical independence and not the definition. This is because the above equation is not true (i.e. undefined) when $\mathbb{P}(B) = 0$ and it is not symmetric in that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ does not necessarily imply $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

Definition 1.9 Mutual statistical independence Arbitrary events A_1, A_2, …, A_n are said to be mutually independent, if and only if:

$$\mathbb{P}(A_1 \bigcap A_2 \bigcap ... \bigcap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_n) \quad (1\text{-}9).$$

Definition 1.10 Processes with Independent and Stationary Increments A stochastic process $\left(X_{t\in[0,T]}\right)$ with values in $\mathbb{R}$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be a process with independent and stationary increments if it satisfies the following conditions:

(1) Its increments are independent. In other words, for $t_1 < t_2 < ... < t_n$:

$$\mathbb{P}(X_{t_2} - X_{t_1} \bigcap X_{t_3} - X_{t_2} \bigcap ... \bigcap X_{t_n} - X_{t_{n-1}}) = \mathbb{P}(X_{t_2} - X_{t_1})\mathbb{P}(X_{t_3} - X_{t_2})...\mathbb{P}(X_{t_n} - X_{t_{n-1}}).$$

(2) Its increments are stationary: i.e. for $\forall h \in \mathbb{R}^+$, $X_{t+h} - X_t$ has the same distribution as X_h. In other words, the distribution of increments does not depend on t (i.e. temporal homogeneity).

Consider an increasing sequence of time $0 < t_1 < t_2 < ... < t_n < t < u < \infty$ where t is the present. As a result of independent increments condition:

$$\mathbb{P}(X_u - X_t \,|\, X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})$$

$$= \frac{\mathbb{P}(X_u - X_t \cap X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})}{\mathbb{P}(X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})}$$

$$= \frac{\mathbb{P}(X_u - X_t)\mathbb{P}(X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})}{\mathbb{P}(X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_t - X_{t_n})}$$

$$= \mathbb{P}(X_u - X_t) \quad (1\text{-}10),$$

which means that there is no correlation (probabilistic dependence structure) on the increments among the past, the present, and the future.

For example, independent increments condition means that when modeling a log stock price $\ln S_t$ as an independent increment process, the probability distribution of a log stock price in year 2005 – 2006 is independent of the way the log stock price increment has evolved over the years. In other words, it doesn't matter if this stock crushes or soars in year 2004 – 2005:

$$\mathbb{P}(\ln S_{2006} - \ln S_{2005} \,|\, ..., \ln S_{2003} - \ln S_{2002}, \ln S_{2004} - \ln S_{2003}, \ln S_{2005} - \ln S_{2004})$$

$$= \mathbb{P}(\ln S_{2006} - \ln S_{2005}) \quad (1\text{-}11).$$

Using the simple relationship $X_u \equiv (X_u - X_t) + X_t$ for an increasing sequence of time $0 < t_1 < t_2 < ... < t_n < t < u < \infty$:

$$\mathbb{P}(X_u \,|\, X_0, X_{t_1}, X_{t_2}, ..., X_{t_n}, X_t) = \mathbb{P}((X_u - X_t) + X_t \,|\, X_0, X_{t_1}, X_{t_2}, ..., X_{t_n}, X_t)$$

$$= \mathbb{P}(X_u \,|\, X_t) \quad (1\text{-}12),$$

which holds because an increment $(X_u - X_t)$ is independent of X_t by definition and the value of X_t depends on its realization $X_t(\omega)$. This is a strong probabilistic structure imposed on a stochastic process because this means that the conditional probability of the future value X_u depends only on the previous realization $X_t(\omega)$ and not on the entire past history of realizations $X_0, X_{t_1}, X_{t_2}, ..., X_{t_n}, X_t$ (i.e. called Markov property which is discussed soon).

Although this condition seems too strong, it imposes a very tractable property on the process. Because if two variables X and Y are independent:

$$E[XY] = E[X]E[Y],$$

$$Var[X + Y] = Var[X] + Var[Y],$$

$$Cov[X, Y] = 0 \text{ (i.e. } Corr[X, Y] = 0) \quad (1\text{-}13).$$

Stationary increments condition means that the distributions of increments $X_{t+h} - X_t$ do not depend on the time t, but they depend on the time-distance h of two observations (i.e. interval of time). In other words, the probability density function of increments does not change over time. For example, if you model a log stock price $\ln S_t$ as a process with stationary increments, the distribution of increment in year 2005 – 2006 is the same as that in year 2050 – 2051:

$$\ln S_{2006} - \ln S_{2005} \overset{d}{=} \ln S_{2051} - \ln S_{2050}\,.$$

There is no doubt that the above independent and stationary increments conditions impose a strong structure on a stochastic process $\left(X_{t\in[0,T]}\right)$, as a result of these restrictions, the mean and variance of the process is tractable:

$$E[X_t] = \mu_0 + \mu_1 t\,,$$

$$Var[X_t] = \sigma_0^{\,2} + \sigma_1^{\,2} t \qquad \text{(1-14)},$$

where $\mu_0 = E[X_0]$, $\mu_1 = E[X_1] - \mu_0$, $\sigma_0^{\,2} = E[(X_0 - \mu_0)^2]$, and $\sigma_1^{\,2} = E[(X_1 - \mu_1)^2] - \sigma_0^{\,2}$.

1.3.4 Martingale: Structure on Conditional Expectation

Originally, the word 'martingale' comes from a French acronym of a gambling strategy. Imagine a coin flip gamble in which you win if a head turns up and you lose if a tail turns up. Martingale strategy requires a gambler to double his bet after every loss. Following martingale strategy, a gambler can recover all the losses he made and end up with an initial amount of his wealth plus an initial bet. Table 1.1 gives a sample path of a martingale strategy in which a gambler initially owns $200 of wealth, start betting with a stake of $2, and due to his bad luck his first win comes at the seventh trial. As you can see, he basically ends up where he started, i.e. his initial wealth of $200 (plus an initial bet of $2). Thus, a martingale strategy tells that after gambling many hours a gambler gains nothing (loses nothing) and his wealth remains constant on average.

In probability theory, a stochastic process is said to be a martingale if its sample path has no trend. Formally, a martingale is defined as the follows.

Definition 1.11 Martingale Consider a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$. A rcll stochastic process $(X_t)_{t\in[0,\,T]}$ is said to be a martingale with respect to the filtration $\mathcal{F}_t$ and under the probability measure $\mathbb{P}$ if it satisfies the following conditions:

(1) X_t is nonanticipating.

(2) $E[|X_t|] < \infty$ for $\forall t \in [0,\ T]$. Finite mean condition.

(3) $E[X_u | \mathcal{F}_t] = X_t$ for $\forall u > t$.

In other words, if a stochastic process is a martingale, then, the best prediction of its future value is its present value. Note that the definition of martingale makes sense only when the underlying probability measure $\mathbb{P}$ and the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ have been specified.

The fundamental property of a martingale process is that its future variations are completely unpredictable with the filtration $\mathcal{F}_t$:

$$\forall u > 0,\ E[x_{t+u} - x_t | \mathcal{F}_t] = E[x_{t+u} | \mathcal{F}_t] - E[x_t | \mathcal{F}_t] = x_t - x_t = 0 \qquad (1\text{-}14).$$

Finite mean condition is necessary to ensure the existence of the conditional expectation.

1.3.5 Example of Continuous Martingale: Standard Brownian Motion

Let $(B_{t\in[0,\infty)})$ be a standard Brownian motion process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, $(B_{t\in[0,\infty)})$ is a continuous martingale with respect to the filtration $\mathcal{F}_{t\in[0,\infty)}$ and the probability measure $\mathbb{P}$.

Proof

By definition, $(B_{t\in[0,\infty)})$ is a nonanticipating process (i.e. $\mathcal{F}_{t\in[0,\infty)}$ - adapted process) with the finite mean $E[|B_t|] = 0 < \infty$ for $\forall t \in [0,\infty)$. For $\forall 0 \le t \le u < \infty$:

$$B_u = B_t + \int_t^u dB_v \qquad (1\text{-}15).$$

Using the fact that a Brownian motion is a nonanticipating process,

i.e. $E[B_t | \mathcal{F}_t] = B_t$:

$$E[B_u - B_t | \mathcal{F}_t] = E[B_u | \mathcal{F}_t] - E[B_t | \mathcal{F}_t] = E[B_t + \int_t^u dB_v | \mathcal{F}_t] - B_t$$

$$E[B_u - B_t | \mathcal{F}_t] = E[B_t | \mathcal{F}_t] + E[\int_t^u dB_v | \mathcal{F}_t] - B_t$$

$$E[B_u - B_t | \mathcal{F}_t] = B_t + 0 - B_t = 0 \qquad (1\text{-}16),$$

or in other words:

$$E[B_u | \mathcal{F}_t] = E[B_t + \int_t^u dB_v | \mathcal{F}_t] = E[B_t | \mathcal{F}_t] + E[\int_t^u dB_v | \mathcal{F}_t] = B_t + 0$$

$$E[B_u | \mathcal{F}_t] = B_t \qquad (1\text{-}17),$$

which is a martingale condition.

□

Let $(B_{t\in[0,\infty)})$ be a standard Brownian motion process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, a Brownian motion with drift $(X_{t\in[0,\infty)}) \equiv (\mu t + \sigma B_{t\in[0,\infty)})$ is not a continuous martingale with respect to the filtration $\mathcal{F}_{t\in[0,\infty)}$ and the probability measure $\mathbb{P}$.

Proof

By definition, $(X_{t\in[0,\infty)})$ is a nonanticipating process (i.e. $\mathcal{F}_{t\in[0,\infty)}$ - adapted process) with the finite mean $E[X_t] = E[\mu t + \sigma B_t] = \mu t < \infty$ for $\forall t \in [0,\infty)$ and $\mu \in \mathbb{R}$. For $\forall 0 \le t \le u < \infty$:

$$X_u = X_t + \int_t^u dX_v .$$

Using the fact that a Brownian motion with drift is a nonanticipating process, i.e. $E[X_t | \mathcal{F}_t] = X_t$:

$$E[X_u | \mathcal{F}_t] = E[X_t + \int_t^u dX_v | \mathcal{F}_t] = E[X_t | \mathcal{F}_t] + E[\int_t^u dX_v | \mathcal{F}_t]$$

$$E[X_u | \mathcal{F}_t] = X_t + \mu(u - t) \qquad (1\text{-}18),$$

which violates a martingale condition.

□

But one way to transform nonmartingales into martingales is to make the process driftless. In other words, eliminating the trend of the process which is sometimes called a detrending. Consider the following example. A detrended Brownian motion with drift defined as:

$$(X_{t\in[0,\infty)} - \mu t) \equiv (\mu t + \sigma B_{t\in[0,\infty)} - \mu t) \equiv (\sigma B_{t\in[0,\infty)}) \qquad (1\text{-}19),$$

is a continuous martingale with respect to the filtration $\mathcal{F}_{t\in[0,\infty)}$ and the probability measure $\mathbb{P}$.

Proof

For $\forall 0 \le t \le u < \infty$:

$$E[X_u - \mu u \,|\, \mathcal{F}_t] = E[(X_t - \mu t) + (\int_t^u dX_v - \mu \int_t^u dv) \,|\, \mathcal{F}_t]$$

$$E[X_u - \mu u \,|\, \mathcal{F}_t] = E[(X_t - \mu t) \,|\, \mathcal{F}_t] + E[(\int_t^u dX_v - \mu \int_t^u dv) \,|\, \mathcal{F}_t]$$

$$E[X_u - \mu u \,|\, \mathcal{F}_t] = X_t - \mu t + \mu(u-t) - \mu(u-t)$$

$$E[X_u - \mu u \,|\, \mathcal{F}_t] = X_t - \mu t,$$

which satisfies a martingale condition.

□

1.3.6 Martingale Asset Pricing

Most of financial asset prices are not martingales because they are not completely unpredictable and most financial time series have trends. Consider a stock price process $\{S_t; 0 \le t \le T\}$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ and let r be the risk-free interest rate. In a small time interval Δ, risk-averse investors expect S_t to grow at some positive rate. This can be written as under actual probability measure $\mathbb{P}$:

$$E^{\mathbb{P}}[S_{t+\Delta} \,|\, \mathcal{F}_t] > S_t.$$

This means that a stock price S_t is not martingale under $\mathbb{P}$ and with respect to $\mathcal{F}_t$. To be more precise, risk-averse investors expect S_t to grow at a rate greater than r because a stock is risky:

$$E^{\mathbb{P}}[e^{-r\Delta}S_{t+\Delta}\,|\,\mathcal{F}_t] > S_t\,.$$

The stock price discounted by the risk-free interest rate $e^{-r\Delta}S_{t+\Delta}$ is not martingale under $\mathbb{P}$ and with respect to $\mathcal{F}_t$. How can we convert a discounted stock price $e^{-r\Delta}S_{t+\Delta}$ into a martingale? First approach is to eliminate the trend. The trend in this case is the risk premium π which risk-averse investors demand for bearing extra amount of risk. If we can estimate π correctly, a discounted stock price $e^{-r\Delta}S_{t+\Delta}$ can be converted into a martingale by detrending:

$$E^{\mathbb{P}}[e^{-\pi\Delta}e^{-r\Delta}S_{t+\Delta}\,|\,\mathcal{F}_t] = E^{\mathbb{P}}[e^{-(r+\pi)\Delta}S_{t+\Delta}\,|\,\mathcal{F}_t] = S_t\,.$$

But this approach involves the rather difficult job of estimating the risk premium π and is not used in quantitative finance. Martingale asset pricing uses the second approach to convert non-martingales into martingales by changing the probability measure. We will try to find an equivalent probability measure $\mathbb{Q}$ (called risk-neutral measure) under which a discounted stock price becomes martingale:

$$E^{\mathbb{Q}}[e^{-r\Delta}S_{t+\Delta}\,|\,\mathcal{F}_t] = S_t \qquad (1\text{-}20).$$

1.3.7 Submartingales and Supermartingales

Definition 1.12 Submartingale Consider a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$. A rcll stochastic process $(X_t)_{t\in[0,\,T]}$ is said to be a submartingale with

respect to the filtration $\mathcal{F}_t$ and under the probability measure $\mathbb{P}$ if it satisfies the following conditions:

(1) X_t is nonanticipating.

(2) $E[|X_t|] < \infty$ for $\forall t \in [0,\ T]$. Finite mean condition.

(3) $E[X_u | \mathcal{F}_t] \geq X_t$ for $\forall u > t$.

Intuitively, a submartingale is a stochastic process with a positive (upward) trend. A submartingale gains or grows on average as time progresses.

Definition 1.13 Supermartingale Consider a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$. A rcll stochastic process $(X_t)_{t\in[0,\ T]}$ is said to be a supermartingale with respect to the filtration $\mathcal{F}_t$ and under the probability measure $\mathbb{P}$ if it satisfies the following conditions:

(1) X_t is nonanticipating.

(2) $E[|X_t|] < \infty$ for $\forall t \in [0,\ T]$. Finite mean condition.

(3) $E[X_u | \mathcal{F}_t] \leq X_t$ for $\forall u > t$.

Intuitively, a supermartingale is a stochastic process with a negative (downward) trend. A supermartingale loses or declines on average as time progresses. By definition, any martingale is a submartingale and a supermartingale.

1.3.8 Markov Processes: Structure on Conditional Probability

This section gives a brief introduction to a class of stochastic processes called Markov processes which impose a restriction on the conditional probabilities. This differs from martingales which impose a structure on conditional expectations.

Definition 1.14 Discrete time Markov chain Consider a discrete time stochastic process $(X_n)_{n\in\mathbb{N}}$ (i.e. $n = 0,1,2,...$) defined on a filtered probability space $(\Omega, \mathcal{F}_{n\in\mathbb{N}}, \mathbb{P})$ which takes values in a countable or a finite set E called a state space of the process. A realization X_n is said to be in state $i \in E$ at time n if $X_n = i$. An E-valued discrete time Markov chain is a stochastic process which satisfies for $\forall n \in \mathbb{N}$ and $\forall i, j \in E$:

$$\mathbb{P}(X_{n+1} = j \,|\, X_0, X_1, X_2, ..., X_n = i) = \mathbb{P}(X_{n+1} = j \,|\, X_n = i) \quad (1\text{-}21).$$

This is called a Markov property. Markov property means that the probability of a random variable X_{n+1} at time $n+1$ (tomorrow) being in a state j conditional on the entire history of the stochastic process $(X_0, X_1, X_2, ..., X_n)$ is equal to the probability of a random variable X_{n+1} at time $n+1$ (tomorrow) being in a state j conditional only on the value of a random variable at time n (today). In other words, the history (sample path) of the stochastic process $(X_0, X_1, X_2, ..., X_n)$ is of no importance in that the way this stochastic process evolved or the dynamics $(X_1 - X_0, X_2 - X_1, ...)$ does not mean a thing in terms of the conditional probability of the process. The only factor which influences the conditional probability of a random variable X_{n+1} at time $n+1$ (tomorrow) is the state of a random variable at time n (today).

The probability $\mathbb{P}(X_{n+1} = j \,|\, X_n = i)$ which is a conditional probability of moving from a state i at time n to a state j at time $n+1$ is called a one step transition probability. In the general cases, transition probabilities are dependent on the states and time such that $\forall m \neq n \in \mathbb{N}$:

$$\mathbb{P}(X_{n+1}=j\,|\,X_n=i)\neq\mathbb{P}(X_{m+1}=j\,|\,X_m=i)\qquad(1\text{-}22).$$

When transition probabilities are independent of time n, discrete time Markov chains are said to be time homogeneous or stationary.

Definition 1.15 Time homogeneous (stationary) discrete time Markov chain

Consider a discrete time stochastic process $(X_n)_{n\in\mathbb{N}}$ (i.e. $n=0,1,2,...$) defined on a filtered probability space $(\Omega,\mathcal{F}_{n\in\mathbb{N}},\mathbb{P})$ which takes values in a countable or a finite set E called a state space of the process. A realization X_n is said to be in state $i\in E$ at time n if $X_n=i$. An E-valued time homogeneous discrete time Markov chain is a stochastic process which satisfies for $\forall n\in\mathbb{N}$ and $\forall i,j\in E$:

$$\begin{aligned}\mathbb{P}(X_{n+1}=j\,|\,X_0,X_1,X_2,...,X_n=i)&=\mathbb{P}(X_{n+1}=j\,|\,X_n=i)\\&=\mathbb{P}(X_1=j\,|\,X_0=i)\\&=\mathbb{P}(j\,|\,i)\qquad(1\text{-}23).\end{aligned}$$

In other words, transition probabilities do not depend on time n and only depend on transition states from i to j. A matrix of transition probabilities $\mathbb{P}=\left\|\mathbb{P}(j\,|\,i)\right\|_{i,j\in E}$ is called a transition probability matrix:

$$\left\|\mathbb{P}(j\,|\,i)\right\|_{i,j\in E}=\begin{Vmatrix}\mathbb{P}(0|0) & \mathbb{P}(1|0) & \mathbb{P}(2|0) & \mathbb{P}(3|0) & \cdots\\ \mathbb{P}(0|1) & \mathbb{P}(1|1) & \mathbb{P}(2|1) & \mathbb{P}(3|1) & \cdots\\ \mathbb{P}(0|2) & \mathbb{P}(1|2) & \mathbb{P}(2|2) & \mathbb{P}(3|2) & \cdots\\ \vdots & \vdots & \vdots & \vdots & \\ \mathbb{P}(0|i) & \mathbb{P}(1|i) & \mathbb{P}(2|i) & \mathbb{P}(3|i) & \cdots\\ \vdots & \vdots & \vdots & \vdots & \end{Vmatrix}\qquad(1\text{-}24).$$

Transition probabilities $\mathbb{P}(j|i)$ satisfy the following conditions:

(1) $\mathbb{P}(j|i) \geq 0$ for $\forall i, j \in E$.

(2) $\sum_{j \in E} \mathbb{P}(j|i) = 1$ for $\forall i \in E$.

Condition (2) guarantees the occurrence of a transition including a case in which the state remains unchanged.

Proposition 1.1 Defining a discrete time Markov chain An E-valued general discrete time Markov chain $(X_n)_{n \in \mathbb{N}}$ is completely defined if it satisfies the following conditions:

(1) All transition probabilities $\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n)$ are known.

(2) The probability distribution of the initial (i.e. time 0) state of the Markov chain $\mathbb{P}(X_0 = i_0) = \mathbb{P}_0$ is known.

Proof

Consider obtaining the joint probability distribution of an E-valued general discrete time Markov chain $(X_n)_{n \in \mathbb{N}}$. From the definition of a conditional probability:

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_n = i_n)$$

$$= \mathbb{P}(X_n = i_n | X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-1} = i_{n-1})$$

$$\times \mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-1} = i_{n-1}) \quad (1\text{-}25).$$

Since $(X_n)_{n \in \mathbb{N}}$ is a Markov chain:

$$\mathbb{P}(X_n = i_n | X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-1} = i_{n-1})$$

$$= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}) = \mathbb{P}(i_n | i_{n-1}) \quad (1\text{-}26).$$

Joint probability can be calculated as:

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_n = i_n)$$

$$= \mathbb{P}(i_n \,|\, i_{n-1}) \mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-1} = i_{n-1})$$

$$= \mathbb{P}(i_n \,|\, i_{n-1}) \mathbb{P}(X_{n-1} = i_{n-1} \,|\, X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-2} = i_{n-2})$$

$$\times \mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-2} = i_{n-2})$$

$$= \mathbb{P}(i_n \,|\, i_{n-1}) \mathbb{P}(i_{n-1} \,|\, i_{n-2}) \mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, ..., X_{n-2} = i_{n-2})$$

$$= \mathbb{P}(i_n \,|\, i_{n-1}) \mathbb{P}(i_{n-1} \,|\, i_{n-2}) ... \mathbb{P}(i_2 \,|\, i_1) \mathbb{P}(i_1 \,|\, i_0) \mathbb{P}_0 \qquad (1\text{-}27)$$

□

Consider a transition probability of a time homogeneous discrete time Markov chain $(X_n)_{n \in \mathbb{N}}$ from a state i at time k (i.e. $X_k = i$) to a state j at time $k+n$. This is called a *n*-step transition probability and expressed as:

$$\mathbb{P}(X_{k+n} = j \,|\, X_k = i) = \mathbb{P}(X_n = j \,|\, X_0 = i) = \mathbb{P}^{(n)}(j \,|\, i) \qquad (1\text{-}28).$$

Proposition 1.2 *n* step transition probability matrix (a special case of Chapman-Kolmogorov equation) Consider a time homogeneous discrete time Markov chain $(X_n)_{n \in \mathbb{N}}$ defined on a filtered probability space $(\Omega, \mathcal{F}_{n \in \mathbb{N}}, \mathbb{P})$ which takes values in a countable or a finite set E called a state space of the process. Then, its *n*-step transition probability matrix from a state i at time k (i.e. $X_k = i$) to a state j at time $k+n$ is given by for $\forall k, n \in \mathbb{N}$ and $\forall i, j \in E$:

$$\mathbb{P}^{(n)}(j \,|\, i) = \mathbb{P}^n(j \,|\, i)$$

$$= \sum_{v \in E} \mathbb{P}^{(y)}(v \,|\, i) \mathbb{P}^{(z)}(j \,|\, v)$$

$$= \sum_{v \in E} \mathbb{P}^{y}(v|i)\mathbb{P}^{z}(j|v) \qquad (1\text{-}29).$$

where $y + z = n$ and $\mathbb{P}^{(0)}(j|i)$ is defined as:

$$\mathbb{P}^{(0)}(j|i) = \begin{cases} 1 & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases}.$$

Proof

When $n = 1$:

$$\mathbb{P}^{(1)}(j|i) = \mathbb{P}(j|i).$$

When $n = 2$:

$$\mathbb{P}^{(2)}(j|i) = \sum_{v \in E} \mathbb{P}(v|i)\mathbb{P}(j|v).$$

By induction:

$$\mathbb{P}^{(n+1)}(j|i) = \sum_{v \in E} \mathbb{P}(v|i)\mathbb{P}^{n}(j|v).$$

□

One interesting topic about this n step transition probability matrix is its asymptotic behavior as $n \to \infty$. As n becomes larger, the initial state i becomes less important and in the limit as $n \to \infty$, $\mathbb{P}^{n}(j|i)$ is independent of i. We recommend Karlin and Taylor (1975) for more details.

Definition 1.16 Markov Processes (Continuous time Markov chains) Consider a continuous time stochastic process $\left(X_{t \in [0,T]}\right)$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t \in [0,T]}, \mathbb{P})$ which takes values in $\mathbb{N}$ (for simplicity) called a state space of the process. $\left(X_{t \in [0,T]}\right)$ is said to be a time homogeneous Markov process if for $\forall h \in \mathbb{R}^{+}$ and $\forall i, j \in \mathbb{N}$:

$$\mathbb{P}_h(j|i) = \mathbb{P}(X_{t+h} = j|\mathcal{F}_t) = \mathbb{P}(X_{t+h} = j|X_t = i) \quad (1\text{-}30).$$

Markov property means that the probability of a random variable X_{t+h} at time $t+h$ (tomorrow) being in a state j conditional on the entire history of the stochastic process $\mathcal{F}_{[0,t]} \equiv X_{[0,t]}$ is equal to the probability of a random variable X_{t+h} at time $t+h$ (tomorrow) being in a state j conditional only on the value of a random variable at time t (today). In other words, the history (sample path) of the stochastic process $\mathcal{F}_{[0,t]}$ is of no importance in that the way this stochastic process evolved or the dynamics does not mean a thing in terms of the conditional probability of the process.

1.3.9 Sample Path Properties of Stochastic Processes

In this section, we give the formal definition of the continuity of a sample path of a stochastic process. Note that the continuity of path and continuity of time are different subjects. In other words, a continuous time stochastic process does not imply continuous stochastic process. For example, a Poisson process is a continuous time stochastic process, but it has discontinuous sample paths. There are different notions of continuity of sample paths which use different notions of convergence of random variables.

Definition 1.17 Continuous in mean square A real valued stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be continuous in mean square if for $\forall t \in [0,\ T]$:

$$\lim_{s\to t} E[|X_s - X_t|^2] = 0 \quad (1\text{-}31).$$

Continuity in mean square implies continuity in probability following Chebyshev's inequality.

Definition 1.18 Continuous in probability A real valued stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be continuous in probability if for $\forall t \in [0,\ T]$ and every $\varepsilon \in \mathbb{R}^+$:

$$\lim_{s\to t} \mathbb{P}\left(\left|X_s - X_t\right| > \varepsilon\right) = 0 \qquad (1\text{-}32),$$

or equivalently:

$$\lim_{s\to t} \mathbb{P}\left(\left|X_s - X_t\right| \le \varepsilon\right) = 1 .$$

Intuitively speaking, continuity in probability means that the probability of X_s getting closer to X_t rises (and eventually converges to 1) as s approaches t. For example, a Brownian motion process is continuous in mean square and continuous in probability although its proof is not that easy. But it turns out that the above definitions of continuity are too loose because a Poisson process can be proven to be continuous in mean square and probability[1]. Therefore, a more strict definition of continuity is used for the definition of a continuity of a sample path of a stochastic process.

Definition 1.19 Continuous stochastic process A real valued nonanticipating stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be (almost surely) continuous if a sample path of the process $(X_{t\in[0,T]}(\omega))$ is almost

[1] Consult Karlin and Taylor (1975) for details.

surely a continuous function for $\forall t \in [0,\ T]$. In other words, a sample path of the process $(X_{t\in[0,T]}(\omega))$ satisfies for $\forall t \in [0,\ T]$:

(1) Right limit of the process as s approaches t from the above (right hand side) exists, $\lim_{s\to t, s>t} X_s = X_{t+}$. Left limit of the process as s approaches t from the below exists, $\lim_{s\to t, s<t} X_s = X_{t-}$.

(2) $X_{t+} = X_{t-} = X_t$.

This means that a continuous stochastic process is a right continuous and left continuous stochastic process.

Definition 1.20 Right continuous with left limit (rcll) stochastic processes A real valued nonanticipating stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be a rcll stochastic process if for $\forall t \in [0,\ T]$:

(1) Right limit of the process as s approaches t from the above (right hand side) exists, $\lim_{s\to t, s>t} X_s = X_{t+}$. Left limit of the process as s approaches t from the below exists, i.e. $\lim_{s\to t, s<t} X_s = X_{t-}$.

(2) $X_{t+} = X_t$.

In other words, only the right continuity is needed. Apparently, a continuous stochastic process implies a rcll stochastic process, but the reverse is not true. What we encounter in finance literatures are all rcll stochastic processes for the modeling of stock price dynamics.

Suppose t is a discontinuity point. The jump of the stochastic process at t is defined as:

$$\Delta X_t = X_t - X_{t-}.$$

A rcll nonanticipating stochastic process $(X_{t\in[0,T]})$ can have a finite number of large jumps and countable number (possibly infinite) of small jumps.

Definition 1.21 Total variation of a function Let $f(x)$ be a bounded function defined in the interval $[a,b]$:

$$f(x):[a,b]\to\mathbb{R}.$$

The interval can be infinite, i.e. $[-\infty,\infty]$. Consider partitioning the interval $[a,b]$ with the points:

$$a = x_0 < x_1 < x_2 ... x_{n-1} < x_n = b.$$

Then, the total variation of a function $f(x)$ is defined by:

$$T(f) = \sup \sum_{i=1}^{n} \left| f(x_i) - f(x_{i-1}) \right| \quad (1\text{-}33),$$

where sup indicates a supremum (least upper bound).

Definition 1.22 Function of finite variation A function $f(x)$ on the interval $[a,b]$ is said to be a function of finite variation, if its total variation on the interval $[a,b]$ is finite:

$$T(f) = \sup \sum_{i=1}^{n} \left| f(x_i) - f(x_{i-1}) \right| < \infty \quad (1\text{-}34).$$

Proposition 1.3 Every bounded increasing or decreasing function is of finite variation on the interval $[a,b]$.

Proof

Consider an increasing function $f(x)$ on the interval $[a,b]$. By its definition, for $\forall i$:

$$f(x_i) - f(x_{i-1}) \geq 0 .$$

And:

$$T(f) = \sup\left\{ f(x_n) - f(x_{n-1}) + f(x_{n-1}) - f(x_{n-2}) + \ldots + f(x_2) - f(x_1) + f(x_1) - f(x_0) \right\}$$

$$T(f) = \sup\left\{ f(x_n) - f(x_0) \right\}$$

$$T(f) = \sup\left\{ f(b) - f(a) \right\},$$

which is finite because $f(x)$ is bounded:

$$-\infty < f(a), f(b) < \infty .$$

□

Definition 1.23 Total variation of a stochastic process Consider a real valued stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$. Partition the time interval $[0,T]$ with the points:

$$0 = t_0 < t_1 < t_2 \ldots < t_{n-1} < t_n = T .$$

Then, the total variation of a stochastic process $(X_{t\in[0,T]})$ on the time interval $[0,T]$ is defined by:

$$T(X) = \sup \sum_{i=1}^{n} \left| X(t_i) - X(t_{i-1}) \right| \quad \text{(1-35)},$$

where sup indicates a supremum (least upper bound).

Definition 1.24 Stochastic process of finite variation A real valued stochastic process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ on the interval $[0,T]$ is said to be a stochastic process of finite variation, if the total variation on the interval $[0,T]$ of a sample path of the process is finite with probability 1:

$$\mathbb{P}(T(X) = \sup \sum_{i=1}^{n} \left| X(t_i) - X(t_{i-1}) \right| < \infty) = 1 \qquad (1\text{-}36).$$

1.4 Lévy Processes

In this section, some theorems and propositions are presented without proofs. This is obviously because it is not the goal of this sequel to prove theorems and some proofs are beyond what we need while consuming too many pages. But for those inquisitive readers, we provide the information about where to look for more details of the subjects and their proofs. Our goal is to present the foundations of the mathematics of Lévy processes as simple as possible.

1.4.1 Definition of Lévy Processes

Definition 1.25 Lévy processes A real valued stochastic process $(X_{t\in[0,\infty)})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ is said to be a Lévy process on $\mathbb{R}$ if it satisfies the following conditions:

(1) Its increments are independent. In other words, for $0 \le t_1 < t_2 < ... < t_n < \infty$:

$$\mathbb{P}(X_{t_0} \bigcap X_{t_1} - X_{t_0} \bigcap X_{t_2} - X_{t_1} \bigcap ... \bigcap X_{t_n} - X_{t_{n-1}})$$

$$= \mathbb{P}(X_{t_0})\mathbb{P}(X_{t_1} - X_{t_0})\mathbb{P}(X_{t_2} - X_{t_1})...\mathbb{P}(X_{t_n} - X_{t_{n-1}}).$$

(2) Its increments are stationary (time homogeneous): i.e. for $h \ge 0$, $X_{t+h} - X_t$ has the same distribution as X_h. In other words, the distribution of increments does not depend on t.

(3) $\mathbb{P}(X_0 = 0) = 1$. The process starts from 0 almost surely (with probability 1).

(4) The process is stochastically continuous: $\forall \varepsilon > 0$, $\lim_{h \to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \varepsilon) = 0$.

(5) Its sample path (trajectory) is rcll almost surely.

Processes satisfying the conditions (1) and (2) are called processes with independent and stationary increments. Independent increments condition is a restriction on the probabilistic dependence structure of increments among the past, present, and future. Stationary increments condition is a restriction on the shape of the distribution of increments among the past, present, and future.

The condition (4) which is implied by the conditions (2), (3), and (5) does not imply the continuous sample paths of the process. It means that if we are at time t, the probability of a jump at time t is zero because there is no uncertainty about the present. Jumps occur at random times. This property is called stochastic continuity or continuity in probability.

Rcll condition (5) does not need to be imposed. This is because a real valued Lévy process in law which is a process satisfying conditions (1), (2), (3), and (4) is modified to a Lévy process which satisfies the conditions (1), (2), (3), (4) , and (5) (theorem 11.5 of Sato (1999)). In other words, the condition (5) results from the conditions (1), (2), (3), and (4) through a theorem.

We saw the definition of a Lévy process. Next we discuss infinite divisibility of a distribution. It turns out that we cannot separate Lévy processes from infinitely divisible distributions because Lévy processes are generated by infinitely divisible distributions.

1.4.2 Infinitely Divisible Random Variable and Distribution

Definition 1.26 Infinitely divisible random variable and distribution A real valued random variable X with the probability density function $\mathbb{P}(x)$ is said to be infinitely divisible if for $\forall n \in \mathbb{N}$ there exist $i.i.d.$ random variables $X_1, X_2, ..., X_n$ satisfying:

$$X \underline{d} X_1 + X_2 + ... + X_n \quad (1\text{-}37).$$

$\mathbb{P}(x)$ is said to be an infinitely divisible distribution.

Definition 1.27 Characteristic function Let X be a random variable with its probability density function $\mathbb{P}(x)$. A characteristic function $\phi(\omega)$ with $\omega \in \mathbb{R}$ is defined as the Fourier transform of the probability density function $\mathbb{P}(x)$ using Fourier transform parameters $(a,b) = (1,1)$:

$$\phi(\omega) \equiv \mathcal{F}[\mathbb{P}(x)] \equiv \int_{-\infty}^{\infty} e^{i\omega x}\mathbb{P}(x)dx \equiv E[e^{i\omega x}] \quad (1\text{-}38).$$

In terms of a characteristic function, infinite divisibility is defined as follows.

Proposition 1.4 Infinitely divisible random variable and distribution A real valued random variable X with the probability density function $\mathbb{P}(x)$ and the characteristic function $\phi_X(\omega)$ is said to be infinitely divisible if for $\forall n \in \mathbb{N}$ there exist $i.i.d.$ random variables $X_1, X_2, ..., X_n$ each with a characteristic function $\phi_{X_i}(\omega)$ such that:

$$\phi_X(\omega) = (\phi_{X_i}(\omega))^n \text{ or } (\phi_X(\omega))^{1/n} = \phi_{X_i}(\omega) \quad (1\text{-}39).$$

$\mathbb{P}(x)$ is said to be an infinitely divisible distribution.

Proof

Consult Applebaum (2004) section 1.2.2.

Examples of infinitely divisible distributions include normal distributions on $\mathbb{R}$, gamma distributions on $\mathbb{R}$, α-stable distributions on $\mathbb{R}$, Poisson distributions on $\mathbb{R}$, compound Poisson distributions on $\mathbb{R}$, geometric distributions on $\mathbb{R}$, negative binomial distributions on $\mathbb{R}$, exponential distributions on $\mathbb{R}$. From Sato (1999), probability measures with bounded supports (e.g. uniform and binomial distributions) are not infinitely divisible.

Suppose that a random variable Y is drawn from a normal distribution with the mean μ and the variance σ^2:

$$\mathbb{P}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right\}.$$

Consider *i.i.d.* normal random variables $Y_1, Y_2, ..., Y_n$ with the mean μ/n and the variance σ^2/n. We learn in any undergraduate statistics courses that the sum of normal distributions is also normal with the linear mean and the linear variance:

$$E[Y_1 + Y_2 + ... + Y_n] = E[n(\mu/n)] = \mu,$$

$$Var[Y_1 + Y_2 + ... + Y_n] = Var[Y_1] + Var[Y_2] + ... + Var[Y_n] = n(\sigma^2/n) = \sigma^2.$$

Therefore, a normal distribution is an infinitely divisible distribution since it satisfies:

$$Y \overset{d}{=} Y_1 + Y_2 + ... + Y_n.$$

This can be shown more formally using characteristic functions. The characteristic function of a normal random variable $Y \sim N(\mu, \sigma^2)$ is:

$$\phi_Y(\omega) = \mathcal{F}_Y[\mathbb{P}(y)](\omega) = \int_{-\infty}^{\infty} e^{i\omega y}\mathbb{P}(y)dy = \exp(i\mu\omega - \frac{\sigma^2\omega^2}{2}).$$

The characteristic function for the *i.i.d. n* summands of *Y*, Y_i, can be computed as:

$$\phi_{Y_i}(\omega) = \mathcal{F}_{Y_i}[\mathbb{P}(y_i)](\omega) = \int_{-\infty}^{\infty} e^{i\omega y_i}\mathbb{P}(y_i)dy$$

$$\phi_{Y_i}(\omega) = \int_{-\infty}^{\infty} e^{i\omega y_i}\frac{1}{\sqrt{2\pi\sigma^2/n}}\exp\left\{-\frac{1}{2}\frac{(y_i-\mu/n)^2}{\sigma^2/n}\right\}dy$$

$$\phi_{Y_i}(\omega) = \exp\left\{i(\mu/n)\omega - \frac{(\sigma^2/n)\omega^2}{2}\right\}.$$

But, we can see the obvious relationship between $\phi_Y(\omega)$ and $\phi_{Y_i}(\omega)$:

$$\phi_{Y_i}(\omega) = (\phi_Y(\omega))^{1/n} = \left\{\exp(i\mu\omega - \frac{\sigma^2\omega^2}{2})\right\}^{1/n} = \exp\left\{i(\mu/n)\omega - \frac{(\sigma^2/n)\omega^2}{2}\right\}.$$

This proves that the *i.i.d. n* summands of $Y \sim N(\mu,\sigma^2)$ are also normally distributed with the mean μ/n and the variance σ^2/n (because a characteristic function uniquely determines a probability distribution) and therefore proving the infinite divisibility of a normal distribution:

$$Y \overset{d}{=} \sum_{k=0}^{n-1} Y_k\ ,\ Y_k \sim i.i.d.\ N(\mu/n, \sigma^2/n).$$

Another example is a Poisson case. Suppose that Z is a Poisson random variable, i.e. $Z \sim Poisson(\lambda)$ with $\lambda \in \mathbb{R}^+$. Its characteristic function is:

$$\phi_Z(\omega) = \mathcal{F}_Z[\mathbb{P}(z)](\omega) = \sum_{z=0}^{\infty}\frac{e^{-\lambda}\lambda^z}{z!}e^{i\omega z} = \exp[\lambda(e^{i\omega}-1)].$$

Many readers learned that the sum of Poisson random variables is also a Poisson random variable. Consider *i.i.d.* Poisson random variables $Z_1, Z_2, ..., Z_n$ with the intensity λ / n. Its characteristic function is:

$$\phi_{Z_i}(\omega) = \mathcal{F}_{Z_i}[\mathbb{P}(z_i)](\omega) = \sum_{z_i=0}^{\infty} \frac{e^{-\lambda/n}\lambda/n^{z_i}}{z_i!} e^{i\omega z_i} = \exp[\frac{\lambda}{n}(e^{i\omega}-1)].$$

But, we can see the obvious relationship between $\phi_Z(\omega)$ and $\phi_{Z_i}(\omega)$:

$$\phi_{Z_i}(\omega) = (\phi_Z(\omega))^{1/n} = [\exp\{\lambda(e^{i\omega}-1)\}]^{1/n} = \exp\{\frac{\lambda}{n}(e^{i\omega}-1)\}.$$

This proves that the *i.i.d.* n summands of $Z \sim Poisson(\lambda)$ are also Poisson distributed with the intensity λ / n (because a characteristic function uniquely determines a probability distribution) and therefore proving the infinite divisibility of a Poisson distribution:

$$Z \overset{d}{=} \sum_{k=0}^{n-1} Z_k, Z_k \sim i.i.d.\ Poisson(\lambda / n).$$

1.4.3 Relationship between Lévy Processes and Infinitely Divisible Distributions

Proposition 1.5 If $(X_{t\in[0,\infty)})$ is a real valued Lévy process on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$, then, X_t has an infinitely divisible distribution for $\forall t \in [0,T]$. This corresponds to the corollary 11.6 of Sato (1999).

Proof

Consider a realization X_t at time t. Partition the time t into $n \in \mathbb{N}$ intervals using $t_i = i(t/n)$:

$$t_0 = 0, t_1 = t/n, t_2 = 2t/n, ..., t_n = nt/n = t\ .$$

A realization X_t can be considered as a sum of $n \in \mathbb{N}$ increments:

$$X_t = (X_{t_1} - X_{t_0}) + (X_{t_2} - X_{t_1}) + ... + (X_{t_n} - X_{t_{n-1}})\ .$$

Note that all these increments are *i.i.d.* increments because $(X_{t\in[0,\infty)})$ is a real valued Lévy process. Let $\phi(\omega; X(t))$ be the characteristic function of X_t and $\phi(\omega\ ; X(t_i - t_{i-1}))$ be the characteristic function of *i.i.d.* increments. Then, we have:

$$\phi(\omega; X(t)) = \phi(\omega; X(t_i - t_{i-1}))^n \qquad (1\text{-}40),$$

which follows from the property of a characteristic function: if $\{X_k, k = 1, ..., n\}$ are independent random variables, the characteristic function of their sum $X_1 + X_2 + ... + X_n$ is the product of their characteristic functions:

$$\phi_{X_1+X_2+...+X_n}(\omega) = \prod_{k=1}^{n} \phi_{X_k}(\omega) \qquad (1\text{-}41).$$

The equation (1-41) satisfies the proposition 1.4. Therefore, the distribution of Lévy process possesses infinite divisibility.

□

Thus, for every $t \in [0,\infty)$ and $h \in \mathbb{R}^+$, increments of a Lévy process $X_{t+h} - X_t$ follows an infinitely divisible law. Its converse is also true.

Proposition 1.6 For every infinitely divisible distribution $\mathbb{P}$ on $\mathbb{R}$, there exists a Lévy process $(X_{t\in[0,\infty)})$ on $\mathbb{R}$ whose distribution of increments $X_{t+h} - X_t$ is governed by $\mathbb{P}$. This is the corollary 11.6 of Sato (1999).

This proposition is extremely important because it means that there is one-to-one correspondence between an infinitely divisible distribution and a Lévy process. Some typical examples are illustrated in Table 1.2.

1.4.4 Lévy-Khinchin Representation

Lévy-Khinchin representation gives the characteristic functions of all infinitely divisible distributions. In other words, it gives the characteristic functions of all processes whose increments follow infinitely divisible distributions – Lévy processes.

Theorem 1.1 General Lévy-Khinchin representation of all infinitely divisible distributions Let $\mathbb{P}(x)$ be a real valued infinitely divisible distribution. Then, its characteristic function $\phi_X(\omega)$ is given by for $\forall\,\omega \in \mathbb{R}$:

$$\phi_X(\omega) = \exp\left(\psi_X(\omega)\right) \qquad (1\text{-}42),$$

where $\psi_X(\omega)$ called a characteristic exponent (or a log characteristic function) is given by:

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma\omega + \int_{-\infty}^{\infty}\left\{\exp(i\omega x) - 1 - i\omega x 1_D\right\}\ell(dx),$$

where $D = \{x : |x| \le 1\}$, A is a unique nonnegative constant (i.e. $A \in \mathbb{R}^+$), γ is a unique constant on $\mathbb{R}$, and ℓ is a unique measure on $\mathbb{R}$ satisfying:

$$\ell(\{0\}) = 0 \text{ and } \int_{-\infty}^{\infty}\min\{|x|^2, 1\}\ell(dx) < \infty.$$

Proof

Consult Sato (1999) p.37-p.47.

Theorem 1.2 Converse of Theorem 1.1 Consider a characteristic function $\phi_X(\omega)$ of a probability distribution $\mathbb{P}(x)$. If there exists a unique nonnegative constant A, a unique real valued constant γ, and a unique real valued measure ℓ satisfying $\ell(\{0\}) = 0$ and $\int_{-\infty}^{\infty} \min\{|x|^2, 1\}\ell(dx) < \infty$ which yield the characteristic function of the form:

$$\phi_X(\omega) = \exp\left(\psi_X(\omega)\right),$$

where:

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma\omega + \int_{-\infty}^{\infty} \left\{\exp(i\omega x) - 1 - i\omega x 1_D\right\} \ell(dx),$$

then, $\mathbb{P}(x)$ is infinitely divisible.

Proof

Consult Sato (1999) p.37-p.47.

Theorem 1.3 General Lévy-Khinchin representation of all Lévy processes Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, for any $\omega \in \mathbb{R}$, the characteristic function $\phi_X(\omega)$ of a Lévy process $(X_{t\in[0,\infty)})$ can be expressed as:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right) \quad (1\text{-}43),$$

where $\psi_X(\omega)$ called a characteristic exponent (or a log characteristic function) is given by:

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma\omega + \int_{-\infty}^{\infty} \left\{\exp(i\omega x) - 1 - i\omega x 1_D\right\} \ell(dx),$$

where $D = \{x : |x| \le 1\}$, A is a unique nonnegative constant (i.e. $A \in \mathbb{R}^+$), γ is a unique constant on $\mathbb{R}$, and ℓ is a unique measure on $\mathbb{R}$ satisfying:

$$\ell(\{0\}) = 0 \text{ and } \int_{-\infty}^{\infty} \min\{|x|^2, 1\}\ell(dx) < \infty .$$

Proof

Consult Sato (1999) p.37-p.47.

Theorem 1.4 Converse of Theorem 3.3 Consider a characteristic function $\phi_X(\omega)$ of a random variable X_t of a real valued stochastic process $(X_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. If there exists a unique nonnegative constant A, a unique real valued constant γ, and a unique real valued measure ℓ satisfying $\ell(\{0\}) = 0$ and $\int_{-\infty}^{\infty} \min\{|x|^2, 1\}\ell(dx) < \infty$ which yield the characteristic function of the form:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right),$$

where:

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma\omega + \int_{-\infty}^{\infty} \left\{\exp(i\omega x) - 1 - i\omega x 1_D\right\} \ell(dx),$$

then, $(X_{t\in[0,\infty)})$ is a real valued Lévy process (a process whose increments follow an infinitely divisible distribution).

Proof

Consult Sato (1999) p.37-p.47.

Definition 1.28 Lévy triplet (generating triplet) In theorem 1.3, a unique nonnegative constant A is called a Gaussian variance and a unique real valued

measure ℓ satisfying $\ell(\{0\}) = 0$ and $\int_{-\infty}^{\infty} \min\{|x|^2, 1\}\ell(dx) < \infty$ is called a Lévy measure. A unique real valued constant γ does not have any intrinsic meaning since it depends on the behavior of a Lévy measure ℓ. It turns out that these triplets uniquely define a Lévy process as a result of Lévy-Itô decomposition. This triplet is called a Lévy triplet and compactly written as (A, ℓ, γ).

Next, we present special cases of the general Lévy-Khinchin representation of all Lévy processes by restricting the behavior of the Lévy measure ℓ.

Theorem 1.5 Lévy-Khinchin representation of Lévy processes whose Lévy measure satisfies the additional condition $\int_{|x|\le 1}|x|\ell(dx) < 0$ Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ whose Lévy measure satisfies the additional condition $\int_{|x|\le 1}|x|\ell(dx) < \infty$. Then, for any $\omega \in \mathbb{R}$, the characteristic function $\phi_X(\omega)$ of a Lévy process $(X_{t\in[0,\infty)})$ can be expressed as:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right) \qquad (1\text{-}44),$$

where $\psi_X(\omega)$ called a characteristic exponent is given by from the equation (1-43):

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma\omega + \int_{-\infty}^{\infty}\left\{\exp(i\omega x) - 1 - i\omega x 1_{|x|\le 1}\right\}\ell(dx)$$

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\left(\gamma - \int_{|x|\le 1}|x|\,\ell(dx)\right)\omega + \int_{-\infty}^{\infty}\left\{\exp(i\omega x) - 1\right\}\ell(dx)$$

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma_0\omega + \int_{-\infty}^{\infty}\left\{\exp(i\omega x) - 1\right\}\ell(dx),$$

where $\gamma_0 \equiv \gamma - \int_{|x|\le 1} |x|\, \ell(dx)$ is a unique real valued constant called a drift, A is a unique nonnegative constant (i.e. $A \in \mathbb{R}^+$) called a Gaussian variance, and ℓ is a unique measure on $\mathbb{R}$ called a Lévy measure satisfying:

$$\ell(\{0\}) = 0 \ , \ \int_{-\infty}^{\infty} \min\{|x|^2, 1\}\, \ell(dx) < \infty \ , \ \int_{|x|\le 1} |x|\, \ell(dx) < \infty \ .$$

Theorem 1.6 Lévy-Khinchin representation of Lévy processes whose Lévy measure satisfies the additional condition $\int_{|x|>1} |x| l(dx) < \infty$ Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ whose Lévy measure satisfies the additional condition $\int_{|x|>1} |x|\, \ell(dx) < \infty$. Then, for any $\omega \in \mathbb{R}$, the characteristic function $\phi_X(\omega)$ of a Lévy process $(X_{t\in[0,\infty)})$ can be expressed as:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right) \qquad (1\text{-}45),$$

where $\psi_X(\omega)$ called a characteristic exponent (or a log characteristic function) is given by from the equation (1-43):

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma\omega + \int_{-\infty}^{\infty} \left\{ \exp(i\omega x) - 1 - i\omega x 1_{|x|\le 1} \right\} \ell(dx)$$

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma_1\omega + \int_{-\infty}^{\infty} \left\{ \exp(i\omega x) - 1 - i\omega x \right\} \ell(dx) \, ,$$

where γ_1 is a unique real valued constant called a center (identical to the mean), A is a unique nonnegative constant (i.e. $A \in \mathbb{R}^+$) called a Gaussian variance, and ℓ is a unique measure on $\mathbb{R}$ called a Lévy measure satisfying:

$$\ell(\{0\}) = 0 \ , \ \int_{-\infty}^{\infty} \min\{|x|^2, 1\}\, \ell(dx) < \infty \ , \ \int_{|x|>1} |x|\, \ell(dx) < \infty \ .$$

Consult Cont and Tankov (2004) page 83 for more details.

Next, we consider a Lévy-Khinchin representation for a subclass of Lévy processes called a finite variation Lévy process. For a detailed description of the concept of variation, read section 1.3.9.

Definition 1.29 Lévy process of finite variation A real valued Lévy process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ on the interval $[0,T]$ is said to be a Lévy process of finite variation, if the total variation on the interval $[0,T]$ of a sample path of the Lévy process is finite with probability 1:

$$\mathbb{P}(T(X) = \sup \sum_{i=1}^{n} \left| X(t_i) - X(t_{i-1}) \right| < \infty) = 1 \qquad (1\text{-}46).$$

Theorem 1.7 Lévy process of finite variation If $(X_{t\in[0,T]})$ is a real valued Lévy process of finite variation on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ on the interval $[0,T]$, then, its Lévy triplet (A, ℓ, b) satisfies:

$$A = 0 \text{ and } \int_{|x|\le 1} |x| \, \ell(dx) < \infty \qquad (1\text{-}47).$$

This corresponds to the proposition 3.9 of Cont and Tankov (2004) where its proof is given.

Theorem 1.8 Lévy-Khinchin representation for Lévy processes of finite variation Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process of finite variation with Lévy triplet $(A = 0, \ell, b)$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, for any $\omega \in \mathbb{R}$, its characteristic function $\phi_X(\omega)$ can be expressed as:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right) \qquad (1\text{-}48),$$

where $\psi_X(\omega)$ called a characteristic exponent is given by:

$$\psi_X(\omega) = ib\omega + \int_{-\infty}^{\infty} \left\{ \exp(i\omega x) - 1 - i\omega x 1_{|x|\leq 1} \right\} \ell(dx)$$

$$\psi_X(\omega) = i\left(b - \int_{|x|\leq 1} x\ell(dx) \right)\omega + \int_{-\infty}^{\infty} \left\{ \exp(i\omega x) - 1 \right\} \ell(dx),$$

Theorem 1.9 Approximation of Lévy processes by compound Poisson processes Every infinitely divisible distribution can be obtained as the weak limit of a sequence of compound Poisson distributions. This means that every Lévy process can be obtained as the weak limit of a sequence of compound Poisson random variables. In other words, every Lévy process can be approximated by a compound Poisson process. This theorem corresponds to corollary 8.8 of Sato (1999) where you can find its proof.

1.4.5 Lévy-Itô Decomposition of Sample Paths of Lévy Processes

Historically speaking, Paul Lévy proposed the original idea of Lévy-Itô decomposition and as a result of Lévy-Itô decomposition, Lévy-Khinchin representation was proposed. In other words, Lévy-Khinchin representation uses the results of Lévy-Itô decomposition. Later, both Lévy-Itô decomposition and Lévy-Khinchin representation were formally proven by Kiyoshi Itô.

Lévy-Itô decomposition basically states that every sample path of Lévy process can be represented as a sum of two independent processes: One is a continuous Lévy process and the other is a compensated sum of independent jumps. Obviously, a continuous Lévy process is a Brownian motion with drift. One trick is

that the jump component has to be a *compensated* sum of independent jumps because a sum of independent jumps at time t may not converge.

Theorem 1.10 General Lévy-Itô Decomposition of Sample Paths of Lévy Processes Consider a real valued Lévy process $(X_{t\in[0,\infty)})$ with Lévy triplet (A,ℓ,γ) defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$. Then, Lévy-Itô decomposition states:

(1) Let $J_X \equiv J(A,\omega)$ be the random jump measure at time t of a Lévy process $(X_{t\in[0,\infty)})$ which contains the information of the timing of jumps and the size of jumps and the jump size belongs to a Borel set, $\Delta X_t(\omega) = X_t(\omega) - X_{t-}(\omega) \in \mathcal{B}_{\mathbb{R}}$. Then, J_X has a Poisson distribution with the mean (intensity) $\ell(dx)dt$.

(2) Every sample path of Lévy process can be represented as a sum of a continuous Lévy process and a discontinuous Lévy process:

$$X_t(\omega) = X_t^C(\omega) + X_t^D(\omega) \qquad (1\text{-}49).$$

(3) The continuous part $X_t^C(\omega)$ has a Lévy triplet $(A,0,\gamma)$. This means that the continuous part is a Brownian motion with drift:

$$X_t^C(\omega) = \gamma t + \sqrt{A}B_t = \gamma t + \sigma B_t \qquad (1\text{-}50),$$

where $Var[\gamma t + \sqrt{A}B_t] = Var[\gamma t + \sigma B_t] = At = \sigma^2 t$. This is why A is called a Gaussian variance.

(4) The discontinuous part $X_t^D(\omega)$ is a Lévy process with Lévy triplet $(0,\ell<\infty,0)$ which is the definition of a compound Poisson process (we will discuss this soon).

The discontinuous part $X_t^D(\omega)$ can be decomposed into a large jumps part $X_t^L(\omega)$ and a limit as $\varepsilon \to 0$ of compensated small jumps part $\tilde{X}_t^S(\omega)$:

$$X_t^D(\omega) = X_t^L(\omega) + \lim_{\varepsilon \downarrow 0} \tilde{X}_t^S(\omega) \qquad (1\text{-}51).$$

We arbitrarily define large jumps as those with absolute size greater than 1 (i.e. this is completely arbitrary) and small jumps as those with absolute size between $\varepsilon \in \mathbb{R}^+$ and 1:

$$\Delta X_t^L(\omega) \equiv \left|\Delta X_t(\omega)\right| > 1 \text{ and } \Delta X_t^S(\omega) \equiv \varepsilon < \left|\Delta X_t(\omega)\right| < 1.$$

A large jumps part $X_t^L(\omega)$ is a sum of large jumps during the time interval $(0,t]$ and $X_t^L(\omega)$ is almost surely finite because Lévy processes have finite number of large jumps by the definition of the Lévy measure:

$$\mathbb{P}\left(X_t^L(\omega) < \infty\right) = \mathbb{P}\left(\sum_{0 \le r \le t} \Delta X_r^L(\omega) < \infty\right) = 1.$$

This implies that there is no convergence issue with respect to a large jumps part $X_t^L(\omega)$.

Next, consider a *non-compensated* small jumps part $X_t^S(\omega)$ which is simply a sum of small jumps during the time interval $(0,t]$:

$$X_t^S(\omega) = \sum_{0 \le r \le t} \Delta X_r^S(\omega) = \int_{s \in (0,t], \varepsilon < \Delta |x| < 1} x J_X(ds \times dx, \omega),$$

which is a compound Poisson process. The reason that we cannot use a *non-compensated* small jumps part $X_t^S(\omega)$ in Lévy-Itô decomposition is that in the limit $\varepsilon \downarrow 0$, $X_t^S(\omega)$ almost surely does not achieve convergence (i.e. $\lim_{\varepsilon \downarrow 0} X_t^S(\omega) = \infty$)

because Lévy processes can have infinitely many number of small jumps (i.e. Lévy measure ℓ can have a singularity at 0). Therefore, a *non-compensated* small jumps part $X_t^S(\omega)$ must be *compensated* by its mean $\ell(dx)ds$ (i.e. because $X_t^S(\omega)$ is Poisson process) as:

$$\tilde{X}_t^S(\omega) \equiv X_t^S(\omega) - \ell(\varepsilon < dx < 1)dt,$$

which makes $\tilde{X}_t^S(\omega)$ a martingale so that $\lim_{\varepsilon \downarrow 0} \tilde{X}_t^S(\omega)$ can now achieve convergence.

In other words, a compensated small jumps part $\tilde{X}_t^S(\omega)$ is a *compensated* sum of small jumps during the time interval $(0,t]$:

$$\tilde{X}_t^S = \sum_{0 \le r \le t} \Delta \tilde{X}_r^S = \int_{s \in (0,t], \varepsilon < \Delta |x| < 1} \{x J_X(\omega) - x\ell(dx)ds\},$$

which always converges.

(5) Two processes $X_t^C(\omega)$ (i.e. a Brownian motion with drift) and $X_t^D(\omega)$ (i.e. a compound Poisson process) are independent.

Proof

Consult Sato (1999) section 19.

1.4.6 Lévy Measure

Definition 1.30 Lévy measure of all Lévy processes Let $(X_{t \in [0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a filtered probability space $(\Omega, \mathcal{F}_{t \in [0,\infty)}, \mathbb{P})$. The Lévy measure ℓ of a Lévy process $(X_{t \in [0,\infty)})$ is defined as a unique positive measure on $\mathbb{R}$ which measures (counts) the expected (average) number of jumps per unit of time:

$$\ell(A) = E[\#\{t \in [0,1] : \Delta X_t = X_t - X_{t-} \neq 0, \Delta X_t \in A\}] \quad (1\text{-}52),$$

where $\Delta X_t \in A$ indicates that the jump size belongs to a set A and a set A is a member of Borel set. This definition is equivalent to stating that the Lévy measure ℓ of a Lévy process $(X_{t\in[0,\infty)})$ is defined as a unique positive measure on $\mathbb{R}$ which measures arrival rate of jumps per unit of time. We like to emphasize that a Lévy measure ℓ measures the expected number of jumps of all sizes (i.e. small jumps plus large jumps) per unit of time. This definition of Lévy measure ℓ is true regardless of the types of Lévy processes (i.e. it doesn't matter whether it is a finite activity Lévy processes or an infinite activity Lévy processes). For more details, consult Cont and Tankov (2004) section 3.3.

By definition, a Lévy measure ℓ satisfies the two conditions:

(1) $\ell(\{0\}) = 0$, i.e. the measure of an empty set is zero.

(2) $\int_{-\infty}^{\infty} \min\{|x|^2, 1\}\ell(dx) < \infty$, mathematicians write this as $\int_{-\infty}^{\infty} (|x|^2 \wedge 1)\ell(dx) < \infty$.

Condtion (2) is extremely important. When the jump size is greater than 1 (i.e. $|x| > 1$), the jump is defined as a large jump using the arbitrary truncation point of 1 (the choice of truncation point is of no importance), then, the condition (2) reduces to:

$$\int_{|x|>1} \ell(dx) < \infty \quad (1\text{-}53),$$

which tells that the expected number of large jumps per unit of time is finite. When the jump size is less than 1 (i.e. $|x| < 1$), the jump is defined as a small jump, then, the condition (2) reduces to:

$$\int_{|x|<1} |x|^2 \, \ell(dx) < \infty \quad (1\text{-}54),$$

which means that a Lévy measure must be square-integrable around the origin. This indicates the following definition of Lévy processes with respect to the number of small and large jumps.

Definition 1.31 Lévy processes Lévy processes have finite expected number of large jumps per unit of time. Lévy processes can have finite or infinite expected number of small jumps per unit of time.

Depending on the behavior of Lévy measures, Lévy processes can be classified into three types.

Definition 1.32 Lévy process with zero Lévy measure $\ell = 0$ Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. If the Lévy measure ℓ of this Lévy process is $\ell = 0$, $(X_{t\in[0,\infty)})$ is a Brownian motion (with drift). Zero Lévy measure $\ell = 0$ means that the Lévy process $(X_{t\in[0,\infty)})$ has no small or large jumps, in other words, sample paths of the process is continuous.

Definition 1.33 Finite Activity Lévy processes Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. A Lévy process $(X_{t\in[0,\infty)})$ is said to be a finite activity Lévy process if its Lévy measure ℓ has a finite integral:

$$\int_{-\infty}^{\infty} \ell(dx) < \infty \quad (1\text{-}55).$$

It is important to emphasize that the above condition implies that a finite activity Lévy process has a finite expected number of small jumps and a finite expected number of large jumps per unit of time:

$$\int_{|x|<1} \ell(dx) < \infty \text{ and } \int_{|x|>1} \ell(dx) < \infty \quad (1\text{-}56).$$

Definition 1.34 Infinite Activity Lévy processes Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A,ℓ,γ) defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$. A Lévy process $(X_{t\in[0,\infty)})$ is said to be an infinite activity Lévy process if its Lévy measure ℓ has an infinite integral:

$$\int_{-\infty}^{\infty} \ell(dx) = \infty \quad (1\text{-}57).$$

It is important to emphasize that the above condition implies that an infinite activity Lévy process has an infinite expected number of small jumps and a finite expected number of large jumps per unit of time:

$$\int_{|x|<1} \ell(dx) = \infty \text{ and } \int_{|x|>1} \ell(dx) < \infty \quad (1\text{-}58).$$

1.4.7 Classification of Lévy Processes: In Terms of Gaussian or Not

Definition 1.35 Gaussian Lévy Process A real valued Lévy process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$ is said to be a Gaussian Lévy process if it satisfies one of the following three equivalent conditions:

(1) $(X_{t\in[0,T]})$ is a Brownian motion (with drift). In other words, for $\forall t\in[0,T)$ and $\forall h\in\mathbb{R}^+$, its increments $X_{t+h}-X_t$ follow a normal distribution.

(2) Its sample path is continuous.

(3) Its Lévy measure is zero, $\ell = 0$.

Definition 1.36 Continuous Lévy Process (same as Gaussian Lévy Process) The only continuous Lévy process is a Gaussian Lévy process which is a Brownian motion (with drift).

Definition 1.37 Non-Gaussian Lévy Processes (same as jump Lévy Processes) A real valued Lévy process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be a non-Gaussian Lévy process if it satisfies one of the following three equivalent conditions:

(1) For $\forall t \in [0,T)$ and $\forall h \in \mathbb{R}^+$, its increments $X_{t+h} - X_t$ do not follow a normal distribution.

(2) Its sample path is discontinuous.

(3) Its Lévy measure is not zero, $\ell \neq 0$.

Definition 1.38 Purely non-Gaussian Lévy Processes (same as pure jump Lévy Processes) A real valued Lévy process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$ is said to be a purely non-Gaussian Lévy process if it satisfies either condition (1) or (2) (i.e. these conditions are equivalent):

(1) Its Gaussian variance A is zero, that is, its Lévy triplet is given by $(A = 0, \ell, b)$.

(2) The process $(X_{t\in[0,T]})$ does not contain a Brownian motion (with drift).

Consider a jump diffusion (JD) process which is a Brownian motion (with drift) plus a compound Poisson process. Firstly, a JD process obviously contains a Brownian motion (with drift), thus, it is not a purely non-Gaussian Lévy process. But the increments $X_{t+h} - X_t$ of a JD process do not follow a normal distribution

(because of the addition of compound Poisson process), thus, it is not a Gaussian Lévy process. Therefore, a JD process is a non-Gaussian Lévy process. Variance process is a purely non-Gaussian Lévy process because it does not contain a Brownian motion (with drift).

1.4.8 Classification of Lévy Processes: In Terms of the Behavior of Lévy Measure

Definition 1.39 Lévy process with zero Lévy measure $\ell = 0$ Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. If the Lévy measure ℓ of this Lévy process is $\ell = 0$, $(X_{t\in[0,\infty)})$ is a Brownian motion (with drift). Zero Lévy measure $\ell = 0$ means that the Lévy process $(X_{t\in[0,\infty)})$ has no small or large jumps, in other words, sample paths of the process is continuous.

Definition 1.40 Finite Activity Lévy processes Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. A Lévy process $(X_{t\in[0,\infty)})$ is said to be a finite activity Lévy process if its Lévy measure ℓ has a finite integral:

$$\int_{-\infty}^{\infty} \ell(dx) < \infty .$$

It is important to emphasize that the above condition implies that a finite activity Lévy process has a finite expected number of small jumps and a finite expected number of large jumps per unit of time:

$$\int_{|x|<1} \ell(dx) < \infty \text{ and } \int_{|x|>1} \ell(dx) < \infty .$$

Definition 1.41 Infinite Activity Lévy processes Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A,ℓ,γ) defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$. A Lévy process $(X_{t\in[0,\infty)})$ is said to be an infinite activity Lévy process if its Lévy measure ℓ has an infinite integral:

$$\int_{-\infty}^{\infty} \ell(dx) = \infty .$$

It is important to emphasize that the above condition implies that an infinite activity Lévy process has an infinite expected number of small jumps and a finite expected number of large jumps per unit of time:

$$\int_{|x|<1} \ell(dx) = \infty \text{ and } \int_{|x|>1} \ell(dx) < \infty .$$

1.4.9 Classification of Lévy Processes: In Terms of the Total Variation of Lévy Process

Definition 1.42 Lévy processes of finite variation A real valued Lévy process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$ on the interval $[0,T]$ is said to be a Lévy process of finite variation, if the total variation on the interval $[0,T]$ of a sample path of the Lévy process is finite with probability 1:

$$\mathbb{P}\left(T(X) = \sup \sum_{i=1}^{n} \left| X(t_i) - X(t_{i-1}) \right| < \infty \right) = 1 .$$

Theorem 1.11 Lévy processes of finite variation If $(X_{t\in[0,T]})$ is a real valued Lévy process of finite variation on the interval $[0,T]$ defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$, then, its Lévy triplet (A,ℓ,γ) satisfies:

$$A=0 \text{ and } \int_{|x|<1}|x|\,\ell(dx)<\infty .$$

This corresponds to the proposition 3.9 of Cont and Tankov (2004) where its proof is given. Also consult a Sato's monograph in Barndorff-Nielsen et al (2001) page 6.

Lévy-Khinchin representation for Lévy processes of finite variation is given by theorem 3.8.

Definition 1.43 Lévy processes of infinite variation A real valued Lévy process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$ on the interval $[0,T]$ is said to be a Lévy process of infinite variation, if the total variation on the interval $[0,T]$ of a sample path of the Lévy process is infinite with probability 1:

$$\mathbb{P}\left(T(X)=\sup\sum_{i=1}^{n}|X(t_i)-X(t_{i-1})|=\infty\right)=1.$$

Theorem 1.12 Lévy processes of infinite variation If $(X_{t\in[0,T]})$ is a real valued Lévy process of infinite variation on the interval $[0,T]$ defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$, then, its Lévy triplet (A,ℓ,γ) satisfies:

$$A\neq 0 \text{ or } \int_{|x|<1}|x|\,\ell(dx)=\infty .$$

Consult a Sato's monograph in Barndorff-Nielsen et al (2001) page 6.

1.4.10 Classification of Lévy Processes: In Terms of the Properties of Lévy Triplet by Sato (1999)

Definition 1.44 Lévy processes of type A, type B, and type C Consider a real valued Lévy process $(X_{t\in[0,T]})$ with the Lévy triplet (A,ℓ,γ) on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$ on the time interval $[0,T]$.

(1) $(X_{t\in[0,T]})$ is said to be a type A Lévy process, if the Lévy triplet (A,ℓ,γ) satisfies:

$$A=0 \text{ and } \int_{-\infty}^{\infty}\ell(dx)<\infty .$$

In other words, a type A Lévy process is a purely non-Gaussian finite activity Lévy process whose sample paths have a finite number of small and large jumps in any finite time interval. A compound Poisson process is a typical example of a type A Lévy process.

(2) $(X_{t\in[0,T]})$ is said to be a type B Lévy process, if the Lévy triplet (A,ℓ,γ) satisfies:

$$A=0,\ \int_{-\infty}^{\infty}\ell(dx)=\infty, \text{ and } \int_{|x|<1}|x|\,\ell(dx)<\infty .$$

In other words, a type B Lévy process is a purely non-Gaussian infinite activity Lévy process of finite variation whose sample paths have an infinite number of small jumps and a finite number of large jumps in any finite time interval. A variance process is one example of a type B Lévy process.

(3) $(X_{t\in[0,T]})$ is said to be a type C Lévy process, if the Lévy triplet (A,ℓ,γ) satisfies:

$$A\neq 0 \text{ or } \int_{|x|<1}|x|\,\ell(dx)=\infty .$$

In other words, a type C Lévy process is a Lévy process of infinite variation. A Brownian motion (with drift) is a typical example of a type C Lévy process. These definitions are from Sato's monograph in Barndorff-Nielsen et al (2001) page 6.

1.4.11 Classification of Lévy Processes: In Terms of the Sample Paths

Properties of Lévy Processes

Theorem 1.13 A Lévy process with continuous sample path Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. If the sample paths of $(X_{t\in[0,\infty)})$ are continuous with probability 1, then, $(X_{t\in[0,\infty)})$ is a Gaussian Lévy process.

This corresponds to theorem 11.7 of Sato (1999). Theorem 3.13 together with Lévy-Itô decomposition proves that a normal distribution is the only distribution which generates a real valued Lévy process with continuous sample paths.

Theorem 1.14 A Lévy process with a piecewise constant sample path Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. If and only if a sample path of $(X_{t\in[0,\infty)})$ is piecewise constant with probability 1, then, $(X_{t\in[0,\infty)})$ is a compound Poisson process.

Theorem 1.15 The Converse of theorem 1.14 Let $(X_{t\in[0,\infty)})$ be a compound Poisson process on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ whose Lévy triplet (A, ℓ, γ) satisfies $A = 0$, $\int_{-\infty}^{\infty} \ell(dx) < \infty$, and $\int_{|x|<1} |x| \, \ell(dx) < \infty$ (by the definition of a

compound Poisson process). Then, a sample path of a compound Poisson process is piecewise constant with probability 1.

Proof

Consult proposition 3.3 of Cont and Tankov (2004).

Definition 1.45 Increasing Lévy processes (Subordinators) Consider a real valued Lévy process $(X_{t\in[0,\infty)})$ with the Lévy triplet (A,ℓ,γ) on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$. A Lévy process $(X_{t\in[0,\infty)})$ is said to be an increasing Lévy process if sample paths of the process $(X_{t\in[0,\infty)})$ are non-decreasing with probability 1:

$$\mathbb{P}(X_{t+h}\geq X_t : h\in\mathbb{R}^+)=1 \qquad (1\text{-}59).$$

Theorem 1.16 Increasing Lévy processes (Subordinators) Let $(X_{t\in[0,\infty)})$ be an increasing Lévy process on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$. Then, its Lévy triplet (A,ℓ,γ) satisfies the following conditions:

(1) $A=0$. Gaussian variance is zero. In other words, sample paths of $(X_{t\in[0,\infty)})$ do not contain Brownian motion (with drift) parts (Lévy-Itô decomposition).

(2) $\int_{-\infty}^{0}\ell(dx)=0$. The Lévy measure of an increasing Lévy process is concentrated on positive half axis. In other words, an increasing Lévy process has no negative jumps.

(3) $\int_{|x|<1}|x|\,\ell(dx)<\infty$. But because we always have positive jumps, this becomes

$\int_{(0,1)} x\ell(dx) < \infty$. This means that an increasing Lévy process is a Lévy process of finite variation (theorem 1.11).

(4) $\gamma_0 \geq 0$. Sample paths of $(X_{t\in[0,\infty)})$ have nonnegative drift following Lévy-Itô decomposition.

For more details and the proofs, consult Sato (1999) page 137-142.

Definition 1.46 Symmetric Lévy processes Consider a real valued Lévy process $(X_{t\in[0,\infty)})$ with the Lévy triplet (A,ℓ,γ) on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$. A Lévy process $(X_{t\in[0,\infty)})$ is said to be symmetric if sample paths of the process $(X_{t\in[0,\infty)})$ satisfy:

$$(X_t) \underline{\underline{d}} (-X_t),$$

which means that (X_t) and $(-X_t)$ have identical distributions.

1.4.12 Lévy Processes as a Subclass of Markov Processes

Definition 1.47 Transition function Consider a continuous time nonanticipating stochastic process $(X_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$ which takes values in a measurable space $(B,\mathcal{B})$ (i.e. $B\in\mathcal{B}(\mathbb{R})$). $(B,\mathcal{B})$ is called a state space of the process and the process is said to be B - valued. Consider an increasing sequence of time $0\leq t\leq u\leq v<\infty$. A real valued transition function $\mathbb{P}_{t,v}(x,B)$ with $x\in\mathbb{R}$ and $B\in\mathcal{B}(\mathbb{R})$ is a mapping which satisfies the following conditions:

(1) $\mathbb{P}_{t,v}(x,B)$ is a probability measure which maps every fixed x into B.

(2) $\mathbb{P}_{t,v}(x,B)$ is $\mathcal{B}$ - measurable for every $B \in \mathcal{B}(\mathbb{R})$.

(3) $\mathbb{P}_{t,t}(x,B) = \delta(B)$.

(4) $\mathbb{P}_{t,v}(x,B) = \int_{\mathbb{R}} \mathbb{P}_{t,u}(x,dy)\mathbb{P}_{u,v}(y,B)$.

The condition (4) is called the Chapman-Kolmogorov identity. Chapman-Kolmogorov identity means that the transition probability $\mathbb{P}_{t,v}(x,B)$ of moving from a state x at time t to a state B at time v can be calculated as a sum (i.e. integral) of the product of the transition probabilities via an intermediate time $t \le u \le v$, i.e. $\mathbb{P}_{t,u}(x,dy)$ and $\mathbb{P}_{u,v}(y,B)$. In the general cases, transition functions are dependent on the states and time.

Definition 1.48 Time homogeneous (temporary homogeneous or stationary) transition function Consider an increasing sequence of time $0 \le t \le u \le v < \infty$. A real valued transition function $\mathbb{P}_{t,v}(x,B)$ with $x \in \mathbb{R}$ and $B \in \mathcal{B}(\mathbb{R})$ is said to be time homogeneous if it satisfies:

$$\mathbb{P}_{t,v}(x,B) = \mathbb{P}_{0,v-t}(x,B) = \mathbb{P}_{v-t}(x,B),$$

which indicates that the transition function $\mathbb{P}_{t,v}(x,B)$ of moving from a state x at time t to a state B at time v is equivalent to the transition function $\mathbb{P}_{0,v-t}(x,B)$ of moving from a state x at time 0 to a state B at time $v-t$. In other words, the transition function is independent of the time t and depends only on the interval of time $v-t$.

Definition 1.49 Chapman-Kolmogorov identity for the time homogeneous transition function Consider an increasing sequence of time $0 \le t \le u < \infty$. Chapman-Kolmogorov identity for the time homogeneous transition function is:

$$\int_{\mathbb{R}} \mathbb{P}_{0,t}(x, dy)\mathbb{P}_{0,u}(y, B) = \int_{\mathbb{R}} \mathbb{P}_{t}(x, dy)\mathbb{P}_{u}(y, B) = \mathbb{P}_{t+u}(x, B).$$

Definition 1.50 Markov Processes (less formal) Consider a continuous time nonanticipating stochastic process $(X_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, the process $(X_{t\in[0,\infty)})$ is said to be a Markov process if it satisfies, for every increasing sequence of time $0 < t_1 \le t_2 \le ... \le t_n \le t \le u < \infty$:

$$\mathbb{P}(X_u \,|\mathcal{F}_t) = \mathbb{P}(X_u \,|X_0, X_{t_1}, X_{t_2}, ..., X_{t_n}, X_t) = \mathbb{P}(X_u \,|X_t),$$

Informally, Markov property means that the probability of a random variable X_u at time $u \ge t$ (tomorrow) conditional on the entire history of the stochastic process $\mathcal{F}_{[0,t]} \equiv X_{[0,t]}$ is equal to the probability of a random variable X_u at time $u \ge t$ (tomorrow) conditional only on the value of a random variable at time t (today). In other words, the history (sample path) of the stochastic process $\mathcal{F}_{[0,t]}$ is of no importance in that the way this stochastic process evolved or the dynamics does not mean a thing in terms of the conditional probability of the process. This property is sometimes called a memoryless property.

Definition 1.51 Markov Processes (formal) Consider a continuous time nonanticipating stochastic process $(X_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ which takes values in a measurable space $(E, \mathcal{E})$. $(E, \mathcal{E})$ is called a state space of the process and the process is said to be E - valued. Then, the process

$(X_{t\in[0,\infty)})$ is said to be a Markov process if it satisfies, for an increasing sequence of time $0 \le t \le u \le v < \infty$ $0 < t \le u < \infty$:

$$E[X_v \,|\, \mathcal{F}_t] = E[X_v \,|\, X_t],$$

with the transition function:

$$\mathbb{P}_{t,v}(x, B) = \int_{\mathbb{R}} \mathbb{P}_{t,u}(x, dy)\mathbb{P}_{u,v}(y, B).$$

Definition 1.52 Spatially homogeneous transition function Consider an increasing sequence of time $0 \le t \le u \le v < \infty$. A real valued transition function $\mathbb{P}_{t,v}(x, B)$ with $x \in \mathbb{R}$ and $B \in \mathcal{B}(\mathbb{R})$ is said to be spatially homogeneous if it satisfies, for $\forall t, v, x, B, B - x \in \{y - x : y \in B\}$:

$$\mathbb{P}_{t,v}(x, B) = \mathbb{P}_{t,v}(0, B - x).$$

Theorem 1.17 Lévy processes Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, the transition functions of Lévy processes satisfy, for $0 \le t \le u \le v < \infty$:

$$\mathbb{P}_{t,v}(x, B) = \mathbb{P}_{0,v-t}(0, B - x),$$

in other words, Lévy processes are a subclass of Markov processes with the time homogeneous and spatially homogeneous transition functions. Its converse is also true.

Theorem 1.18 Relationship between Markov processes and Lévy processes Let $(X_{t\in[0,\infty)})$ be a real valued Markov process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, a Markov process becomes a Lévy process if it satisfies the following additional conditions:

(1) The process is stochastically continuous: $\forall \varepsilon > 0$, $\lim_{h \to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \varepsilon) = 0$.

(2) Its transition functions are time homogeneous and spatially homogeneous:

$$\mathbb{P}_{t,v}(x, B) = \mathbb{P}_{0,v-t}(0, B - x).$$

Proof

For the proofs of theorem 1.17 and 1.18 and for the more details about Lévy processes as a subclass of Markov processes, consult Sato (1999) section 10.

Therefore, we can state that all Lévy processes are Markov process, but the converse is not true.

It turns out that Lévy processes satisfy not only Markov property, but also strong Markov property.

Theorem 1.19 Strong Markov property of Lévy processes Let $(X_{t \in [0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a filtered probability space $(\Omega, \mathcal{F}_{t \in [0,\infty)}, \mathbb{P})$. Define a new stochastic process as:

$$(X_{t+h} - X_h : t \in [0,\infty), h \geq 0).$$

Then, a process $(X_{t+h} - X_h)$ is a Lévy process satisfying:

(1) $(X_{t+h} - X_h : t \in [0,\infty), h \geq 0) \underline{d} (X_{t \in [0,\infty)})$.

(2) $(X_{t+h} - X_h : t \in [0,\infty), h \geq 0)$ and $(X_t : 0 \leq t \leq h)$ are independent.

Proof

Consult Sato (1999) section 10.

1.4.13 Other Important Properties of Lévy Processes

Theorem 1.20 Linear transformation of Lévy processes Consider a real valued Lévy process $(X_{t\in[0,T]})$ on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$ whose Lévy triplet is (A,ℓ,b). Let c be a constant on $\mathbb{R}$. Then, a linear transformation of the original Lévy Process $(cX_{t\in[0,T]})$ is also a real valued Lévy process whose Lévy triplet (A_c,ℓ_c,b_c) is given by:

$$A_c = c^2 A,$$

$$\ell_c = \left[\frac{\ell}{c}\right]_{\mathbb{R}\setminus\{0\}} \text{ (i.e. a restriction of a measure } \ell c^{-1} \text{ to } \mathbb{R}\setminus\{0\}),$$

$$b_c = cb + \int_{-\infty}^{\infty} cx\left(c\mathrm{1}_{|x|\le 1} - x\mathrm{1}_{|x|\le 1}\right)\ell(dx).$$

This corresponds to proposition 11.10 of Sato (1999) where its proof is given.

Theorem 1.21 Independence of Lévy processes

Part 1) Consider a two dimensional real valued non-Gaussian Lévy process $(X_{t\in[0,T]},Y_{t\in[0,T]})$ on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$ whose Lévy triplet is $(A=0,\ell,b)$. Then, X_t and Y_t are independent (i.e. $\mathbb{P}(X_t|Y_t)=\mathbb{P}(X_t)$), if and only if the set $\{(x,y):xy=0\}$ contains the support of its Lévy measure ℓ. In other words, X_t and Y_t are independent, if and only if X_t and Y_t do not jump together.

Part 2) If X_t and Y_t are independent, then, for any arbitrary set E, the Lévy measure of X_t can be expressed as (i.e. when X_t jumps, Y_t never jumps):

$$\ell_X(E_X) = \ell_X(\{x:(x,0)\in E\}),$$

and the Lévy measure of Y_t can be expressed as:

$$\ell_Y(E_Y) = \ell_Y(\{y : (0, y) \in E\}).$$

And, the Lévy measure of a two dimensional independent non-Gaussian Lévy process $(X_{t\in[0,T]}, Y_{t\in[0,T]})$ is given by the sum of these two Lévy measures:

$$\ell(E) = \ell_X(E_X) + \ell_Y(E_Y).$$

This corresponds to proposition 5.3 of Cont and Tankov (2004) where its proof is given.

Theorem 1.22 Sums of independent Lévy processes If two Lévy processes $(X_{t\in[0,T]})$ and $(Y_{t\in[0,T]})$ with Lévy triplets (A_X, ℓ_X, b_X) and (A_Y, ℓ_Y, b_Y) are independent, then, their sum $X_t + Y_t$ is also a Lévy process. For more details and its proof, consult an example 4.1 of Cont and Tankov (2004).

1.5 Examples of Lévy Processes

1.5.1 Brownian Motion: The only Lévy Process with Continuous Sample Paths

Some people (non-mathematicians) believe that all Lévy processes have discontinuous sample paths, in other words, Lévy processes are equivalent to jump processes. But this belief is wrong. It is true to state that most (but one) Lévy processes have discontinuous sample paths. But you can carefully go through the definition of Lévy processes again and you'll notice that the only sample paths condition which Lévy processes must satisfy is being a right continuous with left limit (i.e. rcll) stochastic process. This in turn indicates that a Lévy process can have continuous sample paths because all continuous processes are rcll processes by

definition. It turns out that the only Lévy process having continuous sample paths is a Brownian motion.

Definition 1.53 Standard Brownian motion (Standard Wiener process) A standard Brownian motion $(B_{t\in[0,\infty)})$ is a real valued stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ satisfying:

(1) Its increments are independent. In other words, for $0 \le t_1 < t_2 < ... < t_n < \infty$:

$$\mathbb{P}(B_{t_0} \bigcap B_{t_1} - B_{t_0} \bigcap B_{t_2} - B_{t_1} \bigcap ... \bigcap B_{t_n} - B_{t_{n-1}})$$

$$= \mathbb{P}(B_{t_0})\mathbb{P}(B_{t_1} - B_{t_0})\mathbb{P}(B_{t_2} - B_{t_1})...\mathbb{P}(B_{t_n} - B_{t_{n-1}}).$$

(2) Its increments are stationary (time homogeneous): i.e. for $h \ge 0$, $B_{t+h} - B_t$ has the same distribution as B_h. In other words, the distribution of increments does not depend on t.

(3) $\mathbb{P}(B_0 = 0) = 1$. The process starts from 0 almost surely (with probability 1).

(4) $B_t \sim Normal(0, t)$. Its increments follow a Gaussian distribution with the mean 0 and the variance t.

Theorem 1.23 Standard Brownian motion process (Standard Wiener process) A standard Brownian motion process $(B_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ satisfies the following conditions:

(1) The process is stochastically continuous: $\forall \varepsilon > 0$, $\lim_{h\to 0} \mathbb{P}(\left|X_{t+h} - X_t\right| \ge \varepsilon) = 0$.

(2) Its sample path (trajectory) is continuous in t (i.e. continuous $\in$ rcll) almost surely.

Proof

Consult Karlin (1975). We have to remind you that this proof is not that simple.

As many of readers realize, the above definitions and conditions all appear to define Lévy processes. Let us put this into another words.

Definition 1.54 Standard Brownian motion as a Lévy process A standard Brownian motion $(B_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ is a Lévy process satisfying:

(1) $B_t \sim Normal(0,t)$. Its increments follow a Gaussian distribution with the mean 0 and the variance t.

(2) Its sample path (trajectory) is continuous in t (i.e. continuous $\in$ rcll) almost surely.

The condition (1) implies the condition (2). Or more generally, a Brownian motion with drift is an only Lévy process with continuous sample paths.

Definition 1.55 Brownian motion with drift Let $(B_{t\in[0,\infty)})$ be a standard Brownian motion process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, a Brownian motion with drift is a real valued stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ as:

$$(X_{t\in[0,\infty)}) \equiv (\mu t + \sigma B_{t\in[0,\infty)}),$$

where $\mu \in \mathbb{R}$ is called a drift and $\sigma \in \mathbb{R}^+$ is called a diffusion (volatility) parameter. A Brownian motion with drift satisfies the following conditions:

(1) Its increments are independent. In other words, for $0 \le t_1 < t_2 < ... < t_n < \infty$:

$$\mathbb{P}(X_{t_0} \bigcap X_{t_1} - X_{t_0} \bigcap X_{t_2} - X_{t_1} \bigcap ... \bigcap X_{t_n} - X_{t_{n-1}})$$

$$= \mathbb{P}(X_{t_0})\mathbb{P}(X_{t_1} - X_{t_0})\mathbb{P}(X_{t_2} - X_{t_1})...\mathbb{P}(X_{t_n} - X_{t_{n-1}}).$$

(2) Its increments are stationary (time homogeneous): i.e. for $h \geq 0$, $X_{t+h} - X_t$ has the same distribution as X_h. In other words, the distribution of increments does not depend on t.

(3) $X_t \equiv \mu t + \sigma B_t \sim Normal(\mu t, \sigma^2 t)$. Its increments follow a Gaussian distribution with the mean μt and the variance $\sigma^2 t$.

(4) Its sample path (trajectory) is continuous in t (i.e. continuous $\in$ rcll) almost surely.

Definition 1.56 Brownian motion with drift as a Lévy process A Brownian motion with drift $(X_{t\in[0,\infty)}) \equiv (\mu t + \sigma B_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ is a Lévy process satisfying:

(1) $X_t \equiv \mu t + \sigma B_t \sim Normal(\mu t, \sigma^2 t)$. Its increments follow a Gaussian distribution with the mean μt and the variance $\sigma^2 t$.

(2) Its sample path (trajectory) is continuous in t (i.e. continuous $\in$ rcll) almost surely.

Next, we define a Brownian motion with drift process in terms of the properties of its Lévy triplet (A, ℓ, γ) and sample paths.

Definition 1.57 Brownian motion with drift in terms of Lévy triplet (A, ℓ, γ)

Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a

filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, $(X_{t\in[0,\infty)})$ is a Brownian motion with drift, if it satisfies one of the following conditions:

(1) $\ell = 0$. The Lévy measure ℓ of this Lévy process is zero. Zero Lévy measure $\ell = 0$ means that the Lévy process $(X_{t\in[0,\infty)})$ has no small or large jumps (i.e. no jumps at all), in other words, sample paths of the process is continuous with probability 1.

(2) Its increments follow a Gaussian distribution.

Theorem 1.24 Sample paths properties of Brownian motion with drift

Consider a real valued Brownian motion with drift process $(X_{t\in[0,\infty)}) \equiv (\mu t + \sigma B_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ and let (A, ℓ, γ) be its Lévy triplet. Then, the sample paths of $(X_{t\in[0,\infty)})$ possess following properties:

(1) Sample paths are continuous with probability 1. In other words, the Lévy measure ℓ of the process is zero. Zero Lévy measure $\ell = 0$ means that a Brownian motion with drift process has no small or large jumps, therefore, no jumps at all.

(2) Sample paths are of infinite variation on any finite interval $[0,t]$. In other words, the total variation on any finite interval $[0,t]$ of a sample path of a Brownian motion with drift is infinite with probability 1 in the limit $n \to \infty$ (as the partition becomes finer and finer):

$$\mathbb{P}\left(\lim_{n\to\infty} T(X) = \lim_{n\to\infty} \sup \sum_{i=1}^{n} \left| X(t_i) - X(t_{i-1}) \right| = \infty \right) = 1 .$$

As we saw in section 1.4.9, the infinite variation property is equivalent to the nonzero Gaussian variance in terms of the Lévy triplet (A,ℓ,γ):

$$A \neq 0 .$$

Intuitively speaking, the infinite variation property means highly oscillatory sample paths.

(3) The quadratic variations of sample paths of Brownian motions with drift $(X_{t\in[0,\infty)})$ are finite on any finite interval $[0,t]$ and converge to $\sigma^2 t$ with probability 1 in the limit $n \to \infty$ (as the partition becomes finer and finer):

$$\mathbb{P}\left(\lim_{n\to\infty} T^2(X) = \limsup_{n\to\infty} \sum_{i=1}^{n} \left| X(t_i) - X(t_{i-1}) \right|^2 = \sigma^2 t < \infty \right) = 1 .$$

For more details and proofs about theorem 1.24, consult Sato (1999) page 22 – 28 and Karatzas and Shreve (1991) section 1.5 and 2.9. We also recommend Rogers and Williams (2000) chapter 1.

Consider a real valued Brownian motion with drift process $(X_{t\in[0,\infty)}) \equiv (\mu t + \sigma B_{t\in[0,\infty)})$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ and let (A,ℓ,γ) be its Lévy triplet. Its characteristic function can be obtained by two approaches. First approach is the direct use of the definition of a characteristic function (i.e. Fourier transform of the probability density function with Fourier transform parameters $(1,1)$):

$$\phi_{X_t}(\omega) \equiv \mathcal{F}\left[\mathbb{P}(x)\right] \equiv \int_{-\infty}^{\infty} e^{i\omega x} \mathbb{P}(x) dx$$

$$\phi_{X_t}(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{ -\frac{(x-\mu t)^2}{2\sigma^2 t} \right\} dx$$

$$\phi_{X_t}(\omega) = \exp(i\mu t\omega - \frac{\sigma^2 t\omega^2}{2}) \quad (1\text{-}60).$$

Second approach is the use of Lévy-Khinchin representation (theorem 1.1). Because we know the Lévy triplet of a Brownian motion with drift is given as $(A = \sigma^2, \ell = 0, \gamma = \mu)$, its characteristic exponent is given by:

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma\omega + \int_{-\infty}^{\infty} \{\exp(i\omega x) - 1 - i\omega x 1_D\}\, \ell(dx)$$

$$\psi_X(\omega) = -\frac{\sigma^2\omega^2}{2} + i\mu\omega,$$

thus, its characteristic function $\phi_X(\omega)$ is expressed as:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right),$$

$$\phi_{X_t}(\omega) = \exp(-\frac{\sigma^2 t\omega^2}{2} + i\mu t\omega).$$

1.5.2 Poisson Process

A Poisson process is a continuous time stochastic process with discontinuous sample paths. To be more precise, the sample paths of Poisson process is right continuous with left limit (i.e. rcll) step functions of jump size 1. It can be used as a building block for all Lévy processes.

Definition 1.58 Exponential random variable An exponential random variable X with a parameter $\lambda \in \mathbb{R}^+$ is a positive random variable whose probability density function is given, for $x \in \mathbb{R}^+$, by:

$$f_X(x) = \lambda e^{-\lambda x} \quad (1\text{-}61).$$

Its distribution function is for $x \in \mathbb{R}^+$:

$$F_X(x) = \Pr(X \le x) = 1 - e^{-\lambda x} \qquad (1\text{-}62).$$

Its mean and variance are:

$$E[X] = \frac{1}{\lambda} \text{ and } Var[X] = \frac{1}{\lambda^2}.$$

For example, the probability density function of an exponential random variable X with $\lambda = 0.01$ is plotted in Figure 1.2.

Theorem 1.25 Lack of memory of an exponential random variable If a random variable X is an exponential random variable, then, for $\forall a, b \in \mathbb{R}^+$:

$$\Pr\{X > a + b \,|\, X > b\} = \Pr\{X > a\}.$$

If X is a random arrival time of an event, the probability of $X > a + b$ given $X > b$ is the same as the probability of $X > a$. This concept of lack of memory has nothing to do with the concept of statistical independence. Let $A \equiv a + b$, $B \equiv X > b$, and $C \equiv X > a$. Then, the lack of memory property becomes:

$$\Pr\{A \,|\, B\} = \Pr\{C\}.$$

If two events A and B are independent:

$$\Pr\{A \,|\, B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} = \frac{\Pr\{A\}\Pr\{B\}}{\Pr\{B\}} = \Pr\{A\}.$$

Proof

$$\Pr\{X > a + b \,|\, X > b\} = \frac{\Pr\{X > a + b \cap X > b\}}{\Pr\{X > b\}} = \frac{\Pr\{X > a + b\}}{\Pr\{X > b\}}$$

$$\Pr\{X > a + b \,|\, X > b\} = \frac{1 - F_X(a + b)}{1 - F_X(b)} = \frac{1 - (1 - e^{-\lambda(a+b)})}{1 - F_X(b)} = \frac{e^{-\lambda a} e^{-\lambda b}}{1 - F_X(b)}$$

$$\Pr\{X > a + b \,|\, X > b\} = \frac{\{1 - F_X(a)\}\{1 - F_X(b)\}}{1 - F_X(b)} = 1 - F_X(a)$$

$$\Pr\{X > a+b \,|\, X > b\} = \Pr\{X > a\}$$

□

We will give one illustrating example. Suppose that the first arrival time of a large earthquake X is modeled as an exponential random variable with the mean of $1/\lambda = 100$ years. Using $a = 10$ years and $b = 40$ years, the lack of memory property becomes:

$$\Pr\{X > 50 \,|\, X > 40\} = \Pr\{X > 10\}.$$

This means that the probability that 40 years has past since the last large earthquake and there will be more than 10 years for the next large earthquake to hit (i.e. which is $\Pr\{X > 50 \,|\, X > 40\}$) is equal to the unconditional probability that there will be more than 10 years for the next large earthquake to hit (i.e. which is $\Pr\{X > 10\}$). In other words, an exponential random variable X does not remember the fact that the 40 years has past since the last large earthquake.

Definition 1.59 Poisson random variable A Poisson random variable N is a discrete random variable with a parameter $\lambda \in \mathbb{R}^+$ called an intensity whose probability mass function is given, for $\forall k \in \mathbb{N}$ (i.e. $k = 0, 1, 2, \ldots$), by:

$$\mathbb{P}(N = k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad (1\text{-}63).$$

Its mean and variance are:

$$E[N] = \lambda \text{ and } Var[X] = \lambda.$$

A Poisson random variable N is used for the purpose of counting the number of arrivals of an event in unit time interval. For example, if we model the number of

arrivals of small earthquakes in one year as a Poisson random variable with the intensity $\lambda = 10$, its probability mass function is given by:

$$\mathbb{P}(N = k) = \frac{e^{-10}10^k}{k!}.$$

Suppose we model the number of dogs we see in one day N as a Poisson random variable with the intensity $\lambda \in \mathbb{R}^+$. Then, its probability mass function is:

$$\mathbb{P}(N = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Next, the number of dogs we see in t days N_t can be models as a Poisson random variable with the intensity $\lambda t \in \mathbb{R}^+$. Then, its probability mass function is:

$$\mathbb{P}(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \qquad (1\text{-}64).$$

The intensity λt can be interpreted as the average number of arrivals of an event (i.e. seeing a dog in our case) in the time interval of $[0, t]$. Let T be the waiting time until the first arrival of an event. Then, the probability of zero arrival of an event in $[0, t]$ can be calculated as:

$$\mathbb{P}(T > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t}.$$

The distribution function of a waiting time T of the first arrival of an event in the interval $[0, t]$ can be expressed as:

$$F_T(t) = \mathbb{P}(T < t) = 1 - \mathbb{P}(T > t) = 1 - e^{-\lambda t} \qquad (1\text{-}65).$$

As you can see, the equation (1-65) is identical to the equation (1-62) which means that a waiting time T of the first arrival of an event in the interval $[0, t]$ is an exponential random variable with the parameter λ.

This idea can be extended. Suppose that an event arrives k times in the interval $[0,t]$ which is illustrated by Figure 1.5. Note in this case, T_i indicates the moment when i th event arrives. Waiting time of the arrival of the first event is $T_1 - 0 = T_1$ which means that the waiting time between the arrival of the second event and the first event can be expressed as $T_2 - T_1$ and the waiting time between the arrival of the i th event and the $(i-1)$ th event can be expressed as $T_i - T_{i-1}$. In this case, a Poisson random variable N_t with the intensity λt counts the number of arrival of an event in the interval $[0,t]$ which means that a Poison random variable is a discrete random variable or has discontinuous sample paths (i.e. $N_t \in \mathbb{N}$). And, each waiting time $T_i - T_{i-1}$ is an *i.i.d* exponential random variable with the parameter λ :

$$\mathbb{P}(T_i - T_{i-1}) = \lambda e^{-\lambda(T_i - T_{i-1})},$$

which indicates that $T_i - T_{i-1}$ is a continuous random variable because an event can arrive at any moment.

For example, suppose that the average number of dogs we see in one day (i.e. the intensity) is $\lambda = 12$:

$$E[N_1] = \lambda = 12 .$$

Then, each waiting time $T_i - T_{i-1}$ is an *i.i.d* exponential random variable with the parameter $\lambda = 12$ which means that the average waiting time to see a dog is 1/12 day which is equal to 2 hours:

$$E[T_i - T_{i-1}] = \frac{1}{\lambda} = \frac{1 \text{ day}}{12 \text{dogs}} = \frac{2 \text{ hours}}{1 \text{ dog}}.$$

Definition 1.60 Poisson process A Poisson process $(N_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ is a stochastic process which counts the number of random times T_k (i.e. $T_1, T_2, ..., T_k$) of the arrival of an event in the time interval $[0,t]$ defined as:

$$N_t = \sum_{k\geq 1} 1_{t\geq T_k} \qquad (1\text{-}66),$$

which means that the sample paths of $(N_{t\in[0,\infty]})$ are right continuous step functions of jump size equal to one (i.e. discontinuous sample paths). Its probability mass function follows a Poisson distribution with the parameter λt :

$$\mathbb{P}(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

And, each waiting time $T_i - T_{i-1}$ is an *i.i.d* exponential random variable with the parameter λ :

$$\mathbb{P}(T_i - T_{i-1}) = \lambda \exp\{-\lambda(T_i - T_{i-1})\},$$

which indicates that a Poisson process $(N_{t\in[0,\infty]})$ is a continuous time stochastic process because an event can arrive at any moment. For more details, consult Karlin and Taylor (1975) pages 22-26.

We present the important properties of a Poisson process here. Some of which can be easily seen by the simulated sample paths of a Poisson process which is shown in Figure 1.7. Note that a path 1 has 10 jumps, a path 2 has 7 jumps, and a path 3 has 13 jumps.

Theorem 1.26 Fundamental properties of Poisson processes A Poisson process $(N_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ satisfies the following conditions:

(1) Its increments are independent. In other words, for any increasing sequence of time $0 \le t_1 < t_2 < ... < t_n$:

$$\mathbb{P}(N_0 \bigcap N_{t_1} - N_0 \bigcap N_{t_2} - N_{t_1} \bigcap ... \bigcap N_{t_n} - N_{t_{n-1}})$$

$$= \mathbb{P}(N_0)\mathbb{P}(N_{t_1} - N_0)\mathbb{P}(N_{t_2} - N_{t_1})...\mathbb{P}(N_{t_n} - N_{t_{n-1}}).$$

(2) Its increments are stationary (time homogeneous): i.e. for $h \ge 0$, $N_{t+h} - N_t$ has the same distribution as N_h. In other words, the distribution of increments does not depend on t.

(3) $\mathbb{P}(N_0 = 0) = 1$. The process starts from 0 almost surely (with probability 1).

(4) The process is stochastically continuous: $\forall \varepsilon > 0$, $\lim_{h\to 0} \mathbb{P}(\left|N_{t+h} - N_t\right| \ge \varepsilon) = 0$.

(5) Its sample paths are 1) non-decreasing functions, 2) right continuous with left limit step functions in $\forall t \in [0,\infty]$, and 3) its jump (step) size is 1. Obviously, these sample paths properties are true almost surely.

(6) For $\forall t \in (0,\infty]$, $\mathbb{P}(N_t = k < \infty) = 1$. In other words, the number of arrivals of an event is almost surely finite for any $t > 0$ including an infinite time horizon (i.e. $t = \infty$).

Proof

Theorems (1), (2), and (5) are true by definition. For more details, consult Ross (1983) chapter2 and Cont and Tankov (2004) pages 48-52.

As you may notice, it is very obvious that a Poisson process is a Lévy process because the condition (1) – (5) of theorem 1.26 is the definition of Lévy process which is the definition 1.25).

Definition 1.61 Poisson process as a Lévy process A Poisson process $(N_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ is a Lévy process satisfying:

(1) N_t follows a Poisson distribution with the intensity λt :

$$\mathbb{P}(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

(2) Its sample paths are non-decreasing right continuous with left limit step functions of step size 1 in $\forall t \in [0,\infty]$ with probability 1.

The condition (1) implies the condition (2).

Theorem 1.27 Lévy process whose sample paths are non-decreasing right continuous with left limit step functions of step size 1 Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. If and only if the sample paths of $(X_{t\in[0,\infty)})$ are non-decreasing right continuous with left limit step functions of step size 1 in $\forall t \in [0,\infty]$ with probability 1, then, $(X_{t\in[0,\infty)})$ is a Poisson process.

Proof

Consult proposition 3.3 of Cont and Tankov (2004).

Definition 1.62 Poisson process in terms of Lévy triplet *(A, l, γ)* Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A, ℓ, γ) defined on a filtered

probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Then, $(X_{t\in[0,\infty)})$ is a Poisson process with the intensity $\lambda \in \mathbb{R}^+$, if it satisfies the following conditions:

(1) $A = 0$. Its Gaussian variance is zero (i.e. Lévy-Itô decomposition).

(2) $\gamma = 0$. Its drift is zero (i.e. Lévy-Itô decomposition).

(3) Its Lévy measure is given by:

$$\ell(x) = \lambda\delta(x-1),$$

where $\delta(x-1)$ is a Dirac's delta function satisfying:

$$\delta(x-1) = \begin{cases} \delta(0) & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}.$$

$\delta(x-1)$ is a pulse of unbounded height and zero width with a unit integral:

$$\int_{-\infty}^{\infty} \delta(x-1)dx = 1.$$

Note that a Dirac's delta function $\delta(x-1)$ is a jump size probability density function for Poisson processes because Poisson processes have only one type of jumps which are jumps of size 1. And not surprisingly, the integral of the Lévy measure of a Poisson process is the intensity parameter $\lambda \in \mathbb{R}^+$ because a Lévy measure $\ell(x)$ measures the arrival rate of jumps:

$$\int_{-\infty}^{\infty} \ell(x)dx = \int_{-\infty}^{\infty} \lambda\delta(x-1)dx = \lambda\int_{-\infty}^{\infty} \delta(x-1)dx = \lambda < \infty,$$

which is finite because the number of arrivals of an event is almost surely finite for any $t > 0$ including an infinite time horizon $t = \infty$[2].

[2] The number (6) of theorem 1.26

Theorem 1.28 Finite variation property of Poisson process If $(X_{t\in[0,T]})$ is a Poisson process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{P})$, then, it is a real valued Lévy process of finite variation on the interval $[0,T]$. And, its Lévy triplet (A, ℓ, γ) satisfies:

$$A = 0 \text{ and } \int_{|x|<1} |x| \, \ell(dx) < \infty ,$$

which are finite variation conditions for Lévy processes. Zero Gaussian variance $A = 0$ of a Poisson process is obvious and the second condition for a Poisson process is satisfied because there is no need to truncate jumps and jump sizes are all one:

$$\int_{|x|<1} |x| \, \ell(dx) = \int_{x=1} x\lambda\delta(x-1)dx = \lambda \int_{x=1} x\delta(x-1)dx$$

$$\int_{|x|<1} |x| \, \ell(dx) = \lambda < \infty .$$

Finite variation property of a Poisson process can be easily guessed from its sample paths behavior. Stochastic processes of infinite variation have highly oscillatory sample paths such as a Brownian motion with drift. But, the sample paths of a Poisson process are rcll step functions of step size 1 which implies that a Poisson process is of finite variation:

$$\mathbb{P}(T(X) = \sup \sum_{i=1}^{n} |X(t_i) - X(t_{i-1})| < \infty) = 1 .$$

Table 1.3 summarizes the properties of a Poisson process.

Next, let's try to obtain the characteristic function of a Poisson process. Consider a Poisson process $(X_{t\in[0,\infty]})$ defined on a filtered probability space

$(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ and let (A, ℓ, γ) be its Lévy triplet. Its characteristic function can be obtained by two approaches. First approach is the direct use of the definition of a characteristic function:

$$\phi_X(\omega) \equiv \mathcal{F}\left[\mathbb{P}(X_t = k)\right] \equiv \sum_{k=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!} e^{i\omega k}$$

$$\phi_X(\omega) = \exp[t\lambda(e^{i\omega} - 1)].$$

Note that in the above, a series expansion is used instead of the Fourier integral $\int_{-\infty}^{\infty} e^{i\omega x}\mathbb{P}(x)dx$ because a Poisson distribution is discrete.

Second approach is the use of Lévy-Khinchin representation. Because we know the Lévy triplet of a Poisson process is given by:

$$\left(A = 0, \ell = \lambda\delta(x-1), \gamma = 0\right),$$

its characteristic exponent is given by:

$$\psi_X(\omega) = -\frac{A\omega^2}{2} + i\gamma\omega + \int_{-\infty}^{\infty}\left\{\exp(i\omega x) - 1 - i\omega x 1_D\right\}\ell(dx)$$

$$\psi_X(\omega) = \int_{-\infty}^{\infty}\left\{\exp(i\omega x) - 1\right\}\lambda\delta(x-1)dx$$

$$\psi_X(\omega) = \lambda(\mathrm{e}^{i\omega} - 1).$$

Note that the term $-i\omega x 1_D$ in the Lévy-Khinchin representation drops out for Poisson processes because there is no need to distinguish between large and small jumps for Poisson processes (i.e. all jumps sizes are 1). Thus, its characteristic function $\phi_X(\omega)$ is expressed as:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right),$$

$$\phi_X(\omega) = \exp\{t\lambda(\mathrm{e}^{i\omega} - 1)\}.$$

Theorem 1.29 Lévy measure of Poisson process A Lévy measure of Poisson process $(X_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ is given by:

$$\ell(x) = \lambda\delta(x-1).$$

And, a Poisson process is a finite activity Lévy process because its Lévy measure satisfies:

$$\int_{-\infty}^{\infty} \ell(x)dx = \lambda < \infty.$$

Proof

The Lévy measure $\ell(x)$ of a Poisson process must satisfy two important conditions by the definition. One is that its jump size density is concentrated at $x = 1$ and the other is that the average number of jumps per unit of time must equal to the intensity λ of a Poisson process. Thus, the only Lévy measure satisfying these two conditions is given by:

$$\ell(x) = \lambda\delta(x-1),$$

where $\delta(x-1)$ is a Dirac's delta function. $\delta(x-1)$ is a pulse of unbounded height and zero width with a unit integral:

$$\int_{-\infty}^{\infty} \delta(x-1)dx = 1.$$

Note that a Dirac's delta function $\delta(x-1)$ is a jump size probability density function for Poisson processes because Poisson processes have only one type of jumps which are jumps of size 1. The integral of the Lévy measure of a Poisson

process is the intensity parameter $\lambda \in \mathbb{R}^+$ because a Lévy measure $\ell(x)$ measures the arrival rate of jumps:

$$\int_{-\infty}^{\infty} \ell(x)dx = \int_{-\infty}^{\infty} \lambda\delta(x-1)dx = \lambda\int_{x=1} \delta(x-1)dx$$

$$\int_{-\infty}^{\infty} \ell(x)dx = \lambda < \infty,$$

which is finite because the number of arrivals of an event is almost surely finite for any $t > 0$ including an infinite time horizon $t = \infty$.

□

Theorem 1.30 Poisson process as a time homogeneous and spatially homogeneous Markov process A Poisson process $(X_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ is a Markov process with a time homogeneous and spatially homogeneous transition function.

Proof

For an increasing sequence of time $0 \le t_1 \le t_2 \le ... \le t \le t+h$:

$$\mathbb{P}(X_{t+h} - X_t = 1 \big| \mathcal{F}_t) = \mathbb{P}(X_{t+h} - X_t = 1 \big| X_0 = k_0, X_{t_1} = k_{t_1}, ..., X_t = k_t)$$

$$\mathbb{P}(X_{t+h} - X_t = 1 \big| \mathcal{F}_t) = \mathbb{P}(X_{t+h} - X_t = 1) = \mathbb{P}(X_h = 1) = \frac{e^{-\lambda h}(\lambda h)^1}{1!}$$

$$\mathbb{P}(X_{t+h} - X_t = 1 \big| \mathcal{F}_t) = e^{-\lambda h}\lambda h.$$

In the limit $h \downarrow 0$:

$$\lim_{h\downarrow 0} \mathbb{P}(X_{t+h} - X_t = 1 \big| \mathcal{F}_t) = \lambda h.$$

In other words, the transition function of a Poisson process is time homogeneous and spatially homogeneous:

$$\mathbb{P}_{t,t+h}(x_t, x_t + 1) = \mathbb{P}_{0,h}(0,1) = \lambda h .$$

□

Theorem 1.31 Nonmartingale property of Poisson process A Poisson process $(X_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ is not a martingale.

Proof

A Poisson process has a nonanticipating X_t (i.e. X_t is $\mathcal{F}_t$-adapted) by definition. A Poisson process has a finite mean by definition, for $\forall t \in [0, \infty]$:

$$E[|X_t|] = E[X_t] = \lambda t < \infty ,$$

which is the number (6) of theorem 1.26. For $\forall 0 \le t \le u \le \infty$:

$$E[X_u | \mathcal{F}_t] = E[X_t + X_{u-t} | \mathcal{F}_t] = E[X_t | \mathcal{F}_t] + E[X_{u-t} | \mathcal{F}_t]$$

$$E[X_u | \mathcal{F}_t] = X_t + E[X_{u-t} | \mathcal{F}_t] = X_t + \lambda(u - t)$$

$$E[X_u | \mathcal{F}_t] \neq X_t .$$

□

Theorem 1.32 Martingale property of Compensated Poisson process A compensated Poisson process $(\tilde{X}_{t\in[0,\infty]}) \equiv (X_{t\in[0,\infty]} - \lambda t)$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ is a martingale.

Proof

A compensated Poisson process has a nonanticipating $\tilde{X}_t \equiv X_t - \lambda t$ (i.e. $\tilde{X}_t$ is $\mathcal{F}_t$-adapted) by definition. A compensated Poisson process has a finite mean by definition, for $\forall t \in [0, \infty]$:

$$E[\left|\tilde{X}_t\right|] = E[X_t - \lambda t] = E[X_t] - \lambda t = \lambda t - \lambda t = 0 < \infty .$$

For $\forall 0 \le t \le u \le \infty$:

$$E[\tilde{X}_u \left| \mathcal{F}_t \right.] = E[\tilde{X}_t + \tilde{X}_{u-t} \left| \mathcal{F}_t \right.] = E[\tilde{X}_t \left| \mathcal{F}_t \right.] + E[\tilde{X}_{u-t} \left| \mathcal{F}_t \right.]$$

$$E[\tilde{X}_u \left| \mathcal{F}_t \right.] = \tilde{X}_t + E[\tilde{X}_{u-t} \left| \mathcal{F}_t \right.] = \tilde{X}_t + E[X_{u-t} - \lambda(u-t) \left| \mathcal{F}_t \right.]$$

$$E[\tilde{X}_u \left| \mathcal{F}_t \right.] = \tilde{X}_t + E[X_{u-t} \left| \mathcal{F}_t \right.] - \lambda(u-t)$$

$$E[\tilde{X}_u \left| \mathcal{F}_t \right.] = \tilde{X}_t + \lambda(u-t) - \lambda(u-t) = \tilde{X}_t .$$

□

A compensated Poisson process is not integer-valued and not a counting process because of the compensator λt. It behaves like a standard Brownian motion after rescaling it by $1/\sqrt{\lambda}$ since $\frac{1}{\sqrt{\lambda}}(\tilde{X}_{t\in[0,\infty]})$ satisfies:

$$E[\frac{\tilde{X}_t}{\sqrt{\lambda}}] = \frac{1}{\sqrt{\lambda}} E[X_t - \lambda t] = \frac{1}{\sqrt{\lambda}}(\lambda t - \lambda t) = 0$$

and:

$$Var[\frac{1}{\sqrt{\lambda}} \tilde{X}_t] = \frac{1}{\lambda} Var[\tilde{X}_t] = \frac{1}{\lambda} Var[X_t - \lambda t] = \frac{1}{\lambda} Var[X_t]$$

$$Var[\frac{1}{\sqrt{\lambda}} \tilde{X}_t] = \frac{1}{\lambda} \lambda t = t .$$

Theorem 1.33 Property of Compensated Poisson process Consider a finite time horizon $t \in [0,T]$. In the limit as the intensity of a compensated Poisson process $(\tilde{X}_{t\in[0,T]}) \equiv (X_{t\in[0,T]} - \lambda t)$ approaches infinity:

$$\lambda \in \mathbb{R}^+ \uparrow \infty,$$

a compensated Poisson process has an identical distribution to a standard Brownain motion:

$$\lim_{\lambda\uparrow\infty}(\tilde{X}_{t\in[0,T]}) \equiv (X_{t\in[0,T]} - \lambda t) \underline{\underline{d}} (B_{t\in[0,T]}).$$

For more details, consult Cont and Tankov (2004) page 53.

1.5.3 Compound Poisson Process

A compound Poisson process is a general case of a Poisson process. It is a continuous time stochastic process with discontinuous sample paths. But, unlike a Poisson process, a compound Poisson process is not necessarily an increasing process and jump size density can be of any type (thus, it is more general).

Definition 1.63 Compound Poisson process A compound Poisson process $(X_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ is a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ which is the sum of *i.i.d.* jumps X_i from the jump size density $f(x)$:

$$X_t = \sum_{i=1}^{N_t} X_i \text{ with } X_i \sim i.i.d.f(x) \quad (1\text{-}67),$$

where a Poisson process $(N_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ counts the number of random times T_k (i.e. $T_1, T_2, ..., T_k$) of the arrival of an event in the time interval $[0,t]$ defined as:

$$N_t = \sum_{k\geq 1} 1_{t\geq T_k} .$$

Note that a Poisson process $(N_{t\in[0,\infty]})$ and the jumps sizes $(X_i)_{i\geq 1}$ are assumed to be independent.

This means that if an event arrives $k \in \mathbb{R}^+$ times in the time interval $\forall t \in (0,\infty]$, i.e. $N_t = k$, then, a compound Poisson process is the sum of k *i.i.d.* jumps X_i from the jump size density $f(x)$:

$$X_t = \sum_{i=1}^{N_t=k} X_i = X_1 + X_2 + ... + X_k .$$

Another point is that a Poisson process is considered as a compound Poisson process with the unit jump size $X_i = 1$, since:

$$X_t = \sum_{i=1}^{N_t} X_i \text{ with } X_i = 1$$

$$X_t = \sum_{i=1}^{N_t} 1 = N_t = \sum_{k\geq 1} 1_{t\geq T_k} .$$

We present the important properties of a compound Poisson process in this section. Some of which can be easily seen by its simulated sample paths illustrated by Figure 1.8. In this figure, the jump size density $f(x)$ is a standard normal:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) .$$

Note that a path 1 has 7 jumps, a path 2 has 10 jumps, and a path 3 has 11 jumps.

Theorem 1.34 Fundamental properties of compound Poisson processes A compound Poisson process $(X_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ satisfies the following conditions:

(1) Its increments are independent. In other words, for any increasing sequence of time $0 \le t_1 < t_2 < ... < t_n$:

$$\mathbb{P}(X_0 \bigcap X_{t_1} - X_0 \bigcap X_{t_2} - X_{t_1} \bigcap ... \bigcap X_{t_n} - X_{t_{n-1}})$$

$$= \mathbb{P}(X_0)\mathbb{P}(X_{t_1} - X_0)\mathbb{P}(X_{t_2} - X_{t_1})...\mathbb{P}(X_{t_n} - X_{t_{n-1}}).$$

(2) Its increments are stationary (time homogeneous): i.e. for $h \ge 0$, $X_{t+h} - X_t$ has the same distribution as X_h. In other words, the distribution of increments does not depend on t.

(3) $\mathbb{P}(X_0 = 0) = 1$. The process starts from 0 almost surely (with probability 1).

(4) The process is stochastically continuous: $\forall \varepsilon > 0$, $\lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \ge \varepsilon) = 0$.

(5) Its sample paths are right continuous with left limit step functions in $\forall t \in [0,\infty]$ and the jump sizes X_i are $i.i.d.$ random variables from a density $f(x)$.

(6) For $\forall t \in (0,\infty]$, $\mathbb{P}(N_t = k < \infty) = 1$. In other words, the number of arrivals of an event is almost surely finite for any $t > 0$ including an infinite time horizon (i.e. $t = \infty$).

Proof

Similar to the proof of theorem 1.26.

As you may notice, it is very obvious that a compound Poisson process is a Lévy process because the condition (1) – (5) of theorem 1.34 is the definition of Lévy process (i.e. the definition 3.1).

Definition 1.64 Compound Poisson process as a Lévy process A Compound Poisson process $(X_{t\in[0,\infty]}) \equiv (\sum_{i=1}^{N_t} X_i)$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ is a Lévy process satisfying:

(1) A counter N_t follows a Poisson distribution with the intensity λt :

$$\mathbb{P}(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

(2) The jump sizes are $i.i.d.$ random variables from a density $f(x)$:

$$X_i \sim i.i.d. f(x).$$

(3) Its sample paths are right continuous with left limit step functions.

The condition (3) is implied by the conditions (2) and (3).

Theorem 1.35 Lévy process whose sample paths are right continuous with left limit step functions Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. If and only if the sample paths of $(X_{t\in[0,\infty)})$ are right continuous with left limit step functions in $\forall t \in [0,\infty]$ with probability 1, then, $(X_{t\in[0,\infty)})$ is a compound Poisson process.

Proof

Consult proposition 3.3 of Cont and Tankov (2004).

Definition 1.65 Compound Poisson process in terms of Lévy triplet (A,ℓ,γ)

Let $(X_{t\in[0,\infty)})$ be a real valued Lévy process with Lévy triplet (A,ℓ,γ) defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$. Then, $(X_{t\in[0,\infty)})$ is a compound Poisson process with the intensity $\lambda\in\mathbb{R}^+$, if it satisfies the following conditions:

(1) $A=0$. Its Gaussian variance is zero (i.e. Lévy-Itô decomposition).

(2) $\gamma=0$. Its drift is zero (i.e. Lévy-Itô decomposition).

(3) Its Lévy measure is given by:

$$\ell(x)=\lambda f(x),$$

where $f(x)$ is a jump size density satisfying:

$$\int_{-\infty}^{\infty} f(x)dx=1,$$

which holds because $f(x)$ is a probability density function. The integral of the Lévy measure of a compound Poisson process is the intensity parameter $\lambda\in\mathbb{R}^+$ because a Lévy measure $\ell(x)$ measures the arrival rate of jumps:

$$\int_{-\infty}^{\infty}\ell(x)dx=\int_{-\infty}^{\infty}\lambda f(x)dx=\lambda\int_{-\infty}^{\infty}f(x)dx=\lambda<\infty,$$

which is finite because the number of arrivals of an event is almost surely finite for any $t>0$ including an infinite time horizon $t=\infty$.

Theorem 1.36 Finite variation property of Compound Poisson process If $(X_{t\in[0,T]})$ is a compound Poisson process defined on a filtered probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{P})$, then, it is a real valued Lévy process of finite variation on the interval $[0,T]$. And, its Lévy triplet (A,ℓ,γ) satisfies:

$$A = 0 \text{ and } \int_{|x|<1} |x| \, \ell(dx) < \infty ,$$

which are finite variation conditions for Lévy processes (i.e. theorem 3.7). Finite variation property of a compound Poisson process can be easily guessed from its sample paths behavior. Stochastic processes of infinite variation have highly oscillatory sample paths such as a Brownian motion with drift. But, the sample paths of a compound Poisson process are rcll step functions which imply that a compound Poisson process is of finite variation:

$$\mathbb{P}(T(X) = \sup \sum_{i=1}^{n} |X(t_i) - X(t_{i-1})| < \infty) = 1 .$$

Table 4.3 summarizes the properties of a compound Poisson process.

Theorem 1.37 (with the proof) Characteristic function of Compound Poisson process Consider a compound Poisson process $(X_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ and let (A, ℓ, γ) be its Lévy triplet. Individual jumps X_i of a compound Poisson process are $i.i.d.$ random variables from a density $f(x)$, i.e. $X_i \sim i.i.d. f(x)$. Let ϕ_f be a characteristic function of a jump size density:

$$\phi_f(\omega) \equiv \mathcal{F}\left[f(x)\right] \equiv \int_{-\infty}^{\infty} e^{i\omega x} f(x) dx = E[e^{i\omega x}] .$$

Then, using the Lévy-Khinchin representation (theorem 3.3) with the Lévy triplet of a compound Poisson process $\left(A = 0, \ell = \lambda f(x), \gamma = 0\right)$, the characteristic exponent of a compound Poisson process is given by:

$$\psi_X(\omega) = \int_{-\infty}^{\infty} \left\{\exp(i\omega x) - 1\right\} \lambda f(x) dx = \lambda \int_{-\infty}^{\infty} \left\{f(x) \exp(i\omega x) - f(x)\right\} dx$$

$$\psi_X(\omega) = \lambda\{\int_{-\infty}^{\infty} f(x)\exp(i\omega x)dx - \int_{-\infty}^{\infty} f(x)dx\}$$

$$\psi_X(\omega) = \lambda\{\phi_f(\omega) - 1\}.$$

Thus, the characteristic function of a compound Poisson process $\phi_X(\omega)$ is expressed as:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right)$$

$$\phi_X(\omega) = \exp\left[t\lambda\{\phi_f(\omega) - 1\}\right] \quad (1\text{-}68).$$

For more information regarding the characteristic function of a compound Poisson process, consult Cont and Tankov (2004) pages 74-75 and Sato (1999) pages 18-21:

Theorem 1.38 Lévy measure of compound Poisson process A Lévy measure of a compound Poisson process $(X_{t\in[0,\infty]})$ with the intensity $\lambda \in \mathbb{R}^+$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ is given by:

$$\ell(x) = \lambda f(x),$$

where individual jumps X_i of a compound Poisson process are $i.i.d.$ random variables from a density $f(x)$, i.e. $X_i \sim i.i.d.f(x)$.

A compound Poisson process is a finite activity Lévy process because the integral of the Lévy measure of a compound Poisson process is the intensity parameter $\lambda \in \mathbb{R}^+$:

$$\int_{-\infty}^{\infty} \ell(x)dx = \int_{-\infty}^{\infty} \lambda f(x)dx = \lambda\int_{-\infty}^{\infty} f(x)dx = \lambda < \infty,$$

which is finite because the number of arrivals of an event is almost surely finite for any $t > 0$ including an infinite time horizon $t = \infty$.

Chapter 2: Introduction to Option Pricing with Fourier Transform

2.1 Introduction

This chapter is designed as an introduction to Fourier transform option pricing for readers who have zero previous knowledge of Fourier transform. First part of this sequel is devoted for the basic understanding of Fourier transform and discrete Fourier transform using numerous examples and providing important properties. Second part of this sequel applies FT and DFT option pricing approach for classic Black-Scholes model which is the only continuous exponential Lévy model. Some readers may question that what the need for FT option pricing is since BS model can price options with closed form formulae. The answer is that BS model is a special case of more general exponential Lévy models. Options cannot be priced with general exponential Lévy models using the traditional approach of the use of the risk-neutral density of the terminal stock price because it is not available. Therefore, Carr and Madan (1999) rewrite the option price in terms of a characteristic function of the log terminal stock price by the use of FT. The advantage of FT option pricing is its generality in the sense that the only thing necessary for FT option pricing is a characteristic function of the log terminal stock price. FT option pricing is a standard option pricing method from now on.

2.2 Prerequisite for Fourier Transform

In this section, we present prerequisite knowledge before moving to Fourier transform.

2.2.1 Radian

Radian is the unit of angle. A complete circle has $2\pi = 6.28319$ radians which is equal to $360°$. This in turn means that one radian is equal to:

$$1 \text{ radian } = \frac{360°}{2\pi} = 57.2958° \quad (2\text{-}1).$$

2.2.2 Wavelength

Wavelength λ of a waveform is defined as a distance (d) between peaks or troughs. In other words, wavelength is the distance at which a waveform completes one cycle:

$$\lambda \equiv \frac{\text{distance}}{1 \text{ cycle}} \quad (2\text{-}2).$$

Let v be the speed (distance/second), and f be the frequency (cycles/second, explained soon) of a waveform. These are related by:

$$\lambda(\text{distance} / \text{cycle}) \equiv \frac{v(\text{distance/second})}{f(\text{cycles/second})} \quad (2\text{-}3).$$

2.2.3 Frequency, Angular Frequency, and Period of a Waveform

Period T of oscillation of a wave is the seconds (time) taken for a waveform to complete one wavelength:

$$T \equiv \frac{\text{seconds}}{1 \text{ wavelength (cycle)}} \quad (2\text{-}4).$$

Period is by definition a reciprocal of a frequency. Let f be the frequency of a wave. Then:

$$T(\text{seconds/cycle}) \equiv \frac{1}{f(\text{cycles/second})} \qquad (2\text{-}5).$$

Frequency f of a wave measures the number of times for a wave to complete one wavelength (cycle) per second:

$$f \equiv \frac{\text{number of wavelengths (cycles)}}{1 \text{ second}} \qquad (2\text{-}6).$$

By definition, f is calculated as a reciprocal of the period of a wave:

$$f(\text{cycles/second}) \equiv \frac{1}{T(\text{seconds/cycle})} \qquad (2\text{-}7).$$

Frequency f is measured in Hertz (Hz). 1 Hz wave is a wave which completes one wavelength (cycle) per second. The frequency of the AC (alternating current) in U.S. is 60 Hz. Human beings can hear frequencies from about 20 to 20,000 Hz (called range of hearing). Alternatively, frequency can be calculated using the speed v and wavelength λ of a wave as:

$$f(\text{cycles/second}) \equiv \frac{v(\text{distance/second})}{\lambda(\text{distance/cycle})} \qquad (2\text{-}8).$$

Angular frequency (also called angular speed or radian frequency) ω is a measure of rotation rate (i.e. the speed at which an object rotates). The unit of measurement for ω is radians per 1 second. Since one cycle equals 2π radians, angular frequency ω is calculated as:

$$\omega(\text{radians/second}) = \frac{2\pi(\text{radians})}{T(\text{seconds/cycle})} = 2\pi f(\text{cycles/second}) \qquad (2\text{-}9).$$

Consider a sine wave $g(t) = \sin\left(2\pi(5)t\right)$ which is illustrated in Figure 2.2 for the time between 0 and 2 seconds. This sine wave has frequency $f = 5$ Hz (5 cycles

per second) and angular frequency $\omega = 10\pi$ Hz (10π radians per second). Its period is $T \equiv \frac{1}{f} = \frac{1}{5} = 0.2$ (seconds/cycle).

2.2.4 Sine and Cosine

Let θ be an angle which is measured counterclockwise from the x-axis along an arc of a unit circle. Sine function ($\sin\theta$) is defines as a vertical coordinate of the arc endpoint. Cosine function ($\cos\theta$) is defined as a horizontal coordinate of the arc endpoint. Sine and cosine functions $\sin\theta$ and $\cos\theta$ are periodic functions with period 2π as illustrated in Figure 2.4:

$$\sin\theta = \sin\left(\theta + 2\pi h\right) \quad (2\text{-}10),$$

$$\cos\theta = \cos\left(\theta + 2\pi h\right) \quad (2\text{-}11),$$

where h is any integer.

Following Pythagorean theorem, we have the identity:

$$\sin\theta^2 + \cos\theta^2 = 1 \quad (2\text{-}12).$$

2.2.5 Derivative and Integral of Sine and Cosine Function

Let $\sin(x)$ and $\cos(x)$ be sine and cosine functions for $x \in \mathbb{R}$. The derivative of $\sin(x)$ can be expressed as:

$$\frac{d\sin(x)}{dx} = \cos(x) \quad (2\text{-}13).$$

The derivative of $\cos(x)$ can be expressed as:

$$\frac{d\cos(x)}{dx} = -\sin(x) \quad (2\text{-}14).$$

The integral of $\sin(x)$ can be expressed as:

$$\int_{-\infty}^{\infty} \sin(x)dx = -\cos(x) \qquad (2\text{-}15).$$

Refer to any graduate school level trigonometry textbook for proofs.

2.2.6 Series Definition of Sine Function and Cosine Function

For any $x \in \mathbb{R}$:

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} \qquad (2\text{-}16),$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \qquad (2\text{-}17).$$

Refer to any graduate school level trigonometry textbook for proofs.

2.2.7 Euler's Formula

Euler's formula gives a very important relationship between the complex (imaginary) exponential function and the trigonometric functions. For any $x \in \mathbb{R}$:

$$e^{ix} = \cos(x) + i\sin(x) \qquad (2\text{-}18).$$

From (2-18), variants of Euler's formula are derived:

$$e^{-ix} = \cos(x) - i\sin(x) \qquad (2\text{-}19),$$

$$e^{ix} + e^{-ix} = 2\cos(x) \qquad (2\text{-}20),$$

$$e^{ix} - e^{-ix} = 2i\sin(x) \qquad (2\text{-}21).$$

Consider sine and cosine functions with complex arguments z. Then, Euler's formula tells:

$$\sin z = \mathrm{Im}(e^{iz}) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} = \frac{e^{iz} - e^{-iz}}{2i} \quad (2\text{-}22),$$

$$\cos z = \mathrm{Re}(e^{iz}) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} = \frac{e^{iz} + e^{-iz}}{2} \quad (2\text{-}23).$$

Refer to any graduate school level trigonometry textbook for proofs.

2.2.8 Sine Wave: Sinusoid

Sine wave is generally defined as a function of time t (seconds):

$$g(t) = a \sin\left(2\pi f_0 t + b\right) \quad (2\text{-}24),$$

where a is the amplitude, f_0 is the fundamental frequency (cycles/second, Hz), and b changes the phase (angular position) of a sinusoid. In terms of a fundamental angular frequency ω_0 (radians/second), a sine wave is defined as (i.e. $\omega_0 \equiv 2\pi f_0$):

$$g(t) = a \sin\left(\omega_0 t + b\right) \quad (2\text{-}25).$$

Figure 2.5 illustrates the role of a fundamental frequency f_0. When a fundamental frequency f_0 doubles from 1 (1 cycle /second) to 2 (cycles/second), its period becomes half from 1 to 1/2 seconds as illustrated.

The role of amplitude a is to increase or decrease the magnitude of an oscillation. Figure 2.6 illustrates how magnitude of an oscillation changes for three different amplitudes $a = 1/2$, 1, and 2. In audio study amplitude a determines how loud a sound is.

Consider three 1 Hz sine waves $\sin(2\pi t)$, $\sin(2\pi t + \frac{\pi}{2})$, and $\sin(2\pi t - \frac{\pi}{2})$. A sine wave $\sin(2\pi t)$ has a phase 0, $\sin(2\pi t + \frac{\pi}{2}) = \cos(2\pi t)$ has a phase $\pi/2$, and

$\sin(2\pi t - \frac{\pi}{2})$ has a phase $-\pi/2$. The role of a parameter b is to change the position of a waveform by an amount b as illustrated in Figure 2.7. Thus, $g(t) = a\sin(2\pi f_0 t)$ is a sinusoid at phase zero and $g(t) = a\cos(2\pi f_0 t)$ is a sinusoid at phase $\pi/2$. For the purpose of defining a sinusoid, it really does not matter whether $\sin(\ )$ or $\cos(\ )$ is used.

2.3 Fourier Transform (FT)

2.3.1 Definition of Fourier Transform

We consider Fourier transform of a function $g(t)$ from a time domain t into an angular frequency domain ω (radians/second). This follows the convention in physics. In the field of signal processing which is a major application of FT, frequency f (cycles /second) is used instead of ω. But this difference is not important because ω and f are measuring the same thing (rotation speed) in different units and related by $\omega = 2\pi f$. Table 2.1 gives a clear-cut relationship between frequency f and angular frequency ω.

We start from the most general definition of FT. FT from a function $g(t)$ to a function $\mathcal{G}(\omega)$ (thus, switching domains from t to ω) is defined using two arbitrary constants a and b called FT parameters as:

$$\mathcal{G}(\omega) \equiv \mathcal{F}\left[g(t)\right](\omega) \equiv \sqrt{\frac{|b|}{(2\pi)^{1-a}}}\int_{-\infty}^{\infty} e^{ib\omega t} g(t)dt \qquad (2\text{-}26).$$

Inverse Fourier transform from a function $\mathcal{G}(\omega)$ to a function $g(t)$ (thus, switching domains from ω to t) is defined as (i.e. the reverse procedure of (2-26)):

$$g(t) \equiv \mathcal{F}^{-1}\left[\mathcal{G}(\omega)\right](t) \equiv \sqrt{\frac{|b|}{(2\pi)^{1+a}}} \int_{-\infty}^{\infty} e^{-ib\omega t} \mathcal{G}(\omega) d\omega \qquad (2\text{-}27).$$

For our purpose which is to calculate characteristic functions, FT parameters are set as $(a,b) = (1,1)$. Thus, (2-26) and (2-27) become:

$$\mathcal{G}(\omega) \equiv \mathcal{F}\left[g(t)\right](\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t) dt \qquad (2\text{-}28),$$

$$g(t) \equiv \mathcal{F}^{-1}\left[\mathcal{G}(\omega)\right](t) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \mathcal{G}(\omega) d\omega \qquad (2\text{-}29).$$

Euler's formula (2.18) is for $t \in \mathbb{R}$:

$$e^{it} = \cos t + i \sin t .$$

Thus, the FT of (2-28) can be rewritten as:

$$\mathcal{G}(\omega) \equiv \mathcal{F}\left[g(t)\right](\omega) \equiv \int_{-\infty}^{\infty} \cos(\omega t) g(t) dt + i \int_{-\infty}^{\infty} \sin(\omega t) g(t) dt .$$

Intuitively speaking, FT is a decomposition of a waveform $g(t)$ (i.e. in time domain t) into a sum of sinusoids (i.e. sine and cosine functions) of different frequencies (Hz) which sum to the original waveform. In other words, FT enables any function in time domain to be represented by an infinite number of sinusoidal functions. Therefore, FT is an angular frequency representation (i.e. different look) of a function $g(t)$ and $\mathcal{G}(\omega)$ contains the exact same information as the original function $g(t)$.

Although FT parameters $(a,b)=(1,1)$ are used for calculating characteristic functions, different pairs of (a,b) are used for other purposes. For example, $(a,b)=(1,-1)$ in pure mathematics:

$$\mathcal{G}(\omega)\equiv\mathcal{F}_t\left[g(t)\right](\omega)\equiv\int_{-\infty}^{\infty}e^{-i\omega t}g(t)dt\,,$$

$$g(t)\equiv\mathcal{F}_\omega^{-1}\left[\mathcal{G}(\omega)\right](t)\equiv\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i\omega t}\mathcal{G}(\omega)d\omega\,.$$

Modern physics uses $(a,b)=(0,1)$:

$$\mathcal{G}(\omega)\equiv\mathcal{F}_t\left[g(t)\right](\omega)\equiv\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{i\omega t}g(t)dt\,,$$

$$g(t)\equiv\mathcal{F}_\omega^{-1}\left[\mathcal{G}(\omega)\right](t)\equiv\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-i\omega t}\mathcal{G}(\omega)d\omega\,.$$

In the field of signal processing and most of the standard textbooks of FT, FT parameters $(a,b)=(0,-2\pi)$ are used (i.e. they use frequency f instead of angular frequency ω):

$$\mathcal{G}(f)\equiv\mathcal{F}_t\left[g(t)\right](f)\equiv\int_{-\infty}^{\infty}e^{-2\pi ift}g(t)dt \qquad (2\text{-}30),$$

$$g(t)\equiv\mathcal{F}_f^{-1}[\mathcal{G}(f)](t)\equiv\int_{-\infty}^{\infty}e^{2\pi ift}\mathcal{G}(f)df \qquad (2\text{-}31).$$

FT definition of (2-30) and (2-31) are frequently used because this definition of FT is mathematically simpler to handle for the purpose of proofs.

In general, Fourier transform $\mathcal{G}(\omega)$ is a complex quantity:

$$\mathcal{G}(\omega)=\mathrm{Re}(\omega)+i\,\mathrm{Im}(\omega)=\left|\mathcal{G}(\omega)\right|e^{i\theta(\omega)} \qquad (2\text{-}32),$$

where $\mathrm{Re}(\omega)$ is the real part of the FT $\mathcal{G}(\omega)$, $\mathrm{Im}(\omega)$ is the imaginary part of the FT, $\left|\mathcal{G}(\omega)\right|$ is the amplitude of a time domain function $g(t)$, and $\theta(\omega)$ is the phase angle

of the FT $\mathcal{G}(\omega)$. $|\mathcal{G}(\omega)|$ and $\theta(\omega)$ can be expressed in terms of $\mathrm{Re}(\omega)$ and $\mathrm{Im}(\omega)$ as:

$$|\mathcal{G}(\omega)| = \sqrt{\mathrm{Re}^2(\omega) + \mathrm{Im}^2(\omega)} \qquad (2\text{-}33),$$

$$\theta(\omega) = \tan^{-1}\left[\frac{\mathrm{Im}(\omega)}{\mathrm{Re}(\omega)}\right] \qquad (2\text{-}34).$$

2.3.2 Examples of Fourier Transform

Before discussing important properties of FT, we present representative examples of FT in this section to get the feeling of what FT does.

Example 1) **Double-Sided Exponential Function**

Consider a double-sided exponential function with $A, \alpha \in \mathbb{R}$:

$$g(t) = Ae^{-\alpha|t|}.$$

From (2-28):

$$\mathcal{G}(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t)dt = \int_{-\infty}^{\infty} e^{i\omega t} Ae^{-\alpha|t|}dt = A\left(\int_{-\infty}^{0} e^{i\omega t} e^{\alpha t}dt + \int_{0}^{\infty} e^{i\omega t} e^{-\alpha t}dt\right)$$

$$\mathcal{G}(\omega) = A\left(\frac{1}{\alpha + i\omega} + \frac{1}{\alpha - i\omega}\right) = \frac{2A\alpha}{\alpha^2 + \omega^2}.$$

When $A = 1$ and $\alpha = 3$, the time domain function $g(t) = e^{-3|t|}$ and its Fourier transform $\mathcal{G}(\omega) = \dfrac{6}{9 + \omega^2}$ is plotted in Figure 2.8.

Using the signal processing definition of FT $(a, b) = (0, -2\pi)$ of the definition (2-30), FT of $g(t) = Ae^{-\alpha|t|}$ is computed as (which is simply obtained by substituting $\omega = 2\pi f$ into $\mathcal{G}(\omega)$):

$$\mathcal{G}(f) = \frac{2A\alpha}{\alpha^2 + 4\pi^2 f^2}.$$

When $A = 1$ and $\alpha = 3$, $\mathcal{G}(f) = \frac{6}{9 + 4\pi^2 f^2}$ which is plotted in Panel C of Figure 2.8.

Example 2) **Rectangular Pulse**

Consider a rectangular pulse with $A, T_0 \in \mathbb{R}$:

$$g(t) = \begin{cases} A & -T_0 \le t \le T_0 \\ 0 & |t| > T_0 \end{cases},$$

which is an even function of t (symmetric with respect to t). From (2-28) and use Euler's formula (2-21) $e^{ix} - e^{-ix} = 2i\sin(x)$:

$$\mathcal{G}(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t)dt = A\int_{-T_0}^{T_0} e^{i\omega t} dt = A\left\{\frac{-\left(e^{-i\omega T_0} - e^{i\omega T_0}\right)}{i\omega}\right\},$$

$$= A\left\{\frac{2i\sin(\omega T_0)}{i\omega}\right\} = \frac{2A\sin(\omega T_0)}{\omega}.$$

Using the signal processing definition of $\mathrm{FT}(a,b) = (0, -2\pi)$ of the definition (2-30), FT of a rectangular pulse is computed as:

$$\mathcal{G}(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi i f t} g(t)dt = A\int_{-T_0}^{T_0} e^{-2\pi i f t} dt = \frac{A\sin(2\pi f T_0)}{\pi f}.$$

When $A = 1$ and $T_0 = 2$, the time domain function $g(t)$, Fourier transform $\mathcal{G}(\omega) = \frac{2\sin(2\omega)}{\omega}$ in angular frequency, and Fourier transform $\mathcal{G}(f) = \frac{\sin(4\pi f)}{\pi f}$ in frequency domain are plotted in Figure 2.9.

Example 3) **Dirac's Delta Function (Impulse Function)**

Consider Dirac's delta function scaled by $a \in \mathbb{R}^+$:

$$g(t) = a\delta(t) .$$

From (2-28):

$$\mathcal{G}(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t)dt = \int_{-\infty}^{\infty} e^{i\omega t} a\delta(t)dt = ae^{i\omega 0} = a .$$

Using the signal processing definition of $\mathrm{FT}(a,b) = (0,-2\pi)$ of the definition (2-30), FT of a scaled Dirac's delta is computed as:

$$\mathcal{G}(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi ift} g(t)dt = \int_{-\infty}^{\infty} e^{-2\pi ift} a\delta(t)dt = ae^{-2\pi if\,0} = a .$$

When $a = 1$ (i.e. pure Dirac's delta), the time domain function $g(t) = \delta(t)$ and Fourier transforms $\mathcal{G}(\omega) = 1$ and $\mathcal{G}(f) = 1$ are plotted in Figure 2.10.

Example 4) **Gaussian Function**

Consider a Gaussian function with $A \in \mathbb{R}^+$:

$$g(t) = e^{-At^2} .$$

From (2-28):

$$\mathcal{G}(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t)dt = \int_{-\infty}^{\infty} e^{i\omega t} e^{-At^2} dt .$$

Use Euler's formula (2-18):

$$\mathcal{G}(\omega) = \int_{-\infty}^{\infty} e^{-At^2} \left\{\cos(\omega t) + i\sin(\omega t)\right\} dt$$

$$= \int_{-\infty}^{\infty} e^{-At^2} \cos(\omega t)dt + i\int_{-\infty}^{\infty} e^{-At^2} \sin(\omega t)dt$$

$$= \sqrt{\frac{\pi}{A}} e^{-\omega^2/4A} + i0 = \sqrt{\frac{\pi}{A}} e^{-\omega^2/4A} .$$

Using the signal processing definition of $\mathrm{FT}(a,b) = (0,-2\pi)$ of the definition (2-30), FT of a Gaussian function is computed as:

$$\mathcal{G}(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi ift} g(t)dt = \int_{-\infty}^{\infty} e^{-2\pi ift} e^{-At^2} dt = \sqrt{\frac{\pi}{A}} e^{-\pi^2 f^2 / A}.$$

When $A = 2$, the time domain function $g(t) = e^{-2t^2}$ and its Fourier transforms $\mathcal{G}(\omega) = \sqrt{\frac{\pi}{2}} e^{-\omega^2/8}$ and $\mathcal{G}(f) = \sqrt{\frac{\pi}{2}} e^{-\pi^2 f^2/2}$ are plotted in Figure 2.11. Note that FT of a Gaussian Function is another Gaussian function.

Example 5) **Cosine Wave** ***g(t) = A cos(2πf₀t) = A cos(ω₀t)***

Consider a sinusoid $g(t) = A\cos(2\pi f_0 t) = A\cos(\omega_0 t)$. Using the signal processing definition of $\mathrm{FT}(a,b) = (0,-2\pi)$ of the definition (2-30), FT of a cosine wave is computed as:

$$\mathcal{G}(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi ift} g(t)dt = \int_{-\infty}^{\infty} e^{-2\pi ift} A\cos(2\pi f_0 t)dt.$$

From Euler's formula:

$$\mathcal{G}(f) = A\int_{-\infty}^{\infty} e^{-2\pi ift} \frac{1}{2}\left(e^{i2\pi f_0 t} + e^{-i2\pi f_0 t}\right)dt$$

$$= \frac{1}{2} A\left\{\int_{-\infty}^{\infty} e^{-2\pi ift} e^{i2\pi f_0 t} dt + \int_{-\infty}^{\infty} e^{-2\pi ift} e^{-i2\pi f_0 t} dt\right\}$$

$$= \frac{1}{2} A\left\{\int_{-\infty}^{\infty} e^{-2\pi i(f-f_0)t} dt + \int_{-\infty}^{\infty} e^{-2\pi i(f+f_0)t} dt\right\}.$$

Use the identity of Dirac's delta function:

$$\delta(x-a) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega(x-a)} d\omega.$$

Therefore, we obtain:

$$\mathcal{G}(f) = \frac{1}{2} A\{\delta(f-f_0) + \delta(f+f_0)\}$$

$$\mathcal{G}(f) = \frac{A}{2}\delta(f - f_0) + \frac{A}{2}\delta(f + f_0),$$

which is two impulse functions at $f = \pm f_0$. Thus, FT of a cosine wave (which is an even function) is a real valued even function which means $\mathcal{G}(f)$ is symmetric about $f = 0$.

Next, in terms of angular frequency ω from (2-28):

$$\mathcal{G}(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t)dt = \int_{-\infty}^{\infty} e^{i\omega t} A\cos(\omega_0 t)dt.$$

From Euler's formula:

$$e^{ix} + e^{-ix} = 2\cos(x).$$

Therefore:

$$\mathcal{G}(\omega) = A\int_{-\infty}^{\infty} e^{i\omega t}\cos(\omega_0 t)dt = A\int_{-\infty}^{\infty} e^{i\omega t}\left\{\frac{1}{2}(e^{i\omega_0 t} + e^{-i\omega_0 t})\right\}dt$$

$$= \frac{A}{2}\int_{-\infty}^{\infty} e^{i\omega t}e^{i\omega_0 t}dt + \frac{A}{2}\int_{-\infty}^{\infty} e^{i\omega t}e^{-i\omega_0 t}dt = \frac{A}{2}\int_{-\infty}^{\infty} e^{i(\omega+\omega_0)t}dt + \frac{A}{2}\int_{-\infty}^{\infty} e^{i(\omega-\omega_0)t}dt.$$

Use the identity of Dirac's delta function:

$$\delta(x - a) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega(x-a)}d\omega.$$

Thus:

$$\mathcal{G}(\omega) = \frac{A}{2}2\pi\delta(\omega + \omega_0) + \frac{A}{2}2\pi\delta(\omega - \omega_0)$$

$$\mathcal{G}(\omega) = A\pi\delta(\omega + \omega_0) + A\pi\delta(\omega - \omega_0),$$

which is two impulse functions at $\omega = \pm\omega_0$. Figure 2.12 illustrates a cosine wave $g(t) = A\cos(2\pi f_0 t) = A\cos(\omega_0 t)$ and FTs $\mathcal{G}(\omega) = A\pi\delta(\omega+\omega_0) + A\pi\delta(\omega-\omega_0)$ and $\mathcal{G}(f) = \frac{A}{2}\delta(f-f_0) + \frac{A}{2}\delta(f+f_0)$.

Example 6) **Sine Wave *g(t) = A sin(2πf₀t) = A sin(ω₀t)***

Consider a sinusoid $g(t) = A\sin(2\pi f_0 t) = A\sin(\omega_0 t)$. Using the signal processing definition of $\mathrm{FT}(a,b) = (0,-2\pi)$ of the equation (2-30), FT of a sine wave is computed as:

$$\mathcal{G}(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi i f t} g(t)dt = \int_{-\infty}^{\infty} e^{-2\pi i f t} A\sin(2\pi f_0 t)dt .$$

From Euler's formula:

$$\mathcal{G}(f) = A\int_{-\infty}^{\infty} e^{-2\pi i f t}\frac{1}{2i}\left(e^{i2\pi f_0 t} - e^{-i2\pi f_0 t}\right)dt$$

$$= \frac{1}{2i}A\left\{\int_{-\infty}^{\infty} e^{-2\pi i f t}e^{i2\pi f_0 t}dt - \int_{-\infty}^{\infty} e^{-2\pi i f t}e^{-i2\pi f_0 t}dt\right\}$$

$$= \frac{1}{2i}A\left\{\int_{-\infty}^{\infty} e^{-2\pi i (f-f_0)t}dt - \int_{-\infty}^{\infty} e^{-2\pi i (f+f_0)t}dt\right\}.$$

Multiply $1 = i/i$ to the right hand side:

$$\mathcal{G}(f) = \frac{i}{2i^2}A\left\{\int_{-\infty}^{\infty} e^{-2\pi i (f-f_0)t}dt - \int_{-\infty}^{\infty} e^{-2\pi i (f+f_0)t}dt\right\}$$

$$= \frac{i}{2}A\left\{\int_{-\infty}^{\infty} e^{-2\pi i (f+f_0)t}dt - \int_{-\infty}^{\infty} e^{-2\pi i (f-f_0)t}dt\right\}.$$

Using the identity of Dirac's delta function, we obtain:

$$\mathcal{G}(f) = \frac{i}{2}A\{\delta(f+f_0) - \delta(f-f_0)\}$$

$$= i\frac{A}{2}\delta(f+f_0) - i\frac{A}{2}\delta(f-f_0),$$

which is two complex impulse functions at $f = \pm f_0$ which are not symmetric about $f = 0$.

Next, in terms of angular frequency ω from (2-28):

$$\mathcal{G}(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t)dt = A\int_{-\infty}^{\infty} e^{i\omega t}\sin(\omega_0 t)dt\,.$$

Using Euler's formula:

$$\mathcal{G}(\omega) = A\int_{-\infty}^{\infty} e^{i\omega t}\left\{\frac{1}{2i}(e^{i\omega_0 t} - e^{-i\omega_0 t})\right\}dt = \frac{A}{2i}\int_{-\infty}^{\infty} e^{i\omega t}e^{i\omega_0 t}dt - \frac{A}{2i}\int_{-\infty}^{\infty} e^{i\omega t}e^{-i\omega_0 t}dt$$

$$= \frac{A}{2i}\int_{-\infty}^{\infty} e^{i(\omega+\omega_0)t}dt - \frac{A}{2i}\int_{-\infty}^{\infty} e^{i(\omega-\omega_0)t}dt = \frac{Ai}{2i^2}\int_{-\infty}^{\infty} e^{i(\omega+\omega_0)t}dt - \frac{Ai}{2i^2}\int_{-\infty}^{\infty} e^{i(\omega-\omega_0)t}dt$$

$$= \frac{Ai}{2}\int_{-\infty}^{\infty} e^{i(\omega-\omega_0)t}dt - \frac{Ai}{2}\int_{-\infty}^{\infty} e^{i(\omega+\omega_0)t}dt$$

Use the identity of Dirac's delta function:

$$\mathcal{G}(\omega) = \frac{Ai}{2}2\pi\delta(\omega-\omega_0) - \frac{Ai}{2}2\pi\delta(\omega+\omega_0)$$

$$= Ai\pi\delta(\omega-\omega_0) - Ai\pi\delta(\omega+\omega_0).$$

Figure 2.13 plots a sine wave $g(t) = A\sin(2\pi f_0 t) = A\sin(\omega_0 t)$ and its Fourier transforms $\mathcal{G}(\omega) = Ai\pi\delta(\omega-\omega_0) - Ai\pi\delta(\omega+\omega_0)$ and $\mathcal{G}(f) = i\frac{A}{2}\delta(f+f_0) - i\frac{A}{2}\delta(f-f_0)$.

2.4 Properties of Fourier Transform

We will discuss important properties of Fourier transform in this section starting from Dirac's delta function which is essential to the understanding of properties of Fourier transform.

2.4.1 Dirac's Delta Function (Impulse Function)

Consider a function of the form with $n \in \mathbb{R}^+$:

$$h(x) = \frac{n}{\sqrt{\pi}} \exp(-n^2 x^2) \quad (2\text{-}35).$$

This function is plotted in Figure 2.14 for three different values for n. The function $h(x)$ becomes more and more concentrated around zero as the value of n increases. The function $h(x)$ has a unit integral:

$$\int_{-\infty}^{\infty} h(x)dx = \int_{-\infty}^{\infty} \frac{n}{\sqrt{\pi}} \exp\left(-n^2 x^2\right) dx = 1.$$

Dirac's delta function denoted by $\delta(x)$ can be considered as a limit of $h(x)$ when $n \to \infty$. In other words, $\delta(x)$ is a pulse of unbounded height and zero width with a unit integral:

$$\int_{-\infty}^{\infty} \delta(x)dx = 1.$$

Dirac's delta function $\delta(x)$ evaluates to 0 at all $x \in \mathbb{R}$ other than $x = 0$:

$$\delta(x) = \begin{cases} \delta(0) & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\delta(0)$ is undefined. $\delta(x)$ is called a generalized function not a function because of undefined $\delta(0)$. Therefore, $\delta(x)$ is a distribution with compact support

$\{0\}$ meaning that $\delta(x)$ does not occur alone but occurs combined with any continuous functions $f(x)$ and is well defined only when it is integrated.

Dirac's delta function can be defined more generally by its sampling property. Suppose that a function $f(x)$ is defined at $x=0$. Applying $\delta(x)$ to $f(x)$ yields $f(0)$:

$$\int_{-\infty}^{\infty} f(x)\delta(x)dx = f(0).$$

This is why Dirac's delta function $\delta(x)$ is called a functional because the use of $\delta(x)$ assigns a number $f(0)$ to a function $f(x)$. More generally for $a \in \mathbb{R}$:

$$\delta(x-a) = \begin{cases} \delta(0) & \text{if } x = a \\ 0 & \text{otherwise} \end{cases}$$

and:

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)dx = f(a),$$

or for $\varepsilon > 0$:

$$\int_{a-\varepsilon}^{a+\varepsilon} f(x)\delta(x-a)dx = f(a).$$

$\delta(x)$ has identities such as:

$$\delta(ax) = \frac{1}{|a|}\delta(x),$$

$$\delta(x^2 - a^2) = \frac{1}{2|a|}\left[\delta(x+a) + \delta(x-a)\right].$$

Dirac's delta function $\delta(x)$ can be defined as the limit $n \to \infty$ of a class of delta sequences:

$$\delta(x) = \lim_{n\to\infty} \delta_n(x),$$

such that:

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \delta_n(x) f(x) dx = f(0),$$

where $\delta_n(x)$ is a class of delta sequences. Examples of $\delta_n(x)$ other than (2-35) are:

$$\delta_n(x) = \begin{cases} \text{n} & \text{if} \quad -1/2n < x < 1/2n \\ 0 & \text{otherwise} \end{cases},$$

$$\delta_n(x) = \frac{1}{2\pi} \int_{-n}^{n} \exp(iux)\, du,$$

$$\delta_n(x) = \frac{1}{\pi x} \frac{e^{inx} - e^{-inx}}{2i},$$

$$\delta_n(x) = \frac{1}{2\pi} \frac{\sin\left[\left(n + 1/2\right)x\right]}{\sin\left(x/2\right)}.$$

2.4.2 Useful Identity: Dirac's Delta Function

Dirac's delta function $\delta(x)$ has the following very useful identity which we have used many times before:

$$\delta(x-a) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\{i\omega(x-a)\}\, d\omega \qquad (2\text{-}36).$$

PROOF[3]

First step is to prove a proposition for all $d = 2, 3, 4, \ldots$ and $j = 0, 1, 2, \ldots, d-1$ (note that j depends on d):

$$\frac{1}{d} \sum_{k=0}^{d-1} \exp\left(\frac{2\pi i}{d} jk\right) = 1_{j=0} = \begin{cases} 1 & \text{if} \quad j = 0 \\ 0 & \text{otherwise} \end{cases} \qquad (2\text{-}37).$$

[3] This is based on "Option Pricing Using Integral Transforms" by Carr, P., Geman, H., Madan, D., and Yor, M.

First, we deal an informal proof of a proposition (2-37).

When $d = 2$ and $j = 0$:

$$\frac{1}{d}\sum_{k=0}^{d-1}\exp\left(\frac{2\pi i}{d}jk\right) = \frac{1}{2}\sum_{k=0}^{2-1}\exp\left(\frac{2\pi i}{2}0k\right) = \frac{1}{2}\left\{\exp\left(\frac{2\pi i}{2}00\right) + \exp\left(\frac{2\pi i}{2}01\right)\right\}$$

$$= \frac{1}{2}\left\{\exp(0) + \exp(0)\right\} = 1.$$

When $d = 2$ and $j = 1$:

$$\frac{1}{d}\sum_{k=0}^{d-1}\exp\left(\frac{2\pi i}{d}jk\right) = \frac{1}{2}\sum_{k=0}^{2-1}\exp\left(\frac{2\pi i}{2}1k\right) = \frac{1}{2}\left\{\exp\left(\frac{2\pi i}{2}10\right) + \exp\left(\frac{2\pi i}{2}11\right)\right\}$$

$$= \frac{1}{2}\left\{\exp(0) + \exp(i\pi)\right\} = \frac{1}{2}\{1-1\} = 0.$$

When $d = 3$ and $j = 0$:

$$\frac{1}{d}\sum_{k=0}^{d-1}\exp\left(\frac{2\pi i}{d}jk\right) = \frac{1}{3}\sum_{k=0}^{3-1}\exp\left(\frac{2\pi i}{3}0k\right)$$

$$= \frac{1}{3}\left\{\exp\left(\frac{2\pi i}{3}00\right) + \exp\left(\frac{2\pi i}{3}01\right) + \exp\left(\frac{2\pi i}{3}02\right)\right\}$$

$$= \frac{1}{3}\left\{\exp(0) + \exp(0) + \exp(0)\right\} = 1.$$

When $d = 3$ and $j = 1$:

$$\frac{1}{d}\sum_{k=0}^{d-1}\exp\left(\frac{2\pi i}{d}jk\right) = \frac{1}{3}\sum_{k=0}^{3-1}\exp\left(\frac{2\pi i}{3}1k\right)$$

$$= \frac{1}{3}\left\{\exp\left(\frac{2\pi i}{3}10\right) + \exp\left(\frac{2\pi i}{3}11\right) + \exp\left(\frac{2\pi i}{3}12\right)\right\}$$

$$= \frac{1}{3}\left\{\exp(0) + \exp\left(\frac{2\pi i}{3}\right) + \exp\left(\frac{4\pi i}{3}\right)\right\}$$

$$= \frac{1}{3}\left\{1 + \left(0.866025i - 0.5\right) + \left(-0.866025i - 0.5\right)\right\}$$

$$= 0 .$$

When $d = 3$ and $j = 2$:

$$\frac{1}{d}\sum_{k=0}^{d-1} \exp\left(\frac{2\pi i}{d} jk\right) = \frac{1}{3}\sum_{k=0}^{3-1} \exp\left(\frac{2\pi i}{3} 2k\right)$$

$$= \frac{1}{3}\left\{\exp\left(\frac{2\pi i}{3} 20\right) + \exp\left(\frac{2\pi i}{3} 21\right) + \exp\left(\frac{2\pi i}{3} 22\right)\right\}$$

$$= \frac{1}{3}\left\{\exp\left(0\right) + \exp\left(\frac{4\pi i}{3}\right) + \exp\left(\frac{8\pi i}{3}\right)\right\}$$

$$= \frac{1}{3}\left\{1 + \left(-0.866025i - 0.5\right) + \left(0.866025i - 0.5\right)\right\}$$

$$= 0 .$$ □

(Formal) PROOF of a proposition (2-37)

Rewrite as the below:

$$\frac{1}{d}\sum_{k=0}^{d-1} \exp\left(\frac{2\pi i}{d} jk\right) = \frac{1}{d}\sum_{k=0}^{d-1} \beta^k \qquad (2\text{-}38),$$

where $\beta = \exp\left(\frac{2\pi i}{d} j\right)$. When $j = 0$, $\beta = \exp\left(\frac{2\pi i}{d} 0\right) = 1$. Thus, from (2-38):

$$\frac{1}{d}\sum_{k=0}^{d-1} \beta^k = \frac{1}{d}\sum_{k=0}^{d-1} 1^k = \frac{1}{d}\sum_{k=0}^{d-1} 1 = 1 .$$

When $j \neq 0$, consider the term $\sum_{k=0}^{d-1} \beta^k$ which is a geometric series:

$$S_{d-1} = \sum_{k=0}^{d-1} \beta^k = 1 + \beta + \beta^2 + \beta^3 + ... + \beta^{d-2} + \beta^{d-1} \qquad (2\text{-}39).$$

Multiply β to (2-39):

$$\beta S_{d-1} = \beta + \beta^2 + \beta^3 + ... + \beta^{d-2} + \beta^{d-1} + \beta^d \qquad (2\text{-}40).$$

Subtract (2-40) from (2-39):

$$(1-\beta)S_{d-1} = 1-\beta^d$$

$$S_{d-1} = \frac{1-\beta^d}{(1-\beta)} \qquad (2\text{-}41).$$

Note that for $j \neq 0$:

$$\beta^d = \exp\left(\frac{2\pi i}{d} j\right)^d = \exp\left(2\pi ij\right) = 1.$$

From (2-41):

$$S_{d-1} = \sum_{k=0}^{d-1} \beta^k = \frac{1-1}{(1-\beta)} = 0 .$$

From (2-37):

$$\frac{1}{d}\sum_{k=0}^{d-1} \exp\left(\frac{2\pi i}{d} jk\right) = \frac{1}{d} 0 = 0 .$$

□

Now we have completed the proof of a proposition (2-37) and we use this now.

Multiply d to both sides of a proposition (2-37):

$$d1_{j=0} = \sum_{k=0}^{d-1} \exp\left(\frac{2\pi i}{d} jk\right) \qquad (2\text{-}42).$$

As the limit $d \to \infty$ and plug $j = x - a$, (2-36) becomes:

$$\delta(x-a) = \int_{-\infty}^{\infty} e^{i2\pi f(x-a)} df \qquad (2\text{-}43).$$

Convert frequency f into angular frequency ω which is $\omega = 2\pi f$. From (2-43):

$$\delta(x-a)=\int_{-\infty}^{\infty}e^{i\omega(x-a)}d\frac{\omega}{2\pi}.$$

This completes the proof of an identity (2-36).

□

2.4.3 Linearity of Fourier Transform

Consider time domain functions $f(t)$ and $g(t)$ which have Fourier transforms $\mathcal{F}(\omega)$ and $\mathcal{G}(\omega)$ defined by the equation (2-28). Then:

$$\int_{-\infty}^{\infty}\{af(t)+bg(t)\}e^{i\omega t}dt=a\int_{-\infty}^{\infty}f(t)e^{i\omega t}dt+b\int_{-\infty}^{\infty}g(t)e^{i\omega t}dt$$

$$=a\mathcal{F}(\omega)+b\mathcal{G}(\omega). \quad (2\text{-}44)$$

Or, we can write:

$$\mathcal{F}\left[af(t)+bg(t)\right](\omega)=a\mathcal{F}\left[f(t)\right](\omega)+b\mathcal{F}\left[g(t)\right](\omega).$$

Linearity means two things: homogeneity and additivity. Homogeneity of Fourier transform indicates that if the amplitude is changed in one domain by a, the amplitude in the other domain changes by an exactly the same amount a (i.e. scaling property):

$$\mathcal{F}\left[af(t)\right](\omega)=a\mathcal{F}\left[f(t)\right](\omega).$$

Additivity of Fourier transform indicates that an addition in one domain corresponds to an addition in other domain:

$$\mathcal{F}\left[f(t)+g(t)\right](\omega)=\mathcal{F}\left[f(t)\right](\omega)+\mathcal{F}\left[g(t)\right](\omega).$$

2.4.4 FT of Even and Odd Functions

A function $g(x)$ is said to be even if for $x \in \mathbb{R}$:

$$g(x) = g(-x) \quad (2\text{-}45),$$

which implies that even functions are symmetric with respect to vertical axis. Examples of even functions are illustrated in Panel A of Figure 2.15.

A function $g(x)$ is said to be odd if for $x \in \mathbb{R}$:

$$-g(x) = g(-x) \quad (2\text{-}46),$$

which implies that odd functions are symmetric with respect to the origin. Examples of odd functions are illustrated in Panel B of Figure 2.15.

There are several important properties of even and odd functions. The sum of even functions is even and the sum of odd functions is odd. The product of two even functions is even and the product of two odd functions is also even. The product of an even and an odd function is odd.

Let $even(x)$ be an even function and $odd(x)$ be an odd function. Integral properties of even and odd functions are:

$$\int_{-A}^{A} odd(x)dx = 0 \quad (2\text{-}47),$$

$$\int_{-A}^{A} even(x)dx = 2\int_{0}^{A} even(x)dx \quad (2\text{-}48).$$

Consider FT of an even function $g(t)$. From the definition (2-28) and use Euler's formula:

$$\mathcal{G}(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t)dt = \int_{-\infty}^{\infty} \cos(\omega t)g(t)dt + i\int_{-\infty}^{\infty} \sin(\omega t)g(t)dt \quad (2\text{-}49).$$

Since the imaginary part $\int_{-\infty}^{\infty} \sin(\omega t) g(t) dt$ is zero (because $\sin(\omega t) g(t)$ is odd and use the integral property), FT $\mathcal{G}(\omega)$ is real and symmetric with respect to $\omega = 0$. In other words, FT of an even function is also even.

Next, consider FT of an odd function $g(t)$. Since the term $\int_{-\infty}^{\infty} \cos(\omega t) g(t) dt$ in (2-49) becomes zero ($\cos(\omega t) g(t)$ is odd and the integral property), FT is given as:

$$\mathcal{G}(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} g(t) dt = i \int_{-\infty}^{\infty} \sin(\omega t) g(t) dt .$$

This means that FT of an odd function $\mathcal{G}(\omega)$ is complex and asymmetric with respect to $\omega = 0$.

2.4.5 Symmetry of Fourier Transform

By the definition of an inverse Fourier transform (2-29):

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \mathcal{F}(\omega) d\omega .$$

Change the variable in the integration to y:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iyt} \mathcal{F}(y) dy .$$

Consider $2\pi f(-t)$:

$$2\pi f(-t) = \int_{-\infty}^{\infty} e^{iyt} \mathcal{F}(y) dy \qquad (2\text{-}50).$$

We can say that the right hand side of (2-50) is by definition the Fourier transform of a function $\mathcal{F}(y)$. Replace t by ω and y by t and (2-50) becomes:

$$2\pi f(-\omega) = \int_{-\infty}^{\infty} e^{it\omega} \mathcal{F}(t) dt \qquad (2\text{-}51).$$

The equation (2-51) is called a symmetry property of FT. It means that if a function $f(t)$ has a FT $\mathcal{F}(\omega)$, $\mathcal{F}(t)$ has a FT $2\pi f(-\omega)$. In other words, if $\left(f(t), \mathcal{F}(\omega)\right)$ is a FT pair, $\left(\mathcal{F}(t), 2\pi f(-\omega)\right)$ is another FT pair.

This symmetry property of FT can be shown with mathematically simpler form in the frequency domain f Hz (cycles/second). By the definition of an inverse FT (2-31):

$$g(t) \equiv \int_{-\infty}^{\infty} e^{2\pi i f t} \mathcal{G}(f) df .$$

Change the variable in the integration to y:

$$g(t) \equiv \int_{-\infty}^{\infty} e^{2\pi i y t} \mathcal{G}(y) dy .$$

Consider $g(-t)$:

$$g(-t) \equiv \int_{-\infty}^{\infty} e^{-2\pi i y t} \mathcal{G}(y) dy \qquad (2\text{-}52).$$

We can say that the right hand side of (2-52) is by definition the FT of a function $\mathcal{G}(y)$. Replace t by f and y by t and (2-52) becomes:

$$g(-f) \equiv \int_{-\infty}^{\infty} e^{-2\pi i t f} \mathcal{G}(t) dt \qquad (2\text{-}53).$$

We can state in this case that if $\left(g(t), \mathcal{G}(f)\right)$ is a FT pair, $\left(\mathcal{G}(t), g(-f)\right)$ is another FT pair.

2.4.6 Differentiation of Fourier Transform

By the definition of an inverse Fourier transform (2-29):

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\mathcal{F}(\omega)d\omega .$$

Differentiate with respect to t:

$$\frac{\partial f(t)}{\partial t} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\mathcal{F}(\omega)\frac{\partial e^{-i\omega t}}{\partial t} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\left(-i\omega\mathcal{F}(\omega)\right)e^{-i\omega t}$$

$$= -i\omega\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}(\omega)e^{-i\omega t}d\omega \qquad (2\text{-}54).$$

By the definition of an inverse FT (2-29), the equation (2-54) becomes:

$$\frac{\partial f(t)}{\partial t} = -i\omega f(t) \qquad (2\text{-}55).$$

Equation (2-55) tells us that FT of $\partial f(t)/\partial t$ is equal to a FT of $f(t)$ multiplied by $-i\omega$:

$$\mathcal{F}\left[\partial f(t)/\partial t\right](\omega) = -i\omega\mathcal{F}\left[f(t)\right](\omega) \qquad (2\text{-}56).$$

Next, consider FT in frequency domain f. By the definition of an inverse FT of the equation (2-31):

$$g(t) \equiv \int_{-\infty}^{\infty} e^{2\pi ift}\mathcal{G}(f)df .$$

Differentiate with respect to t:

$$\frac{\partial g(t)}{\partial t} = 2\pi if\int_{-\infty}^{\infty} e^{2\pi ift}\mathcal{G}(f)df = 2\pi ifg(t) \qquad (2\text{-}57).$$

Equation (2-57) tells us that FT of $\partial g(t)/\partial t$ is equal to a Fourier transform of $g(t)$ multiplied by $2\pi if$:

$$\mathcal{F}\left[\partial g(t)/\partial t\right](f) = 2\pi i f\,\mathcal{F}\left[g(t)\right](f).$$

2.4.7 Time Scaling of Fourier Transform

Consider a time domain function $f(t)$ and its Fourier transform $\mathcal{F}(\omega)$ by the equation (2-28):

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} f(t)dt.$$

Then, FT of a function $f(at)$ (i.e. scaled by a real non-zero constant a) can be expressed in terms of $\mathcal{F}(\omega)$ as:

$$\frac{1}{|a|}\mathcal{F}(\frac{\omega}{a}) = \int_{-\infty}^{\infty} e^{i\omega t} f(at)dt \qquad (2\text{-}58).$$

PROOF

Set $at = s$. When $a > 0$:

$$\int_{-\infty}^{\infty} e^{i\omega t} f(at)dt = \int_{-\infty}^{\infty} e^{i\omega s/a} f(s)d\left(\frac{s}{a}\right) = \frac{1}{a}\int_{-\infty}^{\infty} e^{i(\omega/a)s} f(s)ds$$

$$= \frac{1}{a}\mathcal{F}(\frac{\omega}{a}).$$

When $a < 0$:

$$\int_{-\infty}^{\infty} e^{i\omega t} f(at)dt = \int_{-\infty}^{\infty} e^{i\omega s/a} f(s)d\left(\frac{s}{a}\right) = -\frac{1}{a}\int_{-\infty}^{\infty} e^{i(\omega/a)s} f(s)ds$$

$$= -\frac{1}{a}\mathcal{F}(\frac{\omega}{a}).$$

□

Similarly, FT of a function $g(at)$ can be expressed in terms of $\mathcal{G}(f)$ in frequency domain f as:

$$\frac{1}{|a|}\mathcal{G}(\frac{f}{a}) = \int_{-\infty}^{\infty} e^{-2\pi ift} g(at)dt .$$

2.4.8 Time Shifting of Fourier Transform

Consider a function $f(t)$ and its Fourier transform $\mathcal{F}(\omega)$:

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} f(t)dt .$$

Then, FT of a function $f(t-t_0)$ (i.e. time t is shifted by $t_0 \in \mathbb{R}$) can be expressed in terms of $\mathcal{F}(\omega)$ as:

$$\int_{-\infty}^{\infty} e^{i\omega t} f(t-t_0)dt = e^{i\omega t_0}\mathcal{F}(\omega) \qquad (2\text{-}59).$$

PROOF

Set $t - t_0 = t^*$:

$$\int_{-\infty}^{\infty} e^{i\omega t} f(t-t_0)dt = \int_{-\infty}^{\infty} e^{i\omega(t^*+t_0)} f(t^*)d(t^*+t_0)$$

$$= e^{i\omega t_0}\int_{-\infty}^{\infty} e^{i\omega t^*} f(t^*)dt^*$$

$$= e^{i\omega t_0}\mathcal{F}(\omega) .$$

□

Next, consider FT in frequency domain f. FT of a time domain function $g(t)$ is defined by the definition (2-30) as:

$$\mathcal{G}(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi ift} g(t)dt .$$

Then, FT of a function $g(t-t_0)$ (i.e. time t is shifted by $t_0 \in \mathbb{R}$) can be expressed in terms of $\mathcal{G}(f)$ as:

$$\int_{-\infty}^{\infty} e^{-2\pi ift}\, g(t-t_0)dt = e^{-2\pi ift_0}\mathcal{G}(f) \qquad (2\text{-}60).$$

PROOF

Set $t - t_0 = t^*$:

$$\int_{-\infty}^{\infty} e^{-2\pi ift}\, g(t-t_0)dt = \int_{-\infty}^{\infty} e^{-2\pi if(t^*+t_0)}\, g(t^*)d(t^*+t_0)$$

$$= e^{-2\pi ift_0}\int_{-\infty}^{\infty} e^{-2\pi ift^*}\, g(t^*)dt^* = e^{-2\pi ift_0}\mathcal{G}(f).$$

□

2.4.9 Convolution: Time Convolution Theorem

Convolution of time domain functions $f(t)$ and $g(t)$ over a finite interval $[0,t]$ is defined as:

$$f * g \equiv \int_0^t f(\tau)g(t-\tau)d\tau \qquad (2\text{-}61).$$

Convolution of time domain functions $f(t)$ and $g(t)$ over an infinite interval $[-\infty,\infty]$ is defined as:

$$f * g \equiv \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau = \int_{-\infty}^{\infty} g(\tau)f(t-\tau)d\tau \qquad (2\text{-}62).$$

Convolution can be considered as an integral which measures the amount of overlapping of one function $g(t)$ when $g(t)$ is shifted over another function $f(t)$. Website by mathworld provides an excellent description of convolution with animation. For example, suppose $f(t)$ and $g(t)$ are Gaussian functions:

$$f(t) = \frac{1}{\sqrt{2\pi\sigma_1^{\,2}}}\exp\{-\frac{(t-\mu_1)^2}{2\sigma_1^{\,2}}\},$$

$$g(t) = \frac{1}{\sqrt{2\pi{\sigma_2}^2}} \exp\{-\frac{(t-\mu_2)^2}{2{\sigma_2}^2}\}.$$

Then, the convolution of two Gaussian functions is calculated as from (2-62):

$$f * g = \frac{1}{\sqrt{2\pi({\sigma_1}^2 + {\sigma_2}^2)}} \exp\{-\frac{\left(t-(\mu_1+\mu_2)\right)^2}{2({\sigma_1}^2 + {\sigma_2}^2)}\},$$

which is another Gaussian function. The convolution $f * g$ of two Gaussians for the case $\mu_1 = 0,\ \mu_2 = 0,\ \sigma_1 = 1$, and σ_2 =2 is plotted in Figure 2.16.

Consider time domain functions $f(t)$ and $g(t)$ with Fourier transforms $\mathcal{F}(\omega)$ and $\mathcal{G}(\omega)$. FT of the convolution of $f(t)$ and $g(t)$ in the time domain is equal to the multiplication in the angular frequency domain (called time-convolution theorem):

$$\mathcal{F}\left(f * g\right) \equiv \mathcal{F}\left(\int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau\right) = \mathcal{F}(\omega)\mathcal{G}(\omega) \qquad (2\text{-}63).$$

PROOF

$$\mathcal{F}\left(f * g\right) \equiv \mathcal{F}\left(\int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau\right) = \int_{-\infty}^{\infty} e^{i\omega t}\int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau dt$$

$$= \int_{-\infty}^{\infty} f(\tau)d\tau\left(\int_{-\infty}^{\infty} g(t-\tau)e^{i\omega t}dt\right).$$

Use the time shifting property of Fourier transform:

$$\mathcal{F}\left(f * g\right) = \int_{-\infty}^{\infty} f(\tau)d\tau\left(e^{i\omega\tau}\mathcal{G}(\omega)\right) = \mathcal{G}(\omega)\int_{-\infty}^{\infty} f(\tau)e^{i\omega\tau}d\tau$$

$$= \mathcal{F}(\omega)\mathcal{G}(\omega).$$

□

2.4.10 Frequency-Convolution Theorem

Consider time domain functions $f(t)$ and $g(t)$ with Fourier transforms $\mathcal{F}(\omega)$ and $\mathcal{G}(\omega)$. Frequency-convolution theorem states that convolution in the angular frequency domain (scaled by 1/2π) is equal to the multiplication in the time domain. In other words, FT of the product $f(t)g(t)$ in the time domain is equal to the convolution $\mathcal{F}(\omega)*\mathcal{G}(\omega)$ (scaled by 1/2π) in the angular frequency domain:

$$\mathcal{F}\left[f(t)g(t)\right](\omega)=\frac{1}{2\pi}\mathcal{F}(\omega)*\mathcal{G}(\omega)\equiv\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}(\varpi)\mathcal{G}(\omega-\varpi)d\varpi \qquad (2\text{-}64).$$

PROOF

We prove this by showing that the inverse FT of the convolution $\mathcal{F}*\mathcal{G}$ in the angular frequency domain is equal to the multiplication $f(t)g(t)$ (scaled by 2π) in the time domain, although there are several different ways to prove this.

Following the definition of inverse FT (2-29):

$$\begin{aligned}\mathcal{F}_{\omega}^{-1}\left[\mathcal{F}(\omega)*\mathcal{G}(\omega)\right](t)&\equiv\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}\mathcal{F}(\omega)*\mathcal{G}(\omega)d\omega\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}\int_{-\infty}^{\infty}\mathcal{F}(\varpi)\mathcal{G}(\omega-\varpi)d\varpi d\omega\\&=\int_{-\infty}^{\infty}d\varpi\mathcal{F}(\varpi)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}\mathcal{G}(\omega-\varpi)d\omega\right).\end{aligned}$$

Using the frequency shifting (modulation) property of FT:

$$\begin{aligned}\mathcal{F}_{\omega}^{-1}\left[\mathcal{F}(\omega)*\mathcal{G}(\omega)\right](t)&=\int_{-\infty}^{\infty}d\varpi\mathcal{F}(\varpi)e^{-i\varpi t}g(t)\\&=g(t)\int_{-\infty}^{\infty}d\varpi\mathcal{F}(\varpi)e^{-i\varpi t}=g(t)2\pi f(t).\end{aligned}$$

□

Next, consider FT in the frequency domain f. Following the definition of inverse FT (2-31):

$$\mathcal{F}_f^{-1}\left[\mathcal{F}(f)*\mathcal{G}(f)\right](t) \equiv \int_{-\infty}^{\infty} e^{2\pi i f t}\mathcal{F}(f)*\mathcal{G}(f)df$$

$$= \int_{-\infty}^{\infty} e^{2\pi i f t}\int_{-\infty}^{\infty}\mathcal{F}(f^*)\mathcal{G}(f-f^*)df^*df$$

$$= \int_{-\infty}^{\infty} df * \mathcal{F}(f^*)\left(\int_{-\infty}^{\infty} e^{2\pi i f t}\mathcal{G}(f-f^*)df\right).$$

Using the frequency shifting (modulation) property of FT:

$$\mathcal{F}_f^{-1}\left[\mathcal{F}(f)*\mathcal{G}(f)\right](t) = \int_{-\infty}^{\infty} df * \mathcal{F}(f^*)e^{2\pi i f^* t}g(t)$$

$$= g(t)\int_{-\infty}^{\infty} df * \mathcal{F}(f^*)e^{2\pi i f^* t} = f(t)g(t).$$

□

2.4.11 Frequency Shifting: Modulation

Consider time domain function $f(t)$ with Fourier transform $\mathcal{F}(\omega)$. If FT $\mathcal{F}(\omega)$ is shifted by $\omega_0 \in \mathbb{R}$ in the angular frequency domain, then the inverse FT $f(t)$ is multiplied by $e^{i\omega_0 t}$:

$$e^{-i\omega_0 t}f(t) = \mathcal{F}^{-1}(\omega-\omega_0) \quad \text{(2-65)},$$

$$\mathcal{F}\left(e^{-i\omega_0 t}f(t)\right) = \mathcal{F}(\omega-\omega_0) \quad \text{(2-66)}.$$

PROOF

Let $s = \omega - \omega_0$. From the definition of an inverse Fourier transform (2-29):

$$\mathcal{F}^{-1}(\omega-\omega_0) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\mathcal{F}(\omega-\omega_0)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i(s+\omega_0)t}\mathcal{F}(s)d(s+\omega_0)$$

$$= e^{-i\omega_0 t}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ist}\mathcal{F}(s)ds = e^{-i\omega_0 t} f(t).$$

□

Similarly, if FT $\mathcal{F}(f)$ is shifted by $f_0 \in \mathbb{R}$ in the frequency domain, then the inverse FT $g(t)$ is multiplied by $e^{2\pi i f_0 t}$:

$$e^{2\pi i f_0 t} g(t) = \mathcal{F}^{-1}\left[\mathcal{G}(f - f_0)\right] \quad \text{(2-67)},$$

$$\mathcal{F}\left(e^{2\pi i f_0 t} g(t)\right) = \mathcal{G}(f - f_0) \quad \text{(2-68)}.$$

2.4.12 Parseval's Relation

Let $f(t)$ and $g(t)$ be L^2-complex functions. A L^2-function can be informally considered as a square integrable function (i.e. A function $f(t)$ is said to be square-integrable if $\int_{-\infty}^{\infty}\left|f(t)\right|^2 dt < \infty$.) Let $\mathcal{F}(\omega)$ and $\mathcal{G}(\omega)$ be the Fourier transforms of $f(t)$ and $g(t)$ defined by (2-28):

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} f(t)dt,$$

$$\mathcal{G}(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} g(t)dt.$$

Let $\overline{g}(t)$ be a complex conjugate of $f(t)$ and $\overline{\mathcal{G}}(\omega)$ be a complex conjugate of $\mathcal{F}(\omega)$[4]:

$$\left|f(t)\right|^2 = f(t)\overline{g}(t),$$

$$\left|\mathcal{F}(\omega)\right|^2 = \mathcal{F}(\omega)\overline{\mathcal{G}}(\omega).$$

[4] A complex conjugate of a complex number $z \equiv a + bi$ is $\overline{z} \equiv a - bi$.

Then, Parseval's relation is:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathcal{F}(\omega)|^2 \, d\omega ,$$

$$\int_{-\infty}^{\infty} f(t)\overline{g}(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F}(\omega)\overline{\mathcal{G}}(\omega)d\omega \qquad (2\text{-}69).$$

Parseval's relation in the case of FT parameters $(a,b) = (1,1)$ indicates that there is a very simple relationship between the power of a signal function $f(t)$ computed in signal space t or transform space ω of the form (2-69).

Parseval's relation becomes simpler when considered in the frequency domain f instead of angular frequency domain ω. Let $\mathcal{F}(f)$ and $\mathcal{G}(f)$ be the Fourier transforms of $f(t)$ and $g(t)$ defined by (2-30):

$$\mathcal{F}(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi i f t} f(t)dt ,$$

$$\mathcal{G}(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi i f t} g(t)dt .$$

Let $\overline{g}(t)$ be a complex conjugate of $f(t)$ and $\overline{\mathcal{G}}(f)$ be a complex conjugate of $\mathcal{F}(f)$:

$$|f(t)|^2 = f(t)\overline{g}(t) ,$$

$$|\mathcal{F}(f)|^2 = \mathcal{F}(f)\overline{\mathcal{G}}(f) .$$

Then, Parseval's relation is:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = \int_{-\infty}^{\infty} |\mathcal{F}(f)|^2 \, df ,$$

$$\int_{-\infty}^{\infty} f(t)\overline{g}(t)dt = \int_{-\infty}^{\infty} \mathcal{F}(f)\overline{\mathcal{G}}(f)df \qquad (2\text{-}70).$$

This version of Parseval's relation means that the power of a signal function $f(t)$ is same whether it is computed in signal space t or in transform space f.

PROOF

Using inverse Fourier transforms of (2-29):

$$\int_{-\infty}^{\infty}|f(t)|^2\,dt=\int_{-\infty}^{\infty}f(t)\bar{g}(t)dt=\int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}\mathcal{F}(\omega)d\omega\right]\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i\omega' t}\bar{\mathcal{G}}(\omega')d\omega'\right]dt$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}(\omega)\int_{-\infty}^{\infty}\bar{\mathcal{G}}(\omega')\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{it(\omega'-\omega)}dt\right]d\omega' d\omega.$$

Use the identity of Dirac's delta function:

$$\delta(t-a)\equiv\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i\omega(t-a)}d\omega.$$

Thus:

$$\int_{-\infty}^{\infty}|f(t)|^2\,dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}(\omega)\int_{-\infty}^{\infty}\bar{\mathcal{G}}(\omega')\delta(\omega'-\omega)d\omega' d\omega$$

$$\int_{-\infty}^{\infty}|f(t)|^2\,dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}(\omega)\bar{\mathcal{G}}(\omega)d\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}|\mathcal{F}(\omega)|^2\,d\omega.$$

□

2.4.13 Summary of Important Properties of Fourier Transform

Table 2.2 and 2.3 summarizes all the important properties of FT.

2.5 Existence of the Fourier Integral

Thus far, we've assumed that Fourier integral and its inverse of the definitions (2-28) and (2-29) of a FT pair $\left(g(t),\mathcal{G}(\omega)\right)$ do exist (i.e. they are well-defined for all

functions). Sufficient (but not necessary) condition for the existence of Fourier transform and its inverse is:

$$\int_{-\infty}^{\infty} \left| g(t) \right| dt < \infty \qquad (2\text{-}71),$$

which is an integrability condition.

2.6 Characteristic Function

2.6.1 Definition of a Characteristic Function

Let X be a random variable with its probability density function $\mathbb{P}(x)$. A characteristic function $\phi(\omega)$ with $\omega \in \mathbb{R}$ is defined as the Fourier transform of the probability density function $\mathbb{P}(x)$ using Fourier transform parameters $(a,b) = (1,1)$. From the definition (2-28):

$$\phi(\omega) \equiv \mathcal{F}\left[\mathbb{P}(x)\right] \equiv \int_{-\infty}^{\infty} e^{i\omega x} \mathbb{P}(x) dx = E\left[e^{i\omega x}\right] \qquad (2\text{-}72).$$

Using the Euler's formula, $\phi(\omega)$ can be expressed as:

$$\phi(\omega) = E\left[e^{i\omega x}\right] = E\left[\cos(\omega x)\right] + iE\left[\sin(\omega x)\right].$$

Taylor series expansion of a real function $f(x)$ in one dimension about a point $x = b$ is given by:

$$f(x) = f(b) + f'(b)(x-b) + \frac{f''(b)}{2!}(x-b)^2 + \frac{f'''(b)}{3!}(x-b)^3 + \ldots \qquad (2\text{-}73).$$

Taylor series expansion of an exponential function $\exp(i\omega x)$ in one dimension about a point $x = 0$ is given by from (2-73) as:

$$\exp(i\omega x) = \exp(i\omega 0) + \frac{\partial \exp(i\omega x)}{\partial x}\Big|_{x=0} (x-0) + \frac{1}{2!}\frac{\partial^2 \exp(i\omega x)}{\partial x^2}\Big|_{x=0} (x-0)^2$$

$$+\frac{1}{3!}\frac{\partial^3 \exp(i\omega x)}{\partial x^3}\Big|_{x=0}(x-0)^3+\ldots$$

$$\exp(i\omega x)=\exp(i\omega 0)+i\omega\exp(i\omega x)\Big|_{x=0}(x-0)+\frac{1}{2!}(i\omega)^2\exp(i\omega x)\Big|_{x=0}(x-0)^2$$

$$+\frac{1}{3!}(i\omega)^3\exp(i\omega x)\Big|_{x=0}(x-0)^3+\ldots$$

$$\exp(i\omega x)=1+i\omega x+\frac{1}{2!}(i\omega x)^2+\frac{1}{3!}(i\omega x)^3+\frac{1}{4!}(i\omega x)^4+\ldots \quad (2\text{-}74).$$

Therefore, a characteristic function $\phi(\omega)$ can be rewritten as from equations (2-72) and (2-74) as:

$$\phi(\omega)\equiv\int_{-\infty}^{\infty}e^{i\omega x}\mathbb{P}(x)dx$$

$$=\int_{-\infty}^{\infty}\left(1+i\omega x+\frac{1}{2!}(i\omega x)^2+\frac{1}{3!}(i\omega x)^3+\frac{1}{4!}(i\omega x)^4+\ldots\right)\mathbb{P}(x)dx$$

$$=\int_{-\infty}^{\infty}\mathbb{P}(x)dx+i\omega\int_{-\infty}^{\infty}x\mathbb{P}(x)dx+\frac{1}{2!}(i\omega)^2\int_{-\infty}^{\infty}x^2\mathbb{P}(x)dx+\frac{1}{3!}(i\omega)^3\int_{-\infty}^{\infty}x^3\mathbb{P}(x)dx+\ldots$$

$$=r_0+i\omega r_1-\frac{1}{2!}\omega^2 r_2-\frac{1}{3!}i\omega^3 r_3+\frac{1}{4!}\omega^4 r_4+\ldots$$

$$=\sum_{n=0}^{\infty}\frac{(i\omega)^n}{n!}r_n \quad (2\text{-}75),$$

where r_n is the n-th moment about 0 (called raw moment).

Probability density function $\mathbb{P}(x)$ can be obtained by inverse Fourier transform of the characteristic function using the equation (2-29):

$$\mathbb{P}(x)=\mathcal{F}^{-1}\left[\phi(\omega)\right]=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega x}\phi(\omega)d\omega \quad (2\text{-}76).$$

If X is a discrete random variable with possible values $\{x_k\}_{k=0}^{\infty}$ and $\Pr\{X = x_k\} = a_k$, then $\phi(\omega)$ is obtained by a series expansion instead of an integration as:

$$\phi(\omega) \equiv \sum_{k=0}^{\infty} \exp(i\omega x_k) a_k \qquad (2\text{-}77).$$

2.6.2 Properties of a Characteristic Function

Let X be a random variable and $\phi_X(\omega)$ be its characteristic function. A characteristic function $\phi_X(\omega)$ is: 1) bounded by 1 (i.e. $|\phi_X(\omega)| \le 1,\ \omega \in \mathbb{R}$), 2) $\phi_X(0) = 1$, and 3) uniformly continuous in $\mathbb{R}$. A statistical distribution is uniquely determined by its characteristic function, i.e. one-to-one relationship between distribution functions and characteristic functions. In other words, if two random variables X and Y have the same characteristic functions (i.e. $\phi_X(\omega) = \phi_Y(\omega)$), they have the same distribution.

If $\{X_k, k = 1,...,n\}$ are independent random variables, the characteristic function of their sum $X_1 + X_2 + ... + X_n$ is the product of their characteristic functions:

$$\phi_{X_1+X_2+...+X_n}(\omega) = \prod_{k=1}^{n} \phi_{X_k}(\omega) \qquad (2\text{-}78).$$

A random variable X has a symmetric probability density function $\mathbb{P}(x)$ if and only if its characteristic function $\phi_X(\omega)$ is a real-valued function, i.e.:

$$\phi_X(\omega) \in \mathbb{R} \text{ for } \omega \in \mathbb{R}.$$

For $a, b \in \mathbb{R}$:

$$\phi_{aX+b}(\omega) = e^{i\omega b}\phi_X(a\omega) \qquad (2\text{-}79).$$

2.6.3 Characteristic Exponent: Cumulant-Generating Function

A characteristic exponent of a random variable X , $\Psi_X(\omega)$, is defined as a log of a characteristic function $\phi_X(\omega)$:

$$\Psi_X(\omega) \equiv \ln \phi_X(\omega) \qquad (2\text{-}80).$$

The n-th cumulant is defined as:

$$cumulant_n = \frac{1}{i^n}\frac{\partial^n \Psi_X(\omega)}{\partial \omega^n}\Big|_{\omega=0} \qquad (2\text{-}81).$$

Mean, variance, skewness, and excess kurtosis of the random variable X can be obtained from cumulants as follows:

$$\text{Mean of } X = E[X] = cumulant_1 \qquad (2\text{-}82),$$

$$\text{Variance of } X = E\left[\left(X - E(X)\right)^2\right] = cumulant_2,$$

$$\text{Skewness of } X = \frac{E\left[\left(X - E(X)\right)^3\right]}{\left(\sqrt{E\left[\left(X - E(X)\right)^2\right]}\right)^3} = \frac{cumulant_3}{(cumulant_2)^{3/2}},$$

$$\text{Excess kurtosis of } X = \frac{E\left[\left(X - E(X)\right)^4\right]}{\left(\sqrt{E\left[\left(X - E(X)\right)^2\right]}\right)^4} - 3 = \frac{cumulant_4}{(cumulant_2)^2}.$$

Let's consider one fundamental example. A normal random variable X with mean μ and variance σ^2 has a density:

$$\mathbb{P}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{\left(X-\mu\right)^2}{2\sigma^2} \right\}.$$

Its characteristic function can be calculated as from the definition (2-72):

$$\phi(\omega) \equiv \mathcal{F}\left[\mathbb{P}(x)\right] \equiv \int_{-\infty}^{\infty} e^{i\omega x} \mathbb{P}(x) dx = \exp(i\mu\omega - \frac{\sigma^2\omega^2}{2}) .$$

Its characteristic exponent is:

$$\Psi(\omega) \equiv \ln \phi(\omega) = \ln\left\{ \exp(i\mu\omega - \frac{\sigma^2\omega^2}{2}) \right\} = i\mu\omega - \frac{\sigma^2\omega^2}{2} .$$

Cumulants are calculated using $\Psi(\omega)$ from (2-82):

$$cumulant_1 = \mu ,$$

$$cumulant_2 = \sigma^2 ,$$

$$cumulant_3 = 0 ,$$

$$cumulant_4 = 0 .$$

This tells us that a normal random variable X has a mean μ and variance σ^2, zero skewness, and zero excess kurtosis.

2.6.4 Laplace Transform

For nonnegative random variables we replace the Fourier transform by Laplace transform in order to obtain characteristic functions. The (unilateral) Laplace transform $\mathcal{L}$ of a function $f(x)$ is defined as:

$$\mathcal{L}[f(x)] \equiv \int_0^{\infty} f(x) e^{-\omega x} dx \qquad (2\text{-}83),$$

where $f(x)$ is defined for $x \geq 0$. Thus, the characteristic function of a nonnegative random variable X with its density function $\mathbb{P}(x)$ is given by:

$$\phi_X(\omega) = \mathcal{L}[\mathbb{P}(x)] \equiv \int_0^\infty \mathbb{P}(x) e^{-\omega x} dx \qquad (2\text{-}84).$$

2.6.5 Relationship with Moment Generating Function

Let X be a random variable on $\mathbb{R}$ and $\mathbb{P}(x)$ be its probability density function. A function $M(\omega)$ with $\omega \in \mathbb{R}$ is called a moment generating function if there exists an $h > 0$ for $|\omega| < h$ such that (i.e. if the expectation in (2-85) converges):

$$M(\omega) \equiv \int_{-\infty}^\infty e^{\omega x} \mathbb{P}(x) dx \equiv E[\exp(\omega x)] \qquad (2\text{-}85).$$

For a continuous random variable X, again using the equation (2-74):

$$\begin{aligned} M(\omega) &\equiv \int_{-\infty}^\infty e^{\omega x} \mathbb{P}(x) dx \\ &= \int_{-\infty}^\infty \left(1 + \omega x + \frac{1}{2!}(\omega x)^2 + \frac{1}{3!}(\omega x)^3 + \frac{1}{4!}(\omega x)^4 + \ldots\right) \mathbb{P}(x) dx \\ &= 1 + \omega c_1 + \frac{1}{2!}\omega^2 c_2 + \frac{1}{3!}\omega^3 c_3 + \frac{1}{4!}\omega^4 c_4 + \ldots, \end{aligned}$$

where c_n is the n-th central moment.

If $\{X_k, k = 1, \ldots, n\}$ are independent random variables, the moment generating function of their sum $X_1 + X_2 + \ldots + X_n$ is the product of their moment generating functions:

$$M_{X_1 + X_2 + \ldots + X_n}(\omega) = \prod_{k=1}^n M_{X_k}(\omega) \qquad (2\text{-}86).$$

Its proof is very simple:

$$M_{X_1+X_2+\ldots+X_n}(\omega) = E\left[\exp\left\{\omega\left(X_1+X_2+\ldots+X_n\right)\right\}\right]$$

$$= E\left[\exp\left\{\omega X_1+\omega X_2+\ldots+\omega X_n\right\}\right]$$

$$= E\left[\exp\left\{\omega X_1\right\}\exp\left\{\omega X_2\right\}\ldots\exp\left\{\omega X_n\right\}\right]$$

$$= E\left[\exp\left\{\omega X_1\right\}\right]E\left[\exp\left\{\omega X_2\right\}\right]\ldots E\left[\exp\left\{\omega X_n\right\}\right]$$

$$= M_{X_1}(\omega)M_{X_2}(\omega)\ldots M_{X_n}(\omega) = \prod_{k=1}^{n} M_{X_k}(\omega).$$

If the moment generating function $M(\omega)$ is differentiable at zero (defined on a neighborhood $[-\varepsilon,\varepsilon]$ of zero), then the n-th raw moments r_n can be obtained by:

$$r_n = \frac{\partial^n M(\omega)}{\partial \omega^n}\Big|_{\omega=0} \qquad (2\text{-}87).$$

Thus:

$$r_1 = M'(0) = E[X],$$

$$r_2 = M''(0) = E[X^2],$$

$$r_3 = M'''(0) = E[X^3],$$

$$r_4 = M''''(0) = E[X^4].$$

For example, the mean and variance of the random variable X are computed using raw moments as:

$$\text{Mean of } X = E[X] = r_1,$$

$$\text{Variance of } X = E[X^2] - E[X]^2 = r_2 - r_1^2.$$

A characteristic function is always well-defined since it is the Fourier transform of a probability measure. But because the integral in (2-85) ($\forall \omega \in \mathbb{R}$) may not converge for some (all) values of ω, a moment generating function is not always well-defined. When $M(\omega)$ is well-defined, the relationship between the moment generating function $M(\omega)$ and the characteristic function $\phi(\omega)$ is given by:

$$M(\omega) = \phi(-i\omega) \qquad (2\text{-}88).$$

Let's consider one fundamental example. If $X \sim Normal(\mu, \sigma)$, its moment generating function can be calculated as following the definition (2-85):

$$M(\omega) \equiv \int_{-\infty}^{\infty} e^{\omega x} \mathbb{P}(x) dx \equiv E[\exp(\omega x)] = \exp\left(\mu\omega + \frac{\sigma^2\omega^2}{2}\right).$$

You can confirm that (2-88) is true for the normal case. Raw moments are calculated as the following:

$$r_1 = M'(0) = \mu,$$

$$r_2 = M''(0) = \mu^2 + \sigma^2,$$

$$r_3 = M'''(0) = \mu^3 + 3\mu\sigma^2,$$

$$r_4 = M''''(0) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4.$$

Using these raw moments, central moments can be calculated as:

$$E[X] = r_1 = \mu,$$

$$Variance[X] = E[X^2] - E[X]^2 = r_2 - r_1^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2,$$

$$Skewness[X] = \frac{E\{X - E[X]\}^3}{(\sqrt{E\{X - E[X]\}^2})^3} = \frac{2r_1^3 - 3r_1r_2 + r_3}{(r_2 - r_1^2)^{3/2}} = 0,$$

$$Excess\ Kurtosis[X] = \frac{E\{X - E[X]\}^4}{(\sqrt{E\{X - E[X]\}^2})^4} - 3$$

$$= \frac{-6r_1^4 + 12r_1^2 r_2 - 3r_2^2 - 4r_1 r_3 + r_4}{(r_2 - r_1^2)^2} = 0.$$

2.6.6 Summary: How to Calculate Standardized Moments from Characteristic Function and Moment Generating function

Table 2.4 summarizes the relationship between standardized moments, $cumulant_n$, and raw moments r_n. Let X be a random variable. n-th cumulant and n-th raw moment are defined by (2-81) and (2-87).

2.6.7 Examples of Characteristic Functions

Several examples of characteristic functions using the definition (2-72) for continuous distributions and (2-77) for discrete distributions are given in the Table 2.5.

2.7 Option Pricing with Fourier Transform: Black-Scholes Example

In this section, we present Fourier transform option pricing approach by Carr and Madan (1999). To illustrate the idea clearly and simply, we present BS model with Fourier transform pricing methodology.

2.7.1 Motivation

Let $\mathbb{Q} \sim \mathbb{P}$ be an equivalent martingale measure under which the discounted asset price process $\{e^{-rt}S_t; 0 \le t \le T\}$ becomes a martingale and $\{\mathcal{F}_t; 0 \le t \le T\}$ be an

information flow of the asset price S (i.e. filtration). In an arbitrage-free market, prices of any assets can be calculated as expected terminal payoffs under $\mathbb{Q}$ discounted by a risk-free interest rate r:

$$e^{-rt}S_t = E^{\mathbb{Q}}[e^{-rT}S_T \,|\mathcal{F}_t],$$

$$S_t = e^{-r(T-t)}E^{\mathbb{Q}}[S_T \,|\mathcal{F}_t],$$

which are martingale conditions.

Let K be a strike price and T be an expiration of a contingent claim. Plain vanilla call and put option prices are computed as discounted risk-neutral conditional expectations of the terminal payoffs $(S_T - K)^+ \equiv \max(S_T - K, 0)$ and $(K - S_T)^+ \equiv \max(K - S_T, 0)$:

$$C(t, S_t) = e^{-r(T-t)}E^{\mathbb{Q}}\left[(S_T - K)^+ \,|\mathcal{F}_t\right] \quad \text{(2-89)},$$

$$P(t, S_t) = e^{-r(T-t)}E^{\mathbb{Q}}\left[(K - S_T)^+ \,|\mathcal{F}_t\right] \quad \text{(2-90)}.$$

Intrinsic value of a vanilla call (put) is defined as $(S_t - K)^+$ ($(K - S_t)^+$) which is the value of the call (put) exercised immediately. Obviously, the intrinsic value of out of the out-of-the money option is zero. Current option price minus its intrinsic value $C(t, S_t) - (S_t - K)^+$ ($P(t, S_t) - (K - S_t)^+$) is called a time value of the option.

Let $\mathbb{Q}(S_T \,|\mathcal{F}_t)$ be a probability density function of a terminal asset price S_T under $\mathbb{Q}$ conditional on $\mathcal{F}_t$. Using $\mathbb{Q}(S_T \,|\mathcal{F}_t)$, (2-89) and (2-90) can be rewritten as:

$$C(t, S_t) = e^{-r(T-t)}\left\{\int_K^\infty (S_T - K)\mathbb{Q}(S_T \,|\mathcal{F}_t)\,dS_T + \int_0^K (0)\mathbb{Q}(S_T \,|\mathcal{F}_t)\,dS_T\right\}$$

$$= e^{-r(T-t)}\int_K^\infty (S_T - K)\mathbb{Q}(S_T \,|\mathcal{F}_t)\,dS_T \quad \text{(2-91)},$$

$$P(t,S_t)=e^{-r(T-t)}\left\{\int_K^{\infty}(0)\mathbb{Q}\left(S_T\left|\mathcal{F}_t\right.\right)dS_T+\int_0^K\left(K-S_T\right)\mathbb{Q}\left(S_T\left|\mathcal{F}_t\right.\right)dS_T\right\}$$

$$=e^{-r(T-t)}\int_0^K\left(K-S_T\right)\mathbb{Q}\left(S_T\left|\mathcal{F}_t\right.\right)dS_T \qquad (2\text{-}92).$$

Black-Scholes (BS) assumes that a terminal stock price S_T conditional on $\mathcal{F}_t$ is a log-normal random variable with its density given by:

$$\mathbb{Q}\left(S_T\left|\mathcal{F}_t\right.\right)=\frac{1}{S_T\sqrt{2\pi\sigma^2(T-t)}}\exp\left[-\frac{\left\{\ln S_T-\left(\ln S_t+(r-\frac{1}{2}\sigma^2)(T-t)\right)\right\}^2}{2\sigma^2(T-t)}\right].$$

Therefore, BS option pricing formula comes down to single integration problem with respect to S_T since all parameters and variables are known:

$$C_{BS}(t,S_t)$$

$$=e^{-r\tau}\int_K^{\infty}\left(S_T-K\right)\frac{1}{S_T\sqrt{2\pi\sigma^2\tau}}\exp\left[-\frac{\left\{\ln S_T-\left(\ln S_t+(r-\frac{1}{2}\sigma^2)\tau\right)\right\}^2}{2\sigma^2\tau}\right], \qquad (2\text{-}93)$$

$$P_{BS}(t,S_t)$$

$$=e^{-r\tau}\int_0^K\left(K-S_T\right)\frac{1}{S_T\sqrt{2\pi\sigma^2\tau}}\exp\left[-\frac{\left\{\ln S_T-\left(\ln S_t+(r-\frac{1}{2}\sigma^2)\tau\right)\right\}^2}{2\sigma^2\tau}\right]. \qquad (2\text{-}94)$$

This implies that as far as a conditional risk-neutral density of the terminal stock price $\mathbb{Q}(S_T\left|\mathcal{F}_t\right.)$ is given, plain vanilla option pricing reduces to single integration problem.

But for general exponential Lévy models $\mathbb{Q}(S_T|\mathcal{F}_t)$ cannot be expressed using special functions of mathematics or is not known. Therefore, we cannot price plain vanilla options using (2-91) and (2-92). So how do we price options using general exponential Lévy models? The answer is to use a very interesting fact that characteristic functions of general exponential Lévy processes are always known in closed-forms or can be expressed in terms of special functions of mathematics although their probability densities are not. If we can somehow rewrite (2-91) and (2-92) in terms of a characteristic function of $S_T|\mathcal{F}_t$ (i.e. log of $S_T|\mathcal{F}_t$ to be more precise) instead of its probability density $\mathbb{Q}(S_T|\mathcal{F}_t)$, we will be able to price options in general exponential Lévy models.

2.7.2 Derivation of Call Price with Fourier Transform: Carr and Madan (1999)

For simplicity, assume $t = 0$ without loss of generality. From the equation (2-91), use a change of variable technique from S_T to $\ln S_T$:

$$C(T,K) = e^{-rT}\int_{\ln K}^{\infty}\left(e^{\ln S_T} - e^{\ln K}\right)\mathbb{Q}\left(\ln S_T|\mathcal{F}_0\right)d\ln S_T.$$

Let s_T be a log terminal stock price and k be a log strike price, i.e. $s_T \equiv \ln S_T$ and $k \equiv \ln K$. Thus, we have:

$$C(T,k) = e^{-rT}\int_{k}^{\infty}\left(e^{s_T} - e^{k}\right)\mathbb{Q}\left(s_T|\mathcal{F}_0\right)ds_T \qquad (2\text{-}95),$$

where $\mathbb{Q}(s_T) \equiv \mathbb{Q}\left(s_T|\mathcal{F}_0\right)$ (for simplicity) is a risk-neutral density of a log terminal stock price s_T conditional on filtration $\mathcal{F}_0$. From the equation (2-72), a characteristic function of s_T is a Fourier transform of its density function $\mathbb{Q}(s_T)$:

$$\phi_T(\omega) \equiv \mathcal{F}\left[\mathbb{Q}(s_T)\right](\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega s_T}\mathbb{Q}(s_T)ds_T \qquad (2\text{-}96).$$

Consider a function $g(t)$. Sufficient (but not necessary) condition for the existence of Fourier transform and its inverse is the equation (2-71):

$$\int_{-\infty}^{\infty}\left|g(t)\right|dt < \infty \qquad (2\text{-}97)$$

Black-Scholes (1973) models a log terminal stock price s_T as a normal random variable with its normal density given by under $\mathbb{Q}$:

$$\mathbb{Q}(s_T) = \frac{1}{\sqrt{2\pi\sigma^2 T}}\exp\left[-\frac{\left\{s_T - \left(s_0 + (r - \frac{1}{2}\sigma^2)T\right)\right\}^2}{2\sigma^2 T}\right] \qquad (2\text{-}98).$$

From the equations (2-96) and (2-98), a characteristic function of BS log terminal stock price s_T is easily obtained as:

$$\phi_T(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega s_T}\mathbb{Q}(s_T)ds_T = \exp\left[i\left\{s_0 + (r - \frac{1}{2}\sigma^2)T\right\}\omega - \frac{\left(\sigma^2 T\right)\omega^2}{2}\right] \qquad (2\text{-}99).$$

When a call price is expressed in terms of a log strike price $k \equiv \ln K$ in the equation (2-95), k approaches $-\infty$ as a strike price K approaches 0 in the limit. Thus, from (2-95):

$$C(T,k) = e^{-rT}\int_{-\infty}^{\infty}\left(e^{s_T} - e^{-\infty}\right)\mathbb{Q}\left(s_T\left|\mathcal{F}_0\right.\right)ds_T = e^{-rT}\int_{-\infty}^{\infty} e^{s_T}\mathbb{Q}\left(s_T\left|\mathcal{F}_0\right.\right)ds_T$$

$$C(T,k) = e^{-rT}E^{\mathbb{Q}}\left[e^{s_T}\left|\mathcal{F}_0\right.\right] \qquad (2\text{-}100).$$

We know under equivalent martingale measure $\mathbb{Q}$:

$$E^{\mathbb{Q}}\left[S_T \equiv e^{s_T}\left|\mathcal{F}_0\right.\right] = S_0 e^{rT}.$$

Equation (2-100) becomes:

$$C(T,k) = S_0 .$$

Therefore, a call price $C(T,k)$ is not integrable (i.e. $C(T,k)$ does not satisfy (2-97)). Therefore, $C(T,k)$ cannot be Fourier transformed. To solve this problem, CM defines a modified call price as:

$$C_{\text{mod}}(T,k) \equiv e^{\alpha k} C(T,k) \qquad (2\text{-}101),$$

where $C_{\text{mod}}(T,k)$ is expected to satisfy the integrability condition (2-97) by carefully choosing $\alpha > 0$:

$$\int_{-\infty}^{\infty} \left| C_{\text{mod}}(T,k) \right| dk < \infty .$$

Consider a FT of a modified call price by the FT definition (2-28):

$$\psi_T(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega k} C_{\text{mod}}(T,k) dk \qquad (2\text{-}102).$$

From (2-102), call price $C(T,k)$ can be obtained by an inverse FT (i.e. the definition (2-29)) of $\psi_T(\omega)$:

$$C_{\text{mod}}(T,k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega k} \psi_T(\omega) d\omega$$

$$e^{\alpha k} C(T,k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega k} \psi_T(\omega) d\omega$$

$$C(T,k) = \frac{e^{-\alpha k}}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega k} \psi_T(\omega) d\omega \qquad (2\text{-}103).$$

Now CM derives an analytical expression of $\psi_T(\omega)$ in terms of a characteristic function $\phi_T(\omega)$. Substitute (2-101) into (2-102):

$$\psi_T(\omega) = \int_{-\infty}^{\infty} e^{i\omega k} e^{\alpha k} C(T,k) dk \, .$$

Substitute (2-95) and interchanging integrals yields:

$$\begin{aligned}
\psi_T(\omega) &= \int_{-\infty}^{\infty} e^{i\omega k} e^{\alpha k} e^{-rT} \int_{k}^{\infty} \left(e^{s_T} - e^{k}\right) \mathbb{Q}\left(s_T\right) ds_T dk \\
&= \int_{-\infty}^{\infty} e^{-rT} \mathbb{Q}\left(s_T\right) \int_{-\infty}^{s} e^{i\omega k} \left(e^{s_T + \alpha k} - e^{(1+\alpha)k}\right) dk ds_T \\
&= \int_{-\infty}^{\infty} e^{-rT} \mathbb{Q}\left(s_T\right) \left(\frac{e^{(\alpha+1+i\omega)s_T}}{\alpha + i\omega} - \frac{e^{(\alpha+1+i\omega)s_T}}{\alpha + 1 + i\omega} \right) ds_T \\
&= \frac{e^{-rT} \phi_T\left(\omega - (\alpha+1)i\right)}{\alpha^2 + \alpha - \omega^2 + i\left(2\alpha + 1\right)\omega} \qquad \text{(2-104).}
\end{aligned}$$

Thus, a call pricing function is obtained by substituting (2-104) into (2-103):

$$C(T,k) = \frac{e^{-\alpha k}}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega k} \frac{e^{-rT} \phi_T\left(\omega - (\alpha+1)i\right)}{\alpha^2 + \alpha - \omega^2 + i\left(2\alpha + 1\right)\omega} d\omega \qquad \text{(2-105),}$$

where $\phi_T(.)$ is a characteristic function of a log terminal stock price s_T conditional on filtration $\mathcal{F}_0$. We can interpret the equation (2-105) which is single numerical integration problem as a characteristic function $\phi_T(.)$ equivalent of the equation (2-89).

2.7.3 How to Choose Decay Rate Parameter α: Carr and Madan (1999)

We mentioned earlier that when a call price is expressed in terms of a log strike price $k \equiv \ln K$ in the equation (2-95), k approaches $-\infty$ as a strike price K approaches 0 in the limit which is illustrated in Figure 2.17. Thus, a call price function $C(T,k)$ becomes S_0 as $k \to -\infty$ discussed before. In order to make a call price integrable, CM multiplies an exponential function $e^{\alpha k}$ with $\alpha \in \mathbb{R}^+$ to $C(T,k)$

and obtains a modified call price $C_{\text{mod}}(T,k)$. As shown by Figure 2.18, the role of $e^{\alpha k}$ with $\alpha \in \mathbb{R}^+$ is to dampen the size of $C(T,k) = S_0$ in the limit $k \to -\infty$. But this in turn worsens the integrability condition for $k \in \mathbb{R}^+$ (i.e. positive log strike) axis. In order for a modified call price $C_{\text{mod}}(T,k) \equiv e^{\alpha k} C(T,k)$ to be integrable for both positive and negative k axis (i.e. square integrable), CM provides a sufficient condition:

$$\psi_T(0) < \infty .$$

From (2-104):

$$\psi_T(0) = \frac{e^{-rT}\phi_T\left(-(\alpha+1)i\right)}{\alpha^2+\alpha} .$$

Therefore, the sufficient condition of the square-integrability of $C_{\text{mod}}(T,k)$ is:

$$\phi_T\left(-(\alpha+1)i\right) < \infty \qquad \text{(2-106)}.$$

From the definition of a characteristic function:

$$\phi_T(\omega) = E[e^{i\omega s_T}] = E[e^{i\omega \ln S_T}] = E[(e^{\ln S_T})^{i\omega}] = E[{S_T}^{i\omega}] \qquad \text{(2-107)}.$$

From (2-106) and (2-107):

$$\phi_T(-(\alpha+1)i) < \infty$$

$$E[{S_T}^{i\{-(\alpha+1)i\}}] < \infty$$

$$E[{S_T}^{\alpha+1}] < \infty \qquad \text{(2-108)}.$$

CM suggests the use of (2-108) and the analytical expression of the characteristic function to determine an upper bound on α.

2.7.4 Black-Scholes Model with Fourier Transform Pricing Method

By substituting a characteristic function of BS log terminal stock price s_T of the equation (2-98) into the general FT pricing formula of the equation (2-105), we obtain BS-FT call pricing formula:

$$C_{BS-FT}(T,k)=\frac{e^{-\alpha k}}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega k}\frac{e^{-rT}\phi_T\left(\omega-(\alpha+1)i\right)}{\alpha^2+\alpha-\omega^2+i\left(2\alpha+1\right)\omega}d\omega \qquad \text{(2-109)},$$

with $\phi_T(\omega)=\exp\left[i\left\{s_0+(r-\frac{1}{2}\sigma^2)T\right\}\omega-\frac{\left(\sigma^2T\right)\omega^2}{2}\right]$.

Chapter 3: Calibration of Lévy Option Pricing Models: Application to S&P 500 Futures Option

3.1 Introduction

Since the mid 1990's, there was an explosion of literatures regarding so called pure jump Lévy models which try to model the stochastic process of the underlying asset price with pure jump Lévy processes. Matsuda (2005a) defines a Lévy process as follows.

Definition 1 Lévy processes A real valued stochastic process $(X_{t\in[0,\infty)})$ on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ is said to be a Lévy process on $\mathbb{R}$ if it satisfies the following conditions:

(1) Its increments are independent. In other words, for $0 \le t_1 < t_2 < ... < t_n < \infty$:

$$\mathbb{P}(X_{t_0} \bigcap X_{t_1} - X_{t_0} \bigcap X_{t_2} - X_{t_1} \bigcap ... \bigcap X_{t_n} - X_{t_{n-1}})$$

$$= \mathbb{P}(X_{t_0})\mathbb{P}(X_{t_1} - X_{t_0})\mathbb{P}(X_{t_2} - X_{t_1})...\mathbb{P}(X_{t_n} - X_{t_{n-1}}).$$

(2) Its increments are stationary (time homogeneous): i.e. for $h \ge 0$, $X_{t+h} - X_t$ has the same distribution as X_h. In other words, the distribution of increments does not depend on t.

(3) $\mathbb{P}(X_0 = 0) = 1$. The process starts from 0 almost surely (with probability 1).

(4) The process is stochastically continuous: $\forall \varepsilon > 0$, $\lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \ge \varepsilon) = 0$.

(5) Its sample path (trajectory) is right continuous with left limit (i.e. rcll) almost surely.

As you can see, the fact that left continuity is not needed allows the process to have jumps. A continuous stochastic process implies a rcll stochastic process but the reverse is not true. All stochastic processes used in finance literatures for the

modeling of asset price dynamics are rcll stochastic processes. Rcll processes include jump discontinuous process such as Poisson processes and infinite activity Lévy processes.

Let $(L_{t\in[0,T]})$ be a Lévy process defined on a filtered risk-neutral probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{Q})$. Lévy models specify the dynamics of an increment of an underlying asset price as:

$$dS = rSdt + SdL.$$

This means that the classic BS model is a Lévy model when the choice of a Lévy process is a multiplicative Brownian motion:

$$(L_{t\in[0,T]}) \equiv (\sigma B_{t\in[0,T]}).$$

It turns out that a Brownian motion is the only Lévy process with continuous sample paths. This continuous sample paths property of the underlying price is another defect of the BS model which is not supported by the empirical evidence. It is a well established fact that asset prices do jump especially in the downward direction. The first jump Lévy model was developed by Merton (1976). Merton simply adds a compound Poisson process which is a pure jump Lévy process to the Brownian motion. Addition of a compound Poisson process allows the asset price to have rare jumps.

In contrast to the MJD (Merton jump diffusion) model which possesses mostly continuous with rare discontinuous sample paths, more recent Lévy models possess pure jump sample paths. They are labeled as pure jump Lévy models or purely non-Gaussian Lévy models. Carr, Chang, and Madan (1998) developed the variance

gamma (VG) model, Barndorff-Nielsen created the normal inverse Gaussian (NIG) model, and Carr, Geman, Madan, and Yor (2002) developed the CGMY model. Pure jump Lévy models are probably the most drastic revolution in the finance literature since the BS model because they no longer contain a Brownian motion. Note that the well-studied and popular stochastic volatility models which introduce extra parameters by making the volatility random and these extra parameters control the non-normality of log returns still contain a Brownian motion. Consult Hull and White (1987) and Heston (1993) for stochastic volatility models. In contrast to modeling the volatility, the pure jump Lévy models specify the underlying asset price process as pure jump Lévy processes introducing parameters which control the non-normality of log returns.

The goal of this paper is to give an answer to the following simple question, "What do we gain by allowing the asset price process to jump or by specifying the asset price process as a pure jump Lévy process?" For this purpose, we calibrate a total of five different Lévy models to the S&P 500 futures options. The gain of jump Lévy models over the BS model is judged by two main criterions. One is the implied dynamics of the log return probability density and the Lévy density. The other is the out-of-sample fit of the each Lévy model following Bakshi, Cao, and Chen (1997) because of the fact that the better in-sample fit might be due to more parameters of the jump Lévy models than the BS model. In other words, if the extra parameters of the jump Lévy models are redundant, they overfit the data and may produce larger out-of-sample fitting errors.

The calibration result based on 6567 call option prices on S&P 500 futures with March 2005 maturity for the sample period from March 24, 2004, through March 16, 2005 suggests that the extra parameters of various jump Lévy models allow the negative skewness and the excess kurtosis of the log return density over the BS model. On the dynamics of implied Lévy density functions of log returns, the pattern of increasing total mass of the MJD model's symmetric Lévy density is observed. Lévy densities of all pure jump Lévy models are characterized by the asymmetric and infinite activity Lévy densities which all resemble one another. Containing jumps in the underlying price process drastically improves the out-of-sample pricing performance over the BS model. Among jump Lévy models, the CGMY model turned out to be the best performing model for the overall option pricing.

This paper is organized as follows. Section 2 presents three pure jump (i.e. purely non-Gaussian) Lévy models in which the asset price dynamics is modeled using pure jump Lévy process. Section 3 reviews the Gaussian Lévy model and the non-Gaussian Lévy model. Section 4 describes the S&P 500 futures option data set and our calibration methodology. Section 5 provides the implied dynamics of the probability density and the Lévy measure of the log returns. In-sample pricing performance of each model is discussed. Section 6 evaluates the out-of-sample pricing performance of each model. Section 7 concludes.

3.2 Pure Jump Lévy Models

In this section, we present three pure jump Lévy models in which the asset price dynamics is modeled using pure jump Lévy process. Intuitively, this means that the asset price moves only by jumps. Pure jump Lévy processes are also called as purely non-Gaussian Lévy processes because they have zero Gaussian variance following Lévy-Itô decomposition of the sample paths. There are several different ways of building pure jump Lévy processes[5]. The variance gamma process and the normal inverse Gaussian process which are described in section 2.1 and 2.2 are the class of Lévy process called the subordinated Brownian motion with the tempered stable subordinator. The CGMY model in section 2.3 builds the Lévy process by specifying the Lévy measure and the use of Lévy-Khinchin representation.

3.2.1 Variance Gamma Model by Carr, Chang, and Madan (1998)

Consider a fixed filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. Let $(W_{t\in[0,\infty)}) \equiv (\theta t + \sigma B_{t\in[0,\infty)})$ be a Brownian motion with drift where $(B_{t\in[0,\infty)})$ is a standard Brownian motion. Let $(Z_{t\in[0,\infty)})$ be the gamma subordinator[6] with unit mean rate with the Lévy triplet $(A_Z = 0, \ell_Z, \gamma_Z = 0)$ whose probability density is one parameter family of the variance rate $\nu \in \mathbb{R}^+$:

$$\mathbb{P}_G(z_t;\nu) = \frac{z^{\frac{t}{\nu}-1}\exp\left(-\frac{z}{\nu}\right)}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)} 1_{z>0} \qquad (3.1),$$

[5] We recommend Cont and Tankov (2004).

[6] Subordinator is defined as an increasing Lévy process.

its characteristic function is:

$$\phi_Z(\omega) = \frac{1}{\left(1 - i\omega\nu\right)^{\frac{t}{\nu}}} \qquad (3\text{-}2),$$

and its moment generating function[7] is:

$$M_Z(\omega) = \frac{1}{\left(1 - \omega\nu\right)^{\frac{t}{\nu}}} = \exp\left(t\mathcal{L}_Z(\omega)\right) \qquad (3\text{-}3),$$

where the Laplace exponent is given by:

$$\mathcal{L}_Z(\omega) = -\frac{1}{\nu}\ln\left(1 - \omega\nu\right) \qquad (3\text{-}4).$$

Assume the statistical independence between $(W_{t\in[0,\infty)})$ and $(Z_{t\in[0,\infty)})$. Then, the variance gamma (VG) process $(X_{t\in[0,\infty)})$ is defined as a subordinated Brownian motion with the gamma subordinator with unit mean rate:

$$X(t,\omega) \equiv W(Z_t(\omega),\omega) \equiv \theta Z_t(\omega) + \sigma B(Z_t(\omega),\omega) \qquad (3\text{-}5).$$

This VG process $(X_{t\in[0,\infty)})$ with the Lévy triplet $(A_X = 0, \ell_X, \gamma_X)$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ possesses the following properties. It is a pure jump Lévy process (i.e. a purely non-Gaussian Lévy process) which is equivalent to stating that the Gaussian variance term is zero $A_X = 0$ following the Lévy-Itô decomposition of the sample paths. Its characteristic function can be obtained by the use of the subordination theorem of Lévy processes which corresponds to Theorem 30.1 of Sato (1999):

$$\phi_X(\omega) \equiv \int_{-\infty}^{\infty} \mathrm{e}^{i\omega x}\mathbb{P}(x)dx = \exp\left\{t\mathcal{L}_Z\left(\psi_W(\omega)\right)\right\} \qquad (3\text{-}6),$$

[7] Moment generating function is defined as $M_X(\omega) \equiv \int_{-\infty}^{\infty} \mathrm{e}^{\omega x}\mathbb{P}(x)dx$.

where $\psi_W(\omega)$ is the characteristic exponent of the Brownian motion with drift:

$$\psi_W(\omega) = i\theta\omega - \frac{\sigma^2\omega^2}{2} \qquad (3\text{-}7).$$

Substitution of (3-4) and (3-7) into (3-6) produces:

$$\phi_{VG,X}(\omega;\theta,\sigma,\nu) = \left(1 - i\nu\theta\omega + \frac{\nu\sigma^2\omega^2}{2}\right)^{-\frac{t}{\nu}} \qquad (3\text{-}8).$$

Its probability density is given by:

$$\mathbb{P}_{VG}(x_t;\theta,\sigma,\nu) = \frac{\sqrt{2}\exp\left(\frac{\theta}{\sigma^2}x_t\right)}{\nu^{t/\nu}\sigma\sqrt{\pi}\Gamma(t/\nu)}\left(\frac{x_t^{\,2}}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{\sqrt{x_t^{\,2}\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}}{\sigma^2}\right) \qquad (3\text{-}9),$$

where K is the modified Bessel function of the second kind. Note that this probability density can be easily computed using the conditionally normal property of the VG process:

$$\mathbb{P}_{VG}(x_t;\theta,\sigma,\nu) = \int_0^\infty VG(x_t \,|\, t = z_t = z)\mathbb{P}_G(z_t;\nu)dz$$

$$\mathbb{P}_{VG}(x_t;\theta,\sigma,\nu) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 z}}\exp\left\{-\frac{(x_t-\theta z)^2}{2\sigma^2 z}\right\}\frac{z^{\frac{t}{\nu}-1}\exp\left(-\frac{z}{\nu}\right)}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)}dz\,.$$

Its standardized moments are computed by:

$$E[X_t] = \theta t \qquad (3\text{-}10),$$

$$Variance[X_t] = (\nu\theta^2 + \sigma^2)t\,,$$

$$Skewness[X_t] = \frac{\nu\theta(2\nu\theta^2 + 3\sigma^2)}{\sqrt{t}(\nu\theta^2+\sigma^2)^{3/2}}\,,$$

$$Excess\ Kurtosis[X_t] = \frac{3\nu(2\nu^2\theta^4 + 4\nu\theta^2\sigma^2 + \sigma^4)}{t(\nu\theta^2 + \sigma^2)^2}.$$

These standardized moments indicate that $\theta \in \mathbb{R}$ is a location and a skewness parameter, $\sigma \in \mathbb{R}^+$ is a shape parameter, and $\nu \in \mathbb{R}^+$ controls the tail behavior[8] of the probability density (and also the Lévy density) of the VG process.

The Lévy measure of the VG process is of the form:

$$\ell_{VG,X}(x;\theta,\sigma,\nu) = \frac{1}{\nu|x|}\exp\left(\frac{\theta}{\sigma^2}x - \frac{|x|}{\sigma^2}\sqrt{\frac{2\sigma^2}{\nu} + \theta^2}\right) \quad (3\text{-}11).$$

The total mass of the Lévy measure of the VG process is infinite:

$$\int_{-\infty}^{\infty} \ell_{VG,X}(x)dx = \infty.$$

In other words, the VG process is an infinite activity Lévy process which means that the VG process has infinitely many small jumps and a finite number of large jumps inheriting the infinite arrival rate of jumps from the gamma subordinator. The VG process is also a Lévy process of finite variation in the interval $[0,\infty)$. In other words, it is a Lévy process satisfying:

$$A_X = 0 \text{ and } \int_{|x|<1} |x|\ell_{VG,X}(x)dx < \infty.$$

Consult theorem 3.11 of Matsuda (2005a). Examples of the Lévy measure of the VG process (3-11) are plotted in Figure 3.1. The sign of $\theta \in \mathbb{R}$ determines the skewness of the VG Lévy measure as illustrated by Panel A. $\sigma \in \mathbb{R}^+$ controls its shape as shown by Panel B. Larger values of the variance rate $\nu \in \mathbb{R}^+$ of the gamma

[8] The larger value of variance rate ν of the gamma subordinator with unit mean rate (i.e. which determines the degree of randomness of the subordination) makes the probability density of the VG process fatter tailed and higher peaked (i.e. more leptokurtic).

subordinator[9] make the tails of the Lévy measure fatter which indicates the higher arrival rate of large jumps as depicted by Panel C.

The VG model specifies the asset price dynamics $(S_{t\in[0,T]})$ defined on a filtered risk neutral probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{Q})$ as an exponential (geometric) of a Lévy process $(L_{t\in[0,T]})$:

$$S_t = S_0 \exp(L_t),$$

where the choice of the Lévy process is the VG process plus the drift $r - \varpi_{VG,\mathbb{Q}}$:

$$L_t \equiv (r - \varpi_{VG,\mathbb{Q}})t + VG(x_t; \theta_{\mathbb{Q}}, \sigma_{\mathbb{Q}}, \nu_{\mathbb{Q}}) \quad (3\text{-}12),$$

where $r \in \mathbb{R}^+$ is the instantaneous risk-free interest rate and all parameters are under the risk neutral probability measure $\mathbb{Q}$. The term $\varpi_{VG,\mathbb{Q}}$ is the convexity correction which takes the following form in the VG model:

$$\varpi_{VG,\mathbb{Q}} \equiv -\frac{1}{\nu_{\mathbb{Q}}} \ln\left(1 - \theta_{\mathbb{Q}}\nu_{\mathbb{Q}} - \frac{\sigma_{\mathbb{Q}}{}^2 \nu_{\mathbb{Q}}}{2}\right) \quad (3\text{-}13).$$

Define the log return (i.e. log price relative) of the asset price as:

$$R_t \equiv \ln(S_t / S_0) \quad (3\text{-}14).$$

Then, from the equation (3-12):

$$x_t = R_t - (r - \varpi_{VG,\mathbb{Q}})t \quad (3\text{-}15).$$

Since obviously the drift $r - \varpi_{VG,\mathbb{Q}}$ is deterministic, the probability density of the log return in the VG model under the risk neutral probability measure $\mathbb{Q}$ can be expressed using the density (3-9) as:

[9] Which indicates the higher degree of randomness of the subordination.

$$\mathbb{Q}_{VG}(R_t;\theta,\sigma,\nu)$$

$$=\frac{\sqrt{2}\exp\left(\frac{\theta}{\sigma^2}x_t\right)}{\nu^{t/\nu}\sigma\sqrt{\pi}\Gamma(t/\nu)}\left(\frac{x_t^{\,2}}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}}K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{\sqrt{x_t^{\,2}\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}}{\sigma^2}\right) \quad (3\text{-}16),$$

with the equation (3-15). Note that all parameters in the density (3-16) are under the risk neutral probability measure $\mathbb{Q}$.

3.2.2 Normal Inverse Gaussian Model by Barndorff-Nielsen (1998)

Consider a fixed filtered probability space $(\Omega,\mathcal{F}_{t\in[0,\infty)},\mathbb{P})$. Let $(W_{t\in[0,\infty)})\equiv(\beta t+B_{t\in[0,\infty)})$ be a Brownian motion with drift where $(B_{t\in[0,\infty)})$ is a standard Brownian motion. Let $(Z_{t\in[0,\infty)})$ be the inverse Gaussian subordinator with the Lévy triplet $(A_Z=0,\ell_Z,\gamma_Z=0)$ whose probability density is a two parameter family of $\delta\in\mathbb{R}^+$ and $\gamma\equiv\sqrt{\alpha^2-\beta^2}\in\mathbb{R}^+$:

$$\mathbb{P}_{IG}(z_t;\delta,\gamma)=\frac{\delta t}{\sqrt{2\pi}}z_t^{\,-3/2}\exp\left(\delta\gamma t\right)\exp\left\{-\frac{\left(\gamma^2 z_t+\delta^2t^2z_t^{\,-1}\right)}{2}\right\}1_{z>0} \quad (3\text{-}17),$$

its characteristic function is:

$$\phi_Z(\omega)=\exp\left(\delta t\gamma-\delta t\sqrt{\gamma^2-2i\omega}\right) \quad (3\text{-}18),$$

and its moment generating function is:

$$M_Z(\omega)=\exp\left(\delta t\gamma-\delta t\sqrt{\gamma^2-2\omega}\right)=\exp\left(t\mathcal{L}_Z(\omega)\right) \quad (3\text{-}19),$$

where the Laplace exponent is given by:

$$\mathcal{L}_Z(\omega) = \delta\gamma - \delta\sqrt{\gamma^2 - 2\omega} \qquad (3\text{-}20).$$

Assume the statistical independence between $(W_{t\in[0,\infty)})$ and $(Z_{t\in[0,\infty)})$. Then, the normal inverse Gaussian (NIG) process $(X_{t\in[0,\infty)})$ is defined as a subordinated Brownian motion with the IG subordinator plus the drift $\mu \in \mathbb{R}$:

$$X(t,\omega) \equiv \mu t + W(Z_t(\omega),\omega) \equiv \mu t + \beta Z_t(\omega) + B(Z_t(\omega),\omega) \qquad (3\text{-}21).$$

This NIG process $(X_{t\in[0,\infty)})$ with the Lévy triplet $(A_X = 0, \ell_X, \gamma_X)$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ possesses the following properties. It is a pure jump Lévy process. Its characteristic function can be obtained by the use of the subordination theorem of Lévy processes of the equation (6). The characteristic exponent of the Brownian motion with drift is:

$$\psi_W(\omega) = i\beta\omega - \frac{\omega^2}{2} \qquad (3\text{-}22).$$

Substitution of (3-22) and (3-20) into (3-6) after the consideration of the drift μ produces:

$$\phi_{NIG,X}(\omega;\alpha,\beta,\mu,\delta) = \exp\left[i\omega\mu t + t\delta\left\{\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2 - (\beta + i\omega)^2}\right\}\right] \qquad (3\text{-}23).$$

Its probability density is given by:

$$\mathbb{P}_{NIG}(x_t;\alpha,\beta,\mu,\delta)$$

$$= \frac{\alpha}{\pi}\exp\left(\delta t\sqrt{\alpha^2-\beta^2}\right)\exp\left\{\beta(x_t - \mu t)\right\}\frac{K_1\left(\alpha\delta t\sqrt{1+\left(\frac{x_t-\mu t}{\delta t}\right)^2}\right)}{\sqrt{1+\left(\frac{x_t-\mu t}{\delta t}\right)^2}} \qquad (3\text{-}24),$$

where K is the modified Bessel function of the second kind. The domain of the parameters are $\alpha \in \mathbb{R}^+$, $\beta \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\delta \in \mathbb{R}^+$ and $\gamma \equiv \sqrt{\alpha^2 - \beta^2} \in \mathbb{R}^+$. Note that this probability density can be easily computed using the conditionally normal property of the NIG process:

$$\mathbb{P}_{NIG}(x_t; \alpha, \beta, \mu, \delta) = \int_0^\infty NIG(x_t \,|\, t = z_t = z)\mathbb{P}_{IG}(z_t; \delta, \gamma)dz$$

$$\mathbb{P}_{NIG}(x_t; \alpha, \beta, \mu, \delta) =$$

$$\int_0^\infty \frac{1}{\sqrt{2\pi z}} \exp\left\{-\frac{(x_t - \beta z)^2}{2z}\right\} \frac{\delta t z_t^{-3/2}}{\sqrt{2\pi}} \exp\left\{\delta\gamma t - \frac{\left(\gamma^2 z_t + \delta^2 t^2 z_t^{-1}\right)}{2}\right\} dz\,.$$

Its standardized moments are computed by:

$$E[X_t] = \left(\mu + \frac{\beta\delta}{\sqrt{\alpha^2 - \beta^2}}\right)t \qquad \text{(3-25)},$$

$$Variance[X_t] = \frac{\alpha^2 \delta t}{\left(\sqrt{\alpha^2 - \beta^2}\right)^3},$$

$$Skewness[X_t] = \frac{3\beta}{\alpha\sqrt{\delta t}(\alpha^2 - \beta^2)^{1/4}},$$

$$Excess\ Kurtosis[X_t] = \frac{3\left(\alpha^2 + 4\beta^2\right)}{\alpha^2 \delta t \sqrt{\alpha^2 - \beta^2}}.$$

These standardized moments indicate that $\alpha \in \mathbb{R}^+$ is a shape parameter, $\beta \in \mathbb{R}$ is a skewnesss parameter, $\mu \in \mathbb{R}$ is a location parameter, and $\delta \in \mathbb{R}^+$ is an excess kurtosis parameter of the probability density of the NIG process.

The Lévy measure of the NIG process is of the form:

$$\ell_{NIG,X}(x;\alpha,\beta,\delta)=\frac{\alpha\delta}{\pi}\frac{\exp(\beta x)}{|x|}K_1(\alpha|x|) \quad (3\text{-}26).$$

which has the infinite total mass:

$$\int_{-\infty}^{\infty}\ell_{NIG,X}(x)dx=\infty.$$

In other words, the NIG process is an infinite activity Lévy process which means that the NIG process has infinitely many small jumps and a finite number of large jumps inheriting the infinite arrival rate of jumps from the IG subordinator. The NIG process is also a Lévy process of infinite variation in the interval $[0,\infty)$. In other words, it is a Lévy process satisfying:

$$\int_{|x|<1}|x|\ell_{NIG,X}(x)dx=\infty.$$

Consult theorem 3.11 of Matsuda (2005a). Examples of the Lévy measure of the NIG process (3-26) are plotted in Figure 3.2. The sign of $\beta\in\mathbb{R}$ determines the skewness of the NIG Lévy measure as illustrated by Panel A. Larger values of $\alpha\in\mathbb{R}^+$ lead to the smaller overall arrival rate of the NIG Lévy measure as shown by Panel B. Larger values of $\delta\in\mathbb{R}^+$ make the overall arrival rate of jumps larger as depicted by Panel C.

The NIG model specifies the asset price dynamics $(S_{t\in[0,T]})$ defined on a filtered risk neutral probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{Q})$ as an exponential (geometric) of a Lévy process $(L_{t\in[0,T]})$:

$$S_t=S_0\exp(L_t),$$

where the choice of the Lévy process is the NIG process plus the drift $r-\varpi_{VG,\mathbb{Q}}$:

$$L_t \equiv (r - \varpi_{NIG,\mathbb{Q}})t + NIG(x_t; \alpha_{\mathbb{Q}}, \beta_{\mathbb{Q}}, \mu_{\mathbb{Q}}, \delta_{\mathbb{Q}}) \qquad (3\text{-}27),$$

where $r \in \mathbb{R}^+$ is the instantaneous risk-free interest rate and all parameters are under the risk neutral probability measure $\mathbb{Q}$. The term $\varpi_{NIG,\mathbb{Q}}$ is the convexity correction which takes the following form in the NIG model:

$$\varpi_{NIG,\mathbb{Q}} \equiv \mu_{\mathbb{Q}} + \delta_{\mathbb{Q}}\sqrt{\alpha_{\mathbb{Q}}{}^2 - \beta_{\mathbb{Q}}{}^2} - \delta_{\mathbb{Q}}\sqrt{\alpha_{\mathbb{Q}}{}^2 - \left(\beta_{\mathbb{Q}} + 1\right)^2} \qquad (3\text{-}28).$$

Note that this form of the convexity correction (28) necessitates one additional restriction on the parameters:

$$\alpha^2 - \left(\beta + 1\right)^2 \in \mathbb{R}^+ \qquad (3\text{-}29).$$

From the equation (3-27) and the definition (3-14) of the log return, we have the following relationship:

$$x_t = R_t - (r - \varpi_{NIG,\mathbb{Q}})t \qquad (3\text{-}30).$$

Thus, the probability density of the log return in the NIG model under the risk neutral probability measure $\mathbb{Q}$ can be expressed using the density (3-24) as:

$$\mathbb{Q}_{NIG}(R_t; \alpha, \beta, \mu, \delta)$$

$$= \frac{\alpha}{\pi}\exp\left(\delta t\sqrt{\alpha^2 - \beta^2}\right)\exp\left\{\beta\left(x_t - \mu t\right)\right\}\frac{K_1\left(\alpha\delta t\sqrt{1 + \left(\frac{x_t - \mu t}{\delta t}\right)^2}\right)}{\sqrt{1 + \left(\frac{x_t - \mu t}{\delta t}\right)^2}} \qquad (3\text{-}31),$$

with the equation (3-30). Note that all parameters in the density (3-31) are under the risk neutral probability measure $\mathbb{Q}$.

3.2.3 CGMY Model by Carr, Geman, Madan, and Yor (2002)

Koponen (1995) in the field of physics originally proposed a class of Lévy processes called tempered stable processes under the name of truncated Lévy flights. CGMY process is a subclass of tempered stable processes.[10]

Consider a fixed filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$. The CGMY process $(X_{t\in[0,\infty)})$ is defined as a Lévy process with the Lévy triplet $(A_X = 0, \ell_X, \gamma_X = 0)$ whose Lévy measure is given by:

$$\ell_{CGMY,X}(x; C, G, M, Y) = \frac{C\exp\left(-G|x|\right)}{|x|^{1+Y}} 1_{x<0} + \frac{C\exp\left(-Mx\right)}{x^{1+Y}} 1_{x>0} \qquad (3\text{-}32),$$

where $C \in \mathbb{R}^+$, $G \in \mathbb{R}^+$, $M \in \mathbb{R}^+$, and $Y < 2$[11]. Examples of the Lévy measure of the CGMY process (3-32) are plotted in Figure 3.3. The overall arrival rate of jumps is controlled by the parameter C indicating that the larger values of C result in the larger overall arrival rate of jumps as shown by Panel A. Panel B depicts the role of the parameters G and M which are the exponential decay rates of the lower and upper tails of the Lévy measure, respectively. When $G = M$, the Lévy measure is symmetric. When $G < M$, the upper tail of the Lévy measure is decayed more and the lower tail of the Lévy measure is heavier which indicates that the arrival rate of negative jumps is higher than that of large jumps. The most interesting parameter is Y since it controls the activity rate (i.e. total mass) of the Lévy measure and the variation of the CGMY process. Y also determines whether the Lévy measure is a

[10] Consult section 4.5 and remark 4.4 of Cont and Tankov (2004).

[11] The condition $Y < 2$ ensures the integrability of the Lévy measure with respect to x^2 in the neighborhood of zero.

monotonically decreasing function of the jump size $|x|$. Table 3.1 summarizes the role of the parameter Y which is based on the Table 1 of CGMY (2002) and Figure 3.4 illustrates the role of Y.

This CGMY process $(X_{t\in[0,\infty)})$ with the Lévy triplet $(A_X = 0, \ell_X, \gamma_X = 0)$ defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$ possesses the following properties. It is a pure jump Lévy process. Its characteristic function can be obtained by the use of the Lévy-Khinchin representation without truncation of large jumps:

$$\phi_X(\omega) = \exp\left(t\psi_X(\omega)\right),$$

where $\psi_X(\omega)$ is the characteristic exponent of the CGMY process:

$$\psi_X(\omega) = \int_{-\infty}^{\infty} \left\{\exp(i\omega x) - 1 - i\omega x\right\} \ell_{CGMY,X}(x; C, G, M, Y)dx \qquad (3\text{-}33).$$

Following Cont and Tankov (2004), the characteristic exponent of the CGMY process is calculated as:

$$\psi_{CGMY,X}(\omega; C, G, M, Y)$$

$$= CG^Y\Gamma(-Y)\left\{\left(1+\frac{i\omega}{G}\right)^Y - 1 - \frac{i\omega Y}{G}\right\} + CM^Y\Gamma(-Y)\left\{\left(1-\frac{i\omega}{M}\right)^Y - 1 + \frac{i\omega Y}{M}\right\} \qquad (3\text{-}34).$$

In the general case, the probability density function of the CGMY process is not available in closed form. Its standardized moments are computed by:

$$E[X_t] = \gamma_X t = 0 \qquad (3\text{-}35),$$

$$Variance[X_t] = tC\Gamma(2-Y)\left(G^{Y-2} + M^{Y-2}\right),$$

$$Skewness[X_t] = \frac{tC\Gamma(3-Y)\left(-G^{Y-3} + M^{Y-3}\right)}{\left(Variance[X_t]\right)^{3/2}},$$

$$Excess\ Kurtosis[X_t] = \frac{tC\Gamma(4-Y)\left(G^{Y-4} + M^{Y-4}\right)}{\left(Variance[X_t]\right)^2}.$$

The CGMY model specifies the asset price dynamics $(S_{t\in[0,T]})$ defined on a filtered risk neutral probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{Q})$ as an exponential (geometric) of a Lévy process $(L_{t\in[0,T]})$:

$$S_t = S_0 \exp(L_t),$$

where the choice of the Lévy process is the CGMY process plus the drift $r - \varpi_{CGMY,\mathbb{Q}}$:

$$L_t \equiv (r - \varpi_{CGMY,\mathbb{Q}})t + CGMY(x_t; C_{\mathbb{Q}}, G_{\mathbb{Q}}, M_{\mathbb{Q}}, Y_{\mathbb{Q}}) \quad (3\text{-}36),$$

where $r \in \mathbb{R}^+$ is the instantaneous risk-free interest rate. The term $\varpi_{CGMY,\mathbb{Q}}$ is the convexity correction which takes the following form in the CGMY model:

$$\varpi_{CGMY,\mathbb{Q}} = C_{\mathbb{Q}} G_{\mathbb{Q}}{}^{Y_{\mathbb{Q}}} \Gamma(-Y_{\mathbb{Q}}) \left\{ \left(1 + \frac{1}{G_{\mathbb{Q}}}\right)^{Y_{\mathbb{Q}}} - 1 - \frac{Y_{\mathbb{Q}}}{G_{\mathbb{Q}}} \right\} \quad (3\text{-}37)$$

$$+ C_{\mathbb{Q}} M_{\mathbb{Q}}{}^{Y_{\mathbb{Q}}} \Gamma(-Y_{\mathbb{Q}}) \left\{ \left(1 - \frac{1}{M_{\mathbb{Q}}}\right)^{Y_{\mathbb{Q}}} - 1 + \frac{Y_{\mathbb{Q}}}{M_{\mathbb{Q}}} \right\}.$$

Defining the log return as $R_t \equiv \ln(S_t / S_0)$, we have the following relationship from the equation (3-36):

$$x_t = R_t - (r - \varpi_{CGMY,\mathbb{Q}})t \quad (3\text{-}38).$$

3.3 Traditional Lévy Models

In this section, we present two non-pure jump Lévy models in which the asset price dynamics is modeled using the Gaussian Lévy process and the non-Gaussian Lévy process. Intuitively, non-pure jump Lévy processes contain a Brownian motion which is equivalent to stating that the Gaussian variance term A_X is not zero following Lévy-Itô decomposition of the sample paths. The classic Black-Scholes model (1973) is the only continuous (i.e. Gaussian) Lévy model in which the asset price dynamics is modeled by the Brownian motion. Another classic Merton jump diffusion model (1976) is the non-Gaussian Lévy model in which the asset price dynamics is modeled by the jump diffusion process which possesses the discontinuities. Consult Matsuda (2005a) for the detailed treatment of the mathematics of traditional Lévy processes.

3.3.1 Black-Scholes Model (1973)

Let $(B_{t\in[0,\infty)})$ be a standard Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathbb{P})$ with the Lévy triplet $(A_B = 1, \ell_B = 0, \gamma_B = 0)$. Define a multiplicative Brownian motion as $(X_{t\in[0,\infty)}) \equiv (\sigma B_{t\in[0,\infty)})$ with the Lévy triplet $(A_X = \sigma^2, \ell_X = 0, \gamma_X = 0)$. A multiplicative Brownian motion $(X_{t\in[0,\infty)})$ or a Brownian motion in general is the only continuous Lévy process which is equivalent to the Gaussian Lévy process. Its characteristic function can be obtained by the use of the general Lévy-Khinchin representation[12]:

[12] Consult Appendix 3.

$$\phi_X(\omega) = \exp\left(-\frac{\sigma^2 t\omega^2}{2}\right) \quad (3\text{-}39).$$

Its probability density is given by:

$$\mathbb{P}_X(x_t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{x_t^{\,2}}{2\sigma^2 t}\right) \quad (3\text{-}40),$$

where $\sigma \in \mathbb{R}^+$ is a shape parameter.

The Lévy measure of the Brownian motion is zero because it is a stochastic process with continuous sample paths (i.e. no jumps):

$$\ell_X(x) = 0 \quad (3\text{-}41).$$

A Brownian motion is also a Lévy process of infinite variation in the interval $[0,\infty)$.

The BS model specifies the asset price dynamics $(S_{t\in[0,T]})$ defined on a filtered risk neutral probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{Q})$ as an exponential of a Lévy process $(L_{t\in[0,T]})$:

$$S_t = S_0 \exp(L_t),$$

where the choice of the Lévy process is the multiplicative Brownian motion plus the drift $r - \varpi_{BS,\mathbb{Q}}$:

$$L_t \equiv (r - \varpi_{BS,\mathbb{Q}})t + X(x_t; \sigma_{\mathbb{Q}}) \quad (3\text{-}42),$$

where $r \in \mathbb{R}^+$ is the instantaneous risk-free interest rate and all parameters are under the risk neutral probability measure $\mathbb{Q}$. The term $\varpi_{BS,\mathbb{Q}}$ is the convexity correction which takes the following form in the BS model:

$$\varpi_{BS,\mathbb{Q}} \equiv \frac{\sigma_{\mathbb{Q}}^{\,2}}{2} \quad (3\text{-}43).$$

Defining the log return (i.e. log price relative) of the asset price as $R_t \equiv \ln(S_t / S_0)$ and using the equation (3-42):

$$x_t = R_t - (r - \varpi_{BS,\mathbb{Q}})t \qquad (3\text{-}44).$$

Since obviously the drift $r - \varpi_{BS,\mathbb{Q}}$ is deterministic, the probability density of the log return in the BS model under the risk neutral probability measure $\mathbb{Q}$ can be expressed using the probability density (3-40) and the relationship (3-44) as:

$$\mathbb{Q}_{BS}(R_t;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{\left\{R_t - \left(r - \varpi_{BS,\mathbb{Q}}\right)t\right\}^2}{2\sigma^2 t} \right] \qquad (3\text{-}45).$$

3.3.2 Merton Jump Diffusion Model (1976)

Consider a fixed filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$. A jump diffusion process $(X_{t\in[0,\infty]})$ with the Lévy triplet $(A_X = \sigma^2, \ell_X = \lambda f(x), \gamma_X = 0)$ is defined as a Brownian motion plus a compound Poisson process:

$$(X_{t\in[0,\infty]}) \equiv (\sigma B_{t\in[0,\infty]}) + \sum_{i=1}^{N_t} X_i \qquad (3\text{-}46),$$

where $(\sigma B_{t\in[0,\infty]})$ is a multiplicative Brownian motion with the Lévy triplet $(A_B = \sigma^2, \ell_B = 0, \gamma_B = 0)$, $\sum_{i=1}^{N_t} X_i$ is a compound Poisson process with the Lévy triplet $(A_C = 0, \ell_C = \lambda f(x), \gamma_C = 0)$ which is the sum of *i.i.d.* jumps X_i from the jump size probability density $f(x)$, and $(N_{t\in[0,\infty]})$ is a Poisson process with the intensity $\lambda \in \mathbb{R}^+$ which counts the number of random arrival times T_k of an event in the time interval $[0,t]$:

$$N_t = \sum_{k \geq 1} 1_{t \geq T_k} \qquad (3\text{-}47).$$

Note that a Poisson process $(N_{t\in[0,\infty]})$ and the jumps sizes $(X_i)_{i\geq 1}$ are assumed to be independent. This jump diffusion process $(X_{t\in[0,\infty]})$ possesses the following properties. It is a non-Gaussian Lévy process (also called a jump Lévy process), but not a pure jump Lévy process (also called a purely non-Gaussian Lévy process) because the Gaussian variance term of the jump diffusion process A_X is non-zero. In other words, $A_X = \sigma^2 \neq 0$ indicates that the process contains a Brownian motion. The Lévy measure of the jump diffusion process is given by:

$$\ell_{JD,X}(x;\lambda) = \lambda f(x) \qquad (3\text{-}48),$$

where $f(x)$ is the jump size probability density. The total mass of the Lévy measure of the jump diffusion process is the intensity parameter λ because a Lévy measure $\ell(x)$ measures the arrival rate of jumps:

$$\int_{-\infty}^{\infty} \ell_{JD,X}(x)dx = \int_{-\infty}^{\infty} \lambda f(x)dx = \lambda \int_{-\infty}^{\infty} f(x)dx = \lambda < \infty ,$$

which is finite because the number of arrivals of an event is almost surely finite for any $t > 0$ including an infinite time horizon $t = \infty$. In other words, the jump diffusion process is a finite activity Lévy process which means that the process has finite number of small jumps and finite number of large jumps. The jump diffusion process is also a Lévy process of infinite variation in the interval $[0,\infty)$ because $A_X = \sigma^2 \neq 0$. Merton jump diffusion (MJD) model specifies the log return jump size density as the normal, i.e. $X_i \sim i.i.d.\ Normal(\mu, \delta^2)$:

$$f_{MJD}(x) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\delta^2} \right\} \quad (3\text{-}49).$$

Thus, the Lévy measure in MJD model can be expressed as:

$$\ell_{MJD,X}(x;\lambda,\mu,\delta) = \frac{\lambda}{\sqrt{2\pi\delta^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\delta^2} \right\} \quad (3\text{-}50).$$

Figure 3.5 plots the examples of the Lévy measure of the MJD process (3-50). The overall arrival rate of jumps is controlled by the intensity parameter λ indicating that the larger values of λ result in the larger overall arrival rate of jumps. The parameter $\mu \in \mathbb{R}$ controls the location of the Lévy measure and the parameter $\delta \in \mathbb{R}^+$ controls its shape. Apparently, the MJD Lévy measure is always symmetric.

The characteristic function of MJD process can be obtained by the use of the Lévy-Khinchin representation as:

$$\phi_{MJD,X}(\omega;\sigma,\lambda,\mu,\delta) = \exp\left[t\left\{ -\frac{\sigma^2\omega^2}{2} + \lambda\left(\phi_f(\omega) - 1\right) \right\} \right] \quad (3\text{-}51),$$

where ϕ_f is the characteristic function of the jump size density:

$$\phi_f(\omega) = \exp\left(i\omega\mu - \frac{\delta^2\omega^2}{2} \right).$$

The probability density of the MJD process can be computed using the conditionally normal property of the jump diffusion process of the equation (3-46)[13]:

[13] This computation involves the series expansion rather than the integration because a compound Poisson process is a continuous time stochastic process with the discontinuous sample paths.

$$\mathbb{P}_{MJD}(x_t;\sigma,\lambda,\mu,\delta)=\sum_{j=0}^{\infty}\mathbb{P}(x_t\,|N_t=j)\mathbb{P}_{Poisson}(N_t=j)$$

$$=\sum_{j=0}^{\infty}\frac{e^{-\lambda t}(\lambda t)^j}{j!}\frac{1}{\sqrt{2\pi\left(\sigma^2 t+j\delta^2\right)}}\exp\left\{-\frac{\left(x_t-j\mu\right)^2}{2\left(\sigma^2 t+j\delta^2\right)}\right\} \quad (30\text{-}52).$$

Its standardized moments are computed by:

$$E[X_t]=\lambda t\mu \quad (3\text{-}53),$$

$$Variance[X_t]=(\sigma^2+\lambda\delta^2+\lambda\mu^2)t\,,$$

$$Skewness[X_t]=\frac{t\lambda\mu(\mu^2+3\delta^2)}{Variance[X_t]^{3/2}}\,,$$

$$Excess\ Kurtosis[X_t]=\frac{t\lambda(\mu^4+3\delta^4+6\mu^2\delta^2)}{Variance[X_t]^2}\,.$$

These standardized moments indicate that μ is a skewness parameter with $\mu=0$ producing the symmetric probability density. Larger values for λ and σ lead to the larger variance and smaller excess kurtosis of the probability density.

MJD model specifies the asset price dynamics $(S_{t\in[0,T]})$ defined on a filtered risk neutral probability space $(\Omega,\mathcal{F}_{t\in[0,T]},\mathbb{Q})$ as an exponential of a Lévy process $(L_{t\in[0,T]})$:

$$S_t=S_0\exp(L_t)\,,$$

where the choice of the Lévy process is the jump diffusion process plus the drift $r-\varpi_{MJD,\mathbb{Q}}$:

$$L_t\equiv(r-\varpi_{MJD,\mathbb{Q}})t+MJD(x_t;\sigma_{\mathbb{Q}},\lambda_{\mathbb{Q}},\mu_{\mathbb{Q}},\delta_{\mathbb{Q}}) \quad (3\text{-}54),$$

where $r \in \mathbb{R}^+$ is the instantaneous risk-free interest rate and all parameters are under the risk neutral probability measure $\mathbb{Q}$. The term $\varpi_{MJD,\mathbb{Q}}$ is the convexity correction which takes the following form in the MJD model:

$$\varpi_{MJD,\mathbb{Q}} = \lambda_{\mathbb{Q}} \left\{ \exp\left(\mu_{\mathbb{Q}} + \frac{\delta_{\mathbb{Q}}^{\ 2}}{2} \right) - 1 \right\} + \frac{\sigma_{\mathbb{Q}}^{\ 2}}{2} \qquad (3\text{-}55).$$

Defining the log return (i.e. log price relative) of the asset price as $R_t \equiv \ln(S_t / S_0)$ and using the equation (3-54):

$$x_t = R_t - (r - \varpi_{MJD,\mathbb{Q}})t \qquad (3\text{-}56).$$

Since obviously the drift $r - \varpi_{MJD,\mathbb{Q}}$ is deterministic, the probability density of the log return in the MJD model under the risk neutral probability measure $\mathbb{Q}$ can be expressed using the probability density (3-52) and the relationship (3-56) as:

$$\mathbb{Q}_{MJD}(R_t; \sigma, \lambda, \mu, \delta)$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \frac{1}{\sqrt{2\pi\left(\sigma^2 t + j\delta^2\right)}} \exp\left\{ -\frac{\left(x_t - j\mu\right)^2}{2\left(\sigma^2 t + j\delta^2\right)} \right\} \qquad (3\text{-}57),$$

with the equation (3-56). Note that all parameters in the density (3-57) are under the risk neutral probability measure $\mathbb{Q}$.

Table 3.2 summarizes all the models described thus far.

3.4 Calibration Methodology

3.4.1 Data Description

Our data consist of daily settlement call option prices on S&P 500 futures with March 2005 maturity obtained from Chicago Mercantile Exchange (CME) Daily Bulletin for the sample period from March 24, 2004, through March 16, 2005[14] for the total of 248 trading days. These American style option prices are converted to European style option prices using Barone-Adesi and Whaley (1987) quadratic approximation method to adjust for the early exercise premium. After eliminating call prices less than 0.125 due to reliability issues, the data used consist of a total of 6567 call prices.

Daily series of three month Treasury Bill rate are used as appropriate risk-free interest rates.

3.4.2 Introduction to Calibration

Two problems are said to be inverse to each other, if formulating one problem involves the other problem. One of these two problems (the simpler one or the older one) is called a direct problem and the other is called an inverse problem. The inverse problem which we deal with is called a parameter identification problem in which we try to identify physical parameters from observations of the evolution of the system. Let η be a vector of parameters to be identified. In our context, a direct problem is formulated as:

[14] March 16, 2005 is one day before the last trading day.

$$C_i^{\text{model}}(\hat{\eta}, S_t, K_i, T) = e^{-r(T-t)} E^{\mathbb{Q}}\left[\left(S_T - K_i\right)^+ \middle| \mathcal{F}_t\right] \quad \text{(3-58)},$$

where European option prices C_i^{model} across different strikes $K_{i \in I}$ are calculated given a vector of parameters $\hat{\eta}$ and variables such as strikes K_i and a maturity T. An inverse problem is formulated as the reverse of this procedure: Identify a vector of parameters η under measure $\mathbb{Q}$ [15] such that the discounted asset price process $(e^{-rt} S_{t \in 0 \le t \le T})$ becomes a martingale and which reproduce model option prices consistent with market observed option prices:

$$C_i^{\text{model}}(S_t, K_i, T) = C_i^{\text{market}} \quad \text{(3-59)}.$$

Because market option prices $(C_{i \in I}^{\text{market}})$ include noise (i.e. bid-ask spreads), we need not to find a risk-neutral martingale measure $\mathbb{Q} \sim \mathbb{P}$ which exactly reproduces market observed prices. Thus, the practical solution for the calibration problem becomes a best approximation problem between market observed prices and model calibrated prices. In most literatures, this approximation is done in a least squares sense. Thus, a calibration problem can be expressed as a nonlinear least squares problem of the form:

$$\eta^{\mathbb{Q}} = \arg\min_{\mathbb{Q} \in \wp} \sum_{i=1}^{N} \left| C_i^{\text{model}}(S_t, K_i, T) - C_i^{\text{market}} \right|^2 \quad \text{(3-60)},$$

where the implied risk neutral parameter vector $\eta^{\mathbb{Q}}$ is chosen by minimizing the sum of squared dollar pricing errors between market observed prices and model calibrated prices. Note that minimizing the sum of squared dollar pricing errors has the bias toward more expensive in-the-money options over relatively cheaper out-

[15] Called an implied risk neutral measure.

of-the-money options. But this is the standard objective function in the literature such as Dumas, Fleming and Whaley (1995), Bates (1996), Bakshi, Cao and Chen (1997), and Cont and Tankov (2004). Bakshi, Cao and Chen (1997) propose an alternative objective function which is to minimize the sum of squared percentage pricing errors:

$$\eta^{\mathbb{Q}} = \arg\min_{\mathbb{Q}\in\wp} \sum_{i=1}^{N} \left| \frac{C_i^{\text{model}}(S_t, K_i, T) - C_i^{\text{market}}}{C_i^{\text{market}}} \right|^2 ,$$

which has the bias toward cheaper out-of-the-money options.

3.4.3 Pricing Methodology for Lévy Models

We employ Carr and Madan's (1999) Fourier transform option pricing approach with the modified call price. For more details, consult the chapter 8 of Matsuda (2004b). The use of the Fourier transform option pricing is necessary for some Lévy models because their probability densities cannot always be expressed using special functions of mathematics or they are unknown. Therefore, Carr and Madan rewrite the option pricing function involving the probability density into the option pricing function involving the characteristic function because the characteristic function of the Lévy process is always available in a closed form.[16]

[16] There is one-to-one relationship between a probability density and a characteristic functions (i.e. through Fourier transform) and both of which uniquely determine a probability distribution.

3.5 Calibration, Implied Dynamics, and In-Sample Performance of Lévy Models

3.5.1 Dynamics of Calibrated Parameters

Using the data and procedure described in the section 3.4, each model is separately calibrated for each day. Table 3.3 reports the daily average, the standard error, the minimum, and the maximum of the calibrated parameters of each model for the sample period. Time series of calibrated parameters for each model are plotted from the day 1 which is March 24, 2004, through day 248 which is March 16, 2005 in Figure 3.6 through 3.10. Note that we set the NIG model's μ equal to zero because this parameter is a location parameter which only affects the mean of the equation (3-25) and can be incorporated into the parameters $\alpha, \beta,$ and δ. We observe the followings.

As expected and illustrated by Figure 3.6, the BS model's volatility parameter σ declines as the maturity of the S&P 500 futures option nears because the uncertainty regarding the terminal price declines.

The daily average of MJD model's σ is 0.09544 and smaller in size compared to the BS σ which is 0.14437 since the MJD model attributes the volatility to other parameters λ, μ, and δ. The daily average of the intensity parameter λ of the MJD model is 0.77742 which means that the jump occurs 0.77742 times per year on average for this sample. It is very interesting to note that λ tends to increase (i.e. higher frequency of jumps) as the maturity nears as illustrated by Figure 3.7 and this tendency seems to be true regardless of the sample. Bakshi, Cao, and Chen (1997)

reports the same observation. The MJD model's mean log return jump size μ is -14.899% on average and it declines as the maturity nears for the sample. As previously mentioned, the negative value of μ indicates the natively skewed log return probability density function. The MJD model's standard deviation parameter δ of mean log return jump size is 0.09411 on average and it declines as the maturity nears for the sample indicating the less uncertainty regarding the jump size.

The VG model's drift parameter θ is -0.16798 on average and it indicates the natively skewed log return probability density function. Figure 3.8 depicts the declining trend for the volatility parameter of the subordinated Brownian motion σ and the variance rate parameter ν. This contributes to the declining negative skewness and excess kurtosis of the log return probability density function for the sample period.

The NIG model's skewness parameter α is -11.7309 on average and it indicates the natively skewed log return probability density function. Figure 3.9 shows the increasing trend for the shape parameter α (which leads to smaller volatility) and the excess kurtosis parameter δ (which leads to higher excess kurtosis). But later it will be shown that the increase in δ is dominated by the increase in α and the increase in size in β and as a result, the excess kurtosis of the log return probability density function declines as the maturity approaches for the sample.

The CGMY model's parameter C is 0.06675 on average and it declines dramatically as the maturity nears indicating the smaller overall arrival rate of jumps at the near maturity. As expected, the lower tail decay rate parameter G is 2.35246

on average and much smaller than the upper tail decay rate parameter M which is 262.805. Less decayed lower tail means the negative skewness of the log return probability density function. We also observe that the estimates of G and M become extremely volatile as the maturity nears in contrast to that of C. The parameter Y estimated is 1.19952 for the sample which confirms the monotonically decreasing Lévy measure, the infinite total mass of the Lévy measure (i.e. the infinite arrival rate of jumps), and the fact that the underlying S&P 500 futures price process is a process of infinite variation according to Table 3.1. Closer examination reveals that the calibrated Y is above one (the process is infinite variation) for 71% (175/248) of the days. But Figure 3.10 shows the definite pattern of the estimated Y being less than one for the long maturity options and being above one for the short maturity options. For this reason, we divide the entire sample period into three subperiods. The first subperiod is the long term to maturity which include days with greater than or equal to 180 days to maturity (day 1 through day 69), the second subperiod is the medium term to maturity which include days with greater than or equal to 60 but less than 180 days to maturity (day 70 through day 190), and the third subperiod is the short term to maturity which include days with less than 60 days to maturity (day 191 through day 248). Interestingly, for the long term to maturity, the estimated Y is less than one (the underlying process is finite variation) for 68% of the period with the sub-mean 0.941. For the medium term to maturity, the estimated Y is above one for 78.5% of the period with the sub-mean 1.14. And for the short term to maturity, the estimated Y is above one for 100% of the period with the sub-mean 1.63. Thus, the underlying S&P 500 futures price shifted from the finite

variation stochastic process to infinite variation stochastic process as the maturity moves from the long to the short term at least for the sample used here.

3.5.2 Implied Dynamics of Log Return Probability Density Functions

Figure 3.11 through Figure 3.14 reports the calibrated probability density functions of log returns on one day before last trading day March 16, 2005 as the time to maturity nears from 1 to 0 for each model.[17] For the purpose of clear illustration, Figure 3.15 reports the snapshots of Figure 3.11 through Figure 3.14 on six different maturity dates and Figure 3.16 reports the time series of the implied moments of log return probability density function for each model. It is apparent that the extra parameters of various non-Gaussian Lévy models allow the negative skewness (i.e. fatter lower tail and thinner upper tail) and the excess kurtosis (i.e. higher peaked and heavier tailed) of the log return density over the BS model which is the only Gaussian Lévy models. Notice from Figure 3.16 that all non-Gaussian Lévy models have the similar mean and variance, but the CGMY model has the higher negative skewness and the higher excess kurtosis especially for the short term to maturity.

3.5.3 Dynamics of Implied Lévy Density Functions of Log Returns

Figure 3.17 through 3.20 reports the calibrated Lévy density functions of log returns as the time to maturity nears from 1 to 0 for each model. For the purpose of clear illustration, Figure 3.21 reports the snapshots of Figure 3.17 through 3.20 on six different maturity dates.

[17] Remember that the probability density function of the CGMY process is not available in closed form in the general case.

As previously mentioned, we observe the pattern of increasing total mass of the MJD model's Lévy density which is equal to the parameter λ as the maturity nears since it tends to increase. Also notice that the arrival rate of negative jumps is far larger than that of positive jumps.

In contrast to the MJD model's finite activity and symmetric Lévy density, all pure jump Lévy models are characterized by the asymmetric and infinite activity Lévy densities which all look similar except the CGMY model's Lévy density having much thinner upper tail especially in the Panel B) of Figure 3.21 where the arrival rate of positive jumps are actually zero. As expected, all infinite activity Lévy densities have heavier lower tails indicating that the negative jumps are more likely to arrive than the positive jumps.

3.5.4 In-Sample Performance of Lévy Models

In this section, in-sample fit of Lévy option pricing models is discussed although it bears little importance because in general more parameters lead to better in-sample fit. Table 3.4 reports the daily average, the standard error, the minimum, and the maximum of the sum of squared errors (SSE) for the sample period. Figure 3.22 through Figure 3.26 reports the in-sample dollar pricing error defined as $C_i^{\text{model}}(S_0, K_i, T; \eta^{\mathbb{Q}}{}_0) - C_i^{\text{market}}$ across the varying moneyness and the time to maturity using a linear interpolation.

As expected, the BS model is by far the least fit model averaging the SSE of 327.314 because it has only one parameter σ. Figure 3.22 illustrates the well-known bias of the BS model which is to underprice the ITM (in-the-money, the moneyness

K / F_0 less than one in the figure) calls and to overprice the OTM (out-of-the-money, the moneyness K / F_0 larger than one in the figure) calls. Or, through the put-call parity, the BS model underprices the OTM (the moneyness K / F_0 less than) puts and overprices the ITM (the moneyness K / F_0 larger than one in the figure) puts. This bias of the BS model becomes stronger especially for the short term OTM calls and puts in terms of the percentage mispricing errors because of the cheaper prices of the short term OTM calls and puts.

We see from Table 3.4 and Figure 3.23 through Figure 3.26 that incorporating jumps in the underlying price process dramatically improves the in sample fit, but this is no surprise because of more parameters. The CGMY model has the best in-sample fit, the NIG model is the second, the MJD model is the third, and the VG model is the fourth. Figure 3.23 through Figure 3.26 shows no sign of systematic mispricing bias across the moneyness and the time to maturity of each of the jump Levy models.[18]

3.6 Out-of-Sample Pricing Performance of Lévy Models

We saw in the previous section that incorporating jumps in the underlying price process dramatically improves the in sample fit with the CGMY model being the best, the NIG model being the second, the MJD model being the third, and the VG model being the fourth. In this section, the more important out-of-sample fit of

[18] Note that the increasing underpricing for the short term OTM calls in Figure 3.23 through 3.26 is due to the linear extrapolation used and should be ignored. Short term deep OTM calls have prices of almost zero and go in the cabinet.

the each Lévy model is examined following Bakshi, Cao, and Chen (1997) because of the fact that the better in-sample fit might be due to more parameters of the jump Lévy models than the BS model. In other words, if the extra parameters of the jump Lévy models are redundant, they overfit the data and may produce larger out-of-sample fitting errors.

The procedure is as follows. Let t denote today. Firstly, on yesterday ($t-1$), the parameters of each Lévy model are recovered using the yesterday's call prices and variables. Secondly, today's model prices are computed using the yesterday's calibrated parameters $\eta^{\mathbb{Q}}_{t-1}$ and today's variables. Thirdly, using today's model prices and today's call prices, the absolute pricing error is obtained following the equation:

$$\text{absolute pricing error} = \left| C_t^{\text{model}}(F_t, K, T-t; \eta^{\mathbb{Q}}_{t-1}) - C_t^{\text{market}} \right| \qquad (3\text{-}61).$$

We repeat this procedure everyday across all the call strike prices. Table 3.5 reports the out-of-sample absolute pricing errors per call option for total of 15 different subsamples. Note that the entire sample is divided according to its maturity and its moneyness. The long term is defined as more than or equal to 180 days to maturity and the short term is less than 60 days to maturity. We define the deep ITM as the moneyness less than 0.94, the moneyness between 0.94 and 0.98 is ITM, the moneyness between 0.98 and 1.02 is ATM, the moneyness between 1.02 and 1.06 is OTM, and the moneyness greater than 1.06 is deep OTM. Note that we define the moneyness as K / F_0 which is an opposite of the definition of the moneyness used by Bakshi, Cao, and Chen (1997).

We firstly observe from Table 3.5 that in terms of the absolute pricing error per call, the CGMY model is the best with having the least error in 11 subsamples out of 15, the NIG model is the second with having the least error in 4 subsamples and having the second least error in 5 subsamples, the MJD model is the third with having the least error in 1 subsamples and having the second least error in 7 subsamples, and the VG model is the fourth although it is a significant improvement over the BS model. For example, for the out-of-sample pricing of the medium term ATM calls, the CGMY model has the least absolute dollar pricing error per call with \$0.598, the MJD model is the second best with \$0.602, the NIG model is the third with \$0.612, the VG model is the fourth with \$0.685, and the BS model has the largest error of \$1.365. It is interesting to find that the MJD model which is a finite activity Lévy model containing a Brownian motion outperforms the VG model which is an infinite activity Lévy model of finite variation with pure jump sample paths. This indicates that the MJD model is still an excellent model for the purpose of index option pricing.

In order to closer examine the out-of-sample pricing bias associated with the Lévy models, out-of-sample dollar pricing error (not the absolute value, therefore it can show underpricing or overpricing) defined as $C_t^{\text{model}}(F_t, K, T-t; \eta^{\mathbb{Q}}_{t-1}) - C_t^{\text{market}}$ is calculated on each day across different strike prices and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. As expected and previously mentioned, Figure 27 shows that the bias of the BS model which is to underprice the ITM calls and to overprice the OTM calls. Interestingly, the dynamics of out-of-sample dollar pricing error of all jump Lévy

models resemble one another as illustrated by Figure 3.28 through Figure 3.31 which differ significantly from the BS counterpart. We find that jump Lévy models either underprices or overprices all calls on a given day and the size of the dollar pricing error is largest near ATM calls because of the bell-shaped or reverse bell-shaped pattern. Thus, the jump Lévy models still possess the pricing bias associated with the moneyness and the maturity although it is much smaller than the BS model.

Out-of-sample pricing performance of each Lévy model can also be evaluated by the orthogonality test which is based on the idea that a good model should not possess any consistent pattern in pricing errors and the pricing errors of a good model cannot be predicted. Our orthogonality test is conducted by regressing the out-of-sample dollar pricing errors ε_i on the moneyness K_i / F_0, the moneyness squared $(K_i / F_0)^2$, the time to maturity T_i, and the interest rate r_i:

$$\varepsilon_i = \beta_0 + \beta_1 (K_i / F_0) + \beta_2 (K_i / F_0)^2 + \beta_3 T_i + \beta_4 r_i + e_i \qquad (3\text{-}62).$$

Note that Carr, Chang, and Madan (1998) employ this type of orthogonality test for the in-sample pricing performance of the VG model, ours is for the out-of-sample pricing performance. The results of the orthogonality test are shown in Table 3.6. We firstly notice that the orthogonality of the out-of-sample pricing errors for all Lévy models are rejected because of the high values for the $F-$statistic which are in row 8. This means that the out-of-sample pricing errors for all Lévy models can be explained by the moneyness, the moneyness squared, the time to maturity, and the interest rate which all have coefficients significantly different from zero shown in row 1 through row 5. These biases of all Lévy models can be intuitively

seen from Figure 15. But, in terms of the degree of bias, the BS model is the most biased. In other words, the pricing errors of the BS model are the most predictable with the adjusted R^2 of 47%. The least biased model is the CGMY with the adjusted R^2 of 2.4%. The MJD and the NIG models are the second best with the adjusted R^2 of 6.9% which outperform the VG model with the adjusted R^2 of 11.1%.

3.7 Conclusion

We calibrated a total of five different Lévy models to the S&P 500 futures options. The classic Black-Scholes model is the only continuous (i.e. Gaussian) Lévy model in which the asset price dynamics is modeled by the Brownian motion. Another classic Merton jump diffusion model is the non-Gaussian Lévy model in which the asset price dynamics is modeled by the jump diffusion process which possesses the discontinuities although they are rare. And, relatively recently developed three pure jump Lévy models in which the asset price dynamics is modeled using pure jump Lévy process. They are the variance gamma model, the normal inverse Gaussian model, and the CGMY model. Carr and Madan's (1999) Fourier transform option pricing approach with the modified call price was employed for the pricing.

The calibration result suggests that the extra parameters of various jump Lévy models allow the negative skewness and the excess kurtosis of the log return density over the BS model. On the dynamics of implied Lévy density functions of log

returns, we observe the pattern of increasing total mass of the MJD model's symmetric Lévy density. And, the Lévy densities of all pure jump Lévy models are characterized by the asymmetric and infinite activity Lévy densities which all resemble one another. In terms of the out-of-sample pricing performance, containing jumps in the underlying price process drastically improves the performance over the BS model. Which jump Lévy models perform the best in pricing? The CGMY model is the least biased model in terms of overall pricing. Another result is that the classic MJD model which contains a Brownian motion performs as good as the NIG model and consistently outperforms the VG models for the pricing of index options. As Bakshi, Cao, and Chen (1998) point out that all option pricing models are misspecified (biased) because the correctly specified (unbiased) model would produce errors which are on average equal to zero. In our context, all the Lévy models are misspecified with regard to the moneyness and the maturity, most notably. But, we find that adding jumps in the underlying price process drastically reduces the degree of misspecification of the model.

We would like to suggest a couple of topics for the future research. One is that this paper can be reapplied to the other data such as individual stock options, commodity options (such as oil), and the currency options. It is very likely that the empirical result changes.

Chapter 4: Parametric Regularized Calibration of Merton Jump-Diffusion Model with Relative Entropy

4.1 Introduction

A calibration problem is an inverse problem which tries to identify (i.e. back out) a vector of parameters $\theta^{\mathbb{Q}}$ which produce model option prices consistent with market (i.e. observed) option prices. When the objective function which is usually the sum of squared dollar pricing errors between market prices and model prices is not convex, the calibration problem is an illposed problem. In this illposed calibration problem, some solution can always be found (i.e. non-uniqueness), but the solution obtained is very sensitive to the initial values (i.e. instability) and it is not the global minimum with very high likelihood. Traditionally in the field of finance, the gradient descent algorithms such as the BFGS method are generally used to solve this illposed problem. To raise a few examples, Bates (1996), Bakshi, Cao, Chen (1997), Dumas, Fleming, and Whaley (1998), Carr, Chang, and Madan (1998b), and Carr, Geman, Madan, and Yor (2002).

To overcome above mentioned difficulties in illposed problems, statisticians and mathematicians have long been using regularization methods. Regularization methods are not the methods to locate the global solution, but they are the methods to enhance the uniqueness and the stability of the calibration solution by sacrificing its precision. Cont and Tankov (2004a, b) choose the regularization with the relative entropy $\mathcal{E}(\mathbb{Q}|\mathbb{P})$ which is a measure of distance between two probability measures and describes the amount of inefficiency to assume that the true distribution is $\mathbb{Q}$ when the true distribution of the random variable X is $\mathbb{P}$. It is very convenient that the relative entropy for Lévy processes can be explicitly expressed in terms of their

Lévy measures. It is important to realize that using the relative entropy and using the prior mean the introduction of the bias of the calibration solution toward the prior to gain the numerical stability and the uniqueness by making the objective function more convex. It implies that the user has some or strong belief in the use of the prior (i.e. otherwise, why do you bother?). Cont and Tankov develop the non-parametric regularized calibration method for Merton (1976) jump-diffusion model with the relative entropy.

The goal of this paper is to give an answer to the following simple question, "What difference does the regularization with relative entropy make?" For this purpose, we calibrate the Merton (1976) jump-diffusion model which is a Lévy model with occasional but rare jumps to the S&P 500 futures options with or without regularization. The impact of the regularization over the unregularized calibration is judged by the calibrated log return probability density and the calibrated Lévy measure.

Note that we apply Cont and Tankov's (2004 a, b) regularization method with the relative entropy to the index options *parametrically* with the MJD model. This is different from Cont and Tankov's (2004 a, b) research which is to calibrate Merton jump-diffusion model non-parametrically with the relative entropy regularization.

The calibration result suggests that with or without regularization, calibrated risk-neutral parameters, calibrated log return probability densities, and calibrated Lévy measures are not significantly different. It seems that Lévy measures are more sensitive to the regularization than log return probability measures. Notice also that

the difference in calibrated parameters between the regularized and the unregularized become more pronounced especially for near maturity options.

This chapter is organized as follows. Section 4.2 gives the detailed description of the Merton jump-diffusion model. Section 4.3 presents the (unregularized) calibration problem as an inverse problem and as an illposed problem due to the non-convexity of the objective function. Section 4.4 briefly reviews Cont and Tankov's (2004a, b) method of the regularized calibration with the relative entropy which tries to achieve a unique solution and a stable solution. Section 4.5 describes the S&P 500 futures option data set and obtains two different prior probability measures. One is the statistical prior and the other is the risk-neutral prior. Section 4.6 provides our main empirical result of the difference between the regularized calibration and the unregularized calibration in terms of the calibrated parameter vector, the calibrated log return probability density, and the calibrated Lévy measure. Section 4.7 concludes.

4.2 Merton Jump Diffusion (MJD) Model (1976)

Consider a fixed filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$. A jump diffusion process $(X_{t\in[0,\infty]})$ with the Lévy triplet $(A_X = \sigma^2, \ell_X = \lambda f(x), \gamma_X = 0)$ is defined as a Brownian motion plus a compound Poisson process:

$$(X_{t\in[0,\infty]}) \equiv (\sigma B_{t\in[0,\infty]}) + \sum_{i=1}^{N_t} X_i \qquad (4\text{-}1),$$

where $(\sigma B_{t\in[0,\infty]})$ is a multiplicative Brownian motion with the Lévy triplet $(A_B = \sigma^2, \ell_B = 0, \gamma_B = 0)$, $\sum_{i=1}^{N_t} X_i$ is a compound Poisson process with the Lévy triplet $(A_C = 0, \ell_C = \lambda f(x), \gamma_C = 0)$ which is the sum of *i.i.d.* jumps X_i from the jump size probability density $f(x)$, and $(N_{t\in[0,\infty]})$ is a Poisson process with the intensity $\lambda \in \mathbb{R}^+$ which counts the number of random arrival times T_k of an event in the time interval $[0,t]$:

$$N_t = \sum_{k\geq 1} 1_{t\geq T_k} \qquad (4\text{-}2).$$

Note that a Poisson process $(N_{t\in[0,\infty]})$ and the jumps sizes $(X_i)_{i\geq 1}$ are assumed to be independent. This jump diffusion process $(X_{t\in[0,\infty]})$ possesses the following properties. It is a jump Lévy process, but not a pure jump Lévy process because the Gaussian variance term of the jump diffusion process A_X is non-zero. In other words, the process contains a Brownian motion. The Lévy measure of the jump diffusion process is given by:

$$\ell_X(x;\lambda) = \lambda f(x) \qquad (4\text{-}3),$$

where $f(x)$ is the jump size probability density. The total mass of the Lévy measure of the jump diffusion process is the intensity parameter λ because a Lévy measure $\ell(x)$ measures the arrival rate of jumps:

$$\int_{-\infty}^{\infty} \ell_X(x)dx = \int_{-\infty}^{\infty} \lambda f(x)dx = \lambda \int_{-\infty}^{\infty} f(x)dx = \lambda < \infty,$$

which is finite because the number of arrivals of an event is almost surely finite for any $t > 0$ including an infinite time horizon $t = \infty$. In other words, the jump

diffusion process is a finite activity Lévy process which means that the process has finite number of small jumps and finite number of large jumps. The jump diffusion process is also a Lévy process of infinite variation in the interval $[0,\infty)$ because $A_X = \sigma^2 \neq 0$.

Merton jump diffusion (MJD) model specifies the log return jump size density as the normal, i.e. $X_i \sim i.i.d.\ Normal(\mu, \delta^2)$:

$$f_{MJD}(x) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\delta^2} \right\} \quad (4\text{-}4).$$

Thus, the Lévy measure in MJD model can be expressed as:

$$\ell_{MJD,X}(x;\lambda,\mu,\delta) = \frac{\lambda}{\sqrt{2\pi\delta^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\delta^2} \right\} \quad (4\text{-}5).$$

The characteristic function of MJD process can be obtained by the use of the Lévy-Khinchin representation as:

$$\phi_{MJD,X}(\omega;\sigma,\lambda,\mu,\delta) = \exp\left[t\left\{ -\frac{\sigma^2\omega^2}{2} + \lambda\left(\phi_f(\omega) - 1\right) \right\} \right] \quad (4\text{-}6),$$

where ϕ_f is the characteristic function of the jump size density:

$$\phi_f(\omega) = \exp\left(i\omega\mu - \frac{\delta^2\omega^2}{2} \right).$$

The probability density of the MJD process can be computed using the conditionally normal property of the jump diffusion process of the equation (4.1):

$$\mathbb{P}_{MJD}(x_t;\sigma,\lambda,\mu,\delta) = \sum_{j=0}^{\infty} \mathbb{P}(x_t \,|\, N_t = j)\mathbb{P}_{Poisson}(N_t = j)$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \frac{1}{\sqrt{2\pi\left(\sigma^2 t + j\delta^2\right)}} \exp\left\{-\frac{\left(x_t - j\mu\right)^2}{2\left(\sigma^2 t + j\delta^2\right)}\right\} \quad (4\text{-}7).$$

Its standardized moments are computed by:

$$E[X_t] = \lambda t \mu \quad (4\text{-}8),$$

$$Variance[X_t] = (\sigma^2 + \lambda\delta^2 + \lambda\mu^2)t\,,$$

$$Skewness[X_t] = \frac{t\lambda\mu(\mu^2 + 3\delta^2)}{Variance[X_t]^{3/2}}\,,$$

$$Excess\ Kurtosis[X_t] = \frac{t\lambda(\mu^4 + 3\delta^4 + 6\mu^2\delta^2)}{Variance[X_t]^2}\,.$$

These standardized moments indicate that μ is a skewness parameter with $\mu = 0$ producing the symmetric probability density. Larger values for λ and σ lead to the larger variance and smaller excess kurtosis of the probability density.

MJD model specifies the asset price dynamics $(S_{t\in[0,T]})$ defined on a filtered risk neutral probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{Q})$ as an exponential of a Lévy process $(L_{t\in[0,T]})$:

$$S_t = S_0 \exp(L_t)\,,$$

where the choice of the Lévy process is the jump diffusion process plus the drift $r - \varpi_{MJD,\mathbb{Q}}$:

$$L_t \equiv (r - \varpi_{MJD,\mathbb{Q}})t + MJD(x_t; \sigma_{\mathbb{Q}}, \lambda_{\mathbb{Q}}, \mu_{\mathbb{Q}}, \delta_{\mathbb{Q}}) \quad (4\text{-}9),$$

where $r \in \mathbb{R}^+$ is the instantaneous risk-free interest rate and all parameters are under the risk neutral probability measure $\mathbb{Q}$. The term $\varpi_{MJD,\mathbb{Q}}$ is the convexity correction which takes the following form in the MJD model:

$$\varpi_{MJD,\mathbb{Q}} = \lambda_{\mathbb{Q}} \left\{ \exp\left(\mu_{\mathbb{Q}} + \frac{\delta_{\mathbb{Q}}^{\ 2}}{2} \right) - 1 \right\} + \frac{\sigma_{\mathbb{Q}}^{\ 2}}{2} \qquad (4\text{-}10).$$

Defining the log return (i.e. log price relative) of the asset price as $R_t \equiv \ln(S_t / S_0)$ and using the equation (4-9):

$$x_t = R_t - (r - \varpi_{MJD,\mathbb{Q}})t \qquad (4\text{-}11).$$

Since obviously the drift $r - \varpi_{MJD,\mathbb{Q}}$ is deterministic, the probability density of the log return in the MJD model under the risk neutral probability measure $\mathbb{Q}$ can be expressed using the probability density (4-7) and the relationship (4-11) as:

$$\mathbb{Q}_{MJD}(R_t; \sigma, \lambda, \mu, \delta)$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \frac{1}{\sqrt{2\pi\left(\sigma^2 t + j\delta^2\right)}} \exp\left\{ -\frac{\left(x_t - j\mu\right)^2}{2\left(\sigma^2 t + j\delta^2\right)} \right\} \qquad (4\text{-}12),$$

with the equation (4-11). Note that all parameters in the density (4-12) are under the risk neutral probability measure $\mathbb{Q}$.

4.3 Calibration without Regularization

4.3.1 Calibration: An Inverse Problem

Calibration is an inverse problem. Consider two problems. One of which is named as a direct problem and the other is named as an inverse problem. In the context of option pricing, a direct problem is formulated as:

$$C_i^{\text{model}}(\theta^{\mathbb{Q}}, S_t, K_i, T-t) = e^{-r(T-t)} E^{\mathbb{Q}}\left[\left(S_T - K_i\right)^+ \middle| \mathcal{F}_t\right] \quad \text{(4-13)},$$

where European call option prices C_i^{model} across different strikes $K_{i\in I}$ are calculated given a model, a vector of risk-neutral model parameters $\theta^{\mathbb{Q}}$, and variables such as a spot asset price S_t and a maturity $T-t$. Following the martingale asset pricing, a European call price is equal to the discounted value of the terminal payoff under a risk-neutral probability measure $\mathbb{Q}$. For example, $\theta^{\mathbb{Q}} = (\sigma^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \mu^{\mathbb{Q}}, \delta^{\mathbb{Q}})$ in the MJD case. An inverse problem (called a parameter identification problem) is formulated as the reverse of this procedure. It is to identify (i.e. back out) a vector of parameters $\theta^{\mathbb{Q}}$ which produce model option prices consistent with market (i.e. observed) option prices:

$$C_i^{\text{model}}(\theta^{\mathbb{Q}}) = C_i^{\text{market}} \quad \text{(4-14)}.$$

The exact match of the model prices to the market prices is not necessary because of the noise (i.e. bid-ask spreads) contained in the market option prices $(C_{i\in I}^{\text{market}})$. Thus, the calibration problem becomes a best approximation problem between market prices and model prices which is done using a nonlinear least squares (NLS):

$$\theta^{\mathbb{Q}} = \arg\min_{\theta^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 \qquad (4\text{-}15),$$

where the risk-neutral parameter vector $\theta^{\mathbb{Q}}$ is chosen by minimizing the sum of squared dollar pricing errors between market prices and model prices. The gradient descent algorithms such as the BFGS method are generally used to solve the optimization problem of the equation (4-15).

Followings are examples of literatures which use this NLS without regularization. Bates (1996) calibrates stochastic volatility/jump models to a currency option, Bakshi, Cao, Chen (1997) calibrates stochastic volatility/jump models to an index option,and Dumas, Fleming, and Whaley (1998) uses unregularized calibration for fitting a deterministic volatility function. Carr, Chang, and Madan (1998b) calibrate the Variance Gamma model to an index option through the maximum likelihood estimation which is equivalent to the NLS. Carr, Geman, Madan, and Yor (2002) calibrates the extended CGMY model to individual stock options and index options.

4.3.2 Calibration: An Illposed Problem

According to Cont and Tankov (2004a, b), an illposed problem is a problem which possesses the following properties. Firstly, an illposed problem may not have a solution or may have an infinite number of solutions. Secondly, when a solution or solutions of an illposed problem exists (exist) and if some type of an additional criterion is used to choose a solution, it is very sensitive to the initial values. Thirdly, when a solution of an illposed problem exists, it might be difficult to obtain it

because it is likely to get stuck at local minima (due to the non-convex objective function).

Our interest of the calibration problem of the equation (4-15) is an illposed problem whose illposedness is solely caused by the non-convex objective function. One example of the sum of squared dollar pricing error function for the MJD model is illustrated in Figure 4.1 as a function of parameters $\lambda^{\mathbb{Q}}$ and $\mu^{\mathbb{Q}}$ with other parameters being fixed. Due to its non-convex nature, the optimization problem (4-15) possesses the following illposedness. Firstly, some solution can always be found (this is not necessarily a good thing). But, secondly, the solution obtained is very sensitive to the initial values. In other words, the solution is very instable. Thirdly, it is highly likely that the solution obtained is not the global minimum.

Note that the calibration problem of the equation (4-15) is not the only illposed problem. Illposed problems are everywhere which include the numerous maximum likelihood estimation problems. In the past, researchers have used an ad hoc treatment for illposed problems such as repeating the optimization procedure with various initial values.

4.4 Regularized Calibration

4.4.1 Relative Entropy

We saw in the previous section that the NLS calibration problem has the difficulty in achieving a unique solution and a stable solution. To overcome this issue, regularization methods have been developed. Engl, Hanke, and Neubauer

(1996) give a brief summary of regularization methods. In this paper, we focus on the regularization with relative entropy which is used by Cont and Tankov (2004a, b). Note that Cont and Tankov are the first to use relative entropy regularization for the calibration of the exponential Lévy models, but they are not the first in using relative entropy regularization in the finance context. We believe it was Avellaneda, Friedman, Holmes, and Samperi (1997) who used the relative entropy regularization for the calibration of volatility surfaces.

The relative entropy is a measure of distance between two probability measures which is expressed as the expected value of the logarithm of the likelihood ratio. If we knew that the true distribution of the random variable X is $\mathbb{P}$, the relative entropy $\mathcal{E}(\mathbb{Q}|\mathbb{P})$ describes the amount of inefficiency to assume that the true distribution is $\mathbb{Q}$.

DEFINITION Let $(X_{T\in[0,\infty]})$ be a real-valued rcll process defined on a filtered probability space $(\Omega, \mathcal{F}_{T\in[0,\infty]}, \mathbb{P})$. Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probability measures on $(\Omega, \mathcal{F}_{T\in[0,\infty]})$. The relative entropy or Kullback-Leibler distance between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ is defined as:

$$\mathcal{E}(\mathbb{Q}|\mathbb{P}) = E^{\mathbb{Q}}[\ln\frac{d\mathbb{Q}}{d\mathbb{P}}] = E^{\mathbb{P}}[\frac{d\mathbb{Q}}{d\mathbb{P}}\ln\frac{d\mathbb{Q}}{d\mathbb{P}}] \quad (4\text{-}16).$$

Relative entropy is a convex function of $\mathbb{Q}$ because after a rearrangement:

$$\mathcal{E}(\mathbb{Q}|\mathbb{P}) = E^{\mathbb{P}}[f(\frac{d\mathbb{Q}}{d\mathbb{P}})] \quad (4\text{-}17),$$

where $f(x) = x \ln x$ is a strictly convex function as illustrated in Figure 4.2. This strict convexity of $\mathcal{E}(\mathbb{Q}|\mathbb{P})$ plays a crucial role in enhancing the uniqueness of a solution of an illposed problem. Relative entropy is always non-negative (i.e. $\mathcal{E}(\mathbb{Q}|\mathbb{P}) \geq 0$) and $\mathcal{E}(\mathbb{Q}|\mathbb{P}) = 0$ if and only if $d\mathbb{Q}/d\mathbb{P} = 1$ almost surely. Cover and Thomas (1991) provide the proof.

4.4.2 Relative Entropy for Lévy Processes

Very conveniently, relative entropy for Lévy processes can be explicitly expressed in terms of their Lévy measures. The following useful theorem corresponds to the proposition 2 of Cont and Tankov (2004b).

THEOREM: Relative Entropy for Lévy Processes

Let $(X_{T\in[0,\infty]})$ be a real-valued Lévy process defined on a filtered probability space $(\Omega, \mathcal{F}_{T\in[0,\infty]}, \mathbb{P})$ with the Lévy triplet $(A^{\mathbb{P}}, \ell^{\mathbb{P}}, \gamma^{\mathbb{P}})$. Define a probability measure $\mathbb{Q} \sim \mathbb{P}$ under which a Lévy process $(X_{T\in[0,\infty]})$ is described by the triplet $(A^{\mathbb{Q}}, \ell^{\mathbb{Q}}, \gamma^{\mathbb{Q}})$. Note that $A = A^{\mathbb{P}} = A^{\mathbb{Q}}$ because $\mathbb{Q} \sim \mathbb{P}$. Then, for every time horizon $T \in [0,\infty]$, the relative entropy of $\mathbb{Q}|\mathcal{F}_T$ with respect to $\mathbb{P}|\mathcal{F}_T$ can be expressed as:

$$\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = \frac{T}{2A}\left\{\gamma^{\mathbb{Q}} - \gamma^{\mathbb{P}} - \int_{-1}^{1} x(\ell^{\mathbb{Q}} - \ell^{\mathbb{P}})(dx)\right\}^2 1_{A\neq 0}$$

$$+T\int_{-\infty}^{\infty}\left(\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}\ln\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}} + 1 - \frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}\right)\ell^{\mathbb{P}}(dx) \qquad (4\text{-}18).$$

When $\mathbb{P}$ and $\mathbb{Q}$ are both the risk-neutral martingale measures, for every time horizon $T \in [0,\infty]$, the relative entropy of $\mathbb{Q}\big|\mathcal{F}_T$ with respect to $\mathbb{P}\big|\mathcal{F}_T$ can be reduced to:

$$\mathcal{E}_T(\mathbb{Q}\big|\mathbb{P}) = \frac{T}{2A}\left\{\int_{-\infty}^{\infty}(e^x-1)(\ell^{\mathbb{Q}}-\ell^{\mathbb{P}})(dx)\right\}^2 1_{A\neq 0}$$

$$+T\int_{-\infty}^{\infty}\left(\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}\ln\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}+1-\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}\right)\ell^{\mathbb{P}}(dx) \qquad (4\text{-}19).$$

Example of Relative Entropy: Brownian Motion with Drift Let $(X_{T\in[0,\infty]})$ be a Brownian motion with drift defined on a filtered probability space $(\Omega,\mathcal{F}_{T\in[0,\infty]},\mathbb{P})$ with the Lévy triplet $(A^{\mathbb{P}}=(\sigma^{\mathbb{P}})^2,\ell^{\mathbb{P}}=0,\gamma^{\mathbb{P}}=\mu^{\mathbb{P}})$:

$$X_T{}^{\mathbb{P}} = \mu^{\mathbb{P}}T+\sigma^{\mathbb{P}}B_T{}^{\mathbb{P}}.$$

Define a probability measure $\mathbb{Q}\sim\mathbb{P}$ under which a Brownian motion with drift $(X_{T\in[0,\infty]})$ is described by the triplet $(A^{\mathbb{Q}}=(\sigma^{\mathbb{Q}})^2,\ell^{\mathbb{Q}}=0,\gamma^{\mathbb{Q}}=\mu^{\mathbb{Q}})$. Note that $A^{\mathbb{P}}=A^{\mathbb{Q}}=\sigma^2$ because $\mathbb{Q}\sim\mathbb{P}$. Then, for every time horizon $T\in[0,\infty]$, the relative entropy of $\mathbb{Q}\big|\mathcal{F}_T$ with respect to $\mathbb{P}\big|\mathcal{F}_T$ can be expressed as by applying the equation (4-18):

$$\mathcal{E}_T(\mathbb{Q}\big|\mathbb{P}) = \frac{T}{2\sigma^2}\left\{\mu^{\mathbb{Q}}-\mu^{\mathbb{P}}\right\}^2 = \frac{1}{2}\left(\frac{\mu^{\mathbb{Q}}-\mu^{\mathbb{P}}}{\sigma}\right)^2 T \qquad (4\text{-}20).$$

Thus, in Gaussian case, the relative entropy function is symmetric:

$$\mathcal{E}_T(\mathbb{Q}\big|\mathbb{P}) = \mathcal{E}_T(\mathbb{P}\big|\mathbb{Q}),$$

because $(\mu^{\mathbb{Q}} - \mu^{\mathbb{P}})^2 = (\mu^{\mathbb{P}} - \mu^{\mathbb{Q}})^2$. The relative entropy function of the equation (4-20) is plotted in Figure 4.3 by setting $T = 1$ and $\sigma = 1$ for simplicity. Notice its strict convexity.

Example of Relative Entropy: Merton jump-diffusion process Let $(X_{T\in[0,\infty]})$ be a jump diffusion process of the equation (3-1) defined on a filtered probability space $(\Omega, \mathcal{F}_{T\in[0,\infty]}, \mathbb{P})$ with the Lévy triplet $(A^{\mathbb{P}} = (\sigma^{\mathbb{P}})^2, \ell^{\mathbb{P}} = \lambda^{\mathbb{P}} f(x)^{\mathbb{P}}, \gamma^{\mathbb{P}} = 0)$. Define a risk-neutral martingale probability measure $\mathbb{Q} \sim \mathbb{P}$ under which a jump diffusion process $(X_{T\in[0,\infty]})$ is described by the triplet $(A^{\mathbb{Q}} = (\sigma^{\mathbb{Q}})^2, \ell^{\mathbb{Q}} = \lambda^{\mathbb{Q}} f(x)^{\mathbb{Q}}, \gamma^{\mathbb{Q}} = 0)$. Note that $A^{\mathbb{P}} = A^{\mathbb{Q}} = \sigma^2$ because $\mathbb{Q} \sim \mathbb{P}$. From the equation (4-5), the Lévy measures take the form:

$$\ell^{\mathbb{P}} = \frac{\lambda^{\mathbb{P}}}{\sqrt{2\pi(\delta^{\mathbb{P}})^2}} \exp\left\{-\frac{\left(x-\mu^{\mathbb{P}}\right)^2}{2(\delta^{\mathbb{P}})^2}\right\} \quad (4\text{-}21),$$

$$\ell^{\mathbb{Q}} = \frac{\lambda^{\mathbb{Q}}}{\sqrt{2\pi(\delta^{\mathbb{Q}})^2}} \exp\left\{-\frac{\left(x-\mu^{\mathbb{Q}}\right)^2}{2(\delta^{\mathbb{Q}})^2}\right\} \quad (4\text{-}22).$$

Then, after a little bit of algebra, the relative entropy of $\mathbb{Q}|\mathcal{F}_T$ with respect to $\mathbb{P}|\mathcal{F}_T$ can be expressed as by applying the equation (4-19) with the Lévy measures (4-21) and (4-22):

$$T^{-1}\mathcal{E}(\mathbb{Q}|\mathbb{P}) = \frac{1}{2\sigma^2}\left\{\lambda^{\mathbb{Q}}\left(e^{\mu^{\mathbb{Q}}+\frac{\delta^{\mathbb{Q}2}}{2}}-1\right)-\lambda^{\mathbb{P}}\left(e^{\mu^{\mathbb{P}}+\frac{\delta^{\mathbb{P}2}}{2}}-1\right)\right\}^2$$

$$+\lambda^{\mathbb{Q}}\ln\left(\frac{\lambda^{\mathbb{Q}}\delta^{\mathbb{P}}}{\lambda^{\mathbb{P}}\delta^{\mathbb{Q}}}\right)+\lambda^{\mathbb{P}}+\lambda^{\mathbb{Q}}\left(-\frac{3}{2}+\frac{(\mu^{\mathbb{Q}}-\mu^{\mathbb{P}})^2+\delta^{\mathbb{Q}2}}{2\delta^{\mathbb{P}2}}\right) \quad (4.23)$$

The relative entropy function of the equation (4-23) is plotted in Figure 4.4 by setting $T = 0.25$, $\sigma = 0.1$, $\delta^{\mathbb{Q}} = 0.15$, $\lambda^{\mathbb{P}} = 1$, $\mu^{\mathbb{P}} = -0.1$, and $\delta^{\mathbb{P}} = 0.1$. Cont and Tankov (2004b) points out that the relative entropy function in the MJD case of the equation (4-23) is not a concave function in $\lambda^{\mathbb{Q}}$ $\mu^{\mathbb{Q}}$ $\delta^{\mathbb{Q}}$ because its Lévy measure is a nonlinear function in λ, μ, δ.

4.4.3 Regularized Calibration

The uniqueness and the stability of the solution of the NLS calibration problem of the equation (4-15) can be augmented by adding the relative entropy term:

$$\theta^{\mathbb{Q}} = \arg\min_{\theta^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + \alpha \mathcal{E}_T(\mathbb{Q}|\mathbb{P}) \qquad (4\text{-}24),$$

where α is a regularization parameter. The convexity of the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ makes the non-convex objective function more convex, thus, enhancing the uniqueness and the stability of the solution. According to Cont and Tankov (2004a, b), the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ remains convex in the neighborhood of its global minimum as long as the parameterization is well-behaved when $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ is not strictly convex with respect to $\theta^{\mathbb{Q}}$.

By adding and minimizing the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$, we are making the risk-neutral martingale measure $\mathbb{Q}$ as close as possible to the prior measure $\mathbb{P}$. For example, consider a case where $\mathbb{Q}_1$ exactly matches the market option prices. but $\mathbb{Q}_1$ is far away from the prior $\mathbb{P}$:

$$\sum_{i=1}^{N}\left|C_i^{\text{model}}(\theta^{\mathbb{Q}_1})-C_i^{\text{market}}\right|^2+\alpha\mathcal{E}_T(\mathbb{Q}_1\,|\,\mathbb{P})=\alpha\mathcal{E}_T(\mathbb{Q}_1\,|\,\mathbb{P}),$$

In this case, we are willing to sacrifice the precision of the calibration and choose $\mathbb{Q}_2$ which is closer to the prior $\mathbb{P}$ within the error bounds of the bid-ask spread:

$$\sum_{i=1}^{N}\left|C_i^{\text{model}}(\theta^{\mathbb{Q}_2})-C_i^{\text{market}}\right|^2+\alpha\mathcal{E}_T(\mathbb{Q}_2\,|\,\mathbb{P})>\alpha\mathcal{E}_T(\mathbb{Q}_1\,|\,\mathbb{P}).$$

In other words, the regularization by the relative entropy means the introduction of the bias of the calibrated parameter vector $\theta_{\mathbb{Q}}$ under the risk-neutral martingale measure $\mathbb{Q}$ toward the prior parameter vector $\theta_{\mathbb{P}}$ under prior probability measure $\mathbb{P}$ rather than relying solely on the new information contained in the quoted option prices.

There are two important parameters which should be chosen with care. Those are the prior parameter vector $\theta_{\mathbb{P}}$ and the regularization parameter α. With respect to the choice of the prior $\theta_{\mathbb{P}}$, Cont and Tankov (2004a, b) suggest the followings. The first is to employ the historical prior which is estimated using the time series of the underlying price by the statistical method (i.e. maximum likelihood estimation). The second approach which does not require the time series data is to run an unregularized calibration of the equation (4-15) and use the successively updated unregularized risk-neutral parameter solution $\theta_{\mathbb{Q}}$ as the prior $\theta_{\mathbb{P}}$. The third is to use the long-run average of the unregularized risk-neutral parameter solutions $\theta_{\mathbb{Q}}$. The third approach is preferable over the second because the role of the prior is to enforce the stability in the solution.

A regularization parameter α is a weight assigned to the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ and cannot be approximated by *a priori* fixed number because of its dependence on the level of noise present in the data. When α is large, we relatively trust the prior information more than the new information contained in the market option prices. When α is small, we relatively trust the new information more. If $\alpha = 0$, the calibration problem reduces to a simple NLS. There are several approaches to compute the regularization parameter α a posteriori, but Cont and Tankov (2004a, b) propose a use of discrepancy principle by Morozov (1966) which is the oldest, very popular, and a simple posteriori choice rule for α. The first step is to run the unregularized NLS of the equation (4-15) and obtain the unregularized risk-neutral diffusion parameter estimate $\hat{\sigma}_{\alpha=0}$. The second step is to compute the model intrinsic *a priori* quadratic pricing error $\hat{e}_{\alpha=0}{}^2$ by running the unregularized NLS of the equation (4-15) with the fixed $\hat{\sigma}_{\alpha=0}$. Let $e^2(\alpha, \theta_\alpha)$ be a posteriori model intrinsic quadratic pricing error for a given regularization parameter $\alpha > 0$ with $\theta_\alpha = (\sigma_\alpha, \lambda_\alpha, \mu_\alpha, \delta_\alpha)$. We expect:

$$e^2(\alpha, \theta_\alpha) > \hat{e}_{\alpha=0}{}^2,$$

because of the addition of the relative entropy. Morozov (1966) is willing to trade off the precision of the calibration for the numerical stability and the uniqueness of the calibration solution within the error bounds of the model intrinsic *a priori* quadratic pricing error $\hat{e}_{\alpha=0}{}^2$, thus, a regularization parameter α can be estimated by numerically solving the following equation with gradient-descent algorithms:

$$e^2(\alpha,\theta_\alpha) = c\hat{e}_{\alpha=0}{}^2 \qquad (4\text{-}25),$$

where $c = 1.2$, for example. The final regularized calibration result $\theta^{\mathbb{Q}}$ can be obtained by numerically solving the following optimization problem with $\hat{\alpha}$ using gradient-descent algorithms:

$$\theta^{\mathbb{Q}} = \arg\min_{\theta^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + \hat{\alpha}\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) \qquad (4\text{-}26).$$

4.5 Empirical Example: Obtaining the Prior

The first and probably the most important step toward the implementation regularized calibration with relative entropy is to obtain the prior probability measure $\mathbb{P}$ because the addition of the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ means the introduction of the bias of the risk-neutral martingale measure $\mathbb{Q}$ to the prior measure $\mathbb{P}$. In this section, two very different ways of obtaining the prior $\mathbb{P}$ are implemented using index options.

4.5.1 Computation of the Statistical Prior with Time-Series Data

Our data consist of daily closing prices of the futures contract on the S&P 500 index with March 2005 maturity obtained from Chicago Mercantile Exchange (CME) Daily Bulletin for the period March 24, 2004, through March 17, 2005 for the total of 248 trading days. Thus, we have a sample of log return series $(R_{t=1,2,\ldots,N})$ of size $N = 247$ where the log return is defined as:

$$R_t = \ln\left(\frac{F_t}{F_{t-1}}\right),$$

where F_t is the futures price.

According to Table 4.2, the true daily probability distribution of the log return of S&P 500 futures price for this sample shows a slight negative skewness -0.01644 and almost zero excess kurtosis of -0.0003468. True distribution is plotted in Figure 4.5 using the kernel density estimator with the Gaussian kernel with the bandwidth 0.002:

$$\mathbb{P}_N(R) = \frac{1}{0.002N}\sum_{t=1}^{N}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{R-R_t}{0.002}\right)^2\right\} \quad (4\text{-}27).$$

Anderson-Darling (AD) test developed by Stephens (1974) is employed to perform the goodness-of-fit test of the following hypotheses:

H_0 : The sample of log return $(R_{t=1,2,\ldots,N})$ comes from a population with a normal distribution function .

H_1 : H_0 is not true.

The advantage of AD test over the Kolmogorov-Smirnov test is its sensitivity for the tails of the distribution. AD test statistic is calculated as:

$$A^2 = -N - \frac{1}{N}\left[\sum_{i=1}^{N}(2i-1)\left\{\ln\Phi\left(\frac{R_i-\bar{R}}{s_R}\right)+\ln\left(1-\Phi\left(\frac{R_{N+1-i}-\bar{R}}{s_R}\right)\right)\right\}\right] \quad (4\text{-}28),$$

where Φ is the standard normal cumulative density function, $\bar{R}$ is the sample mean, and s_R is the sample standard deviation. Note that the sample values are rearranged in ascending order:

$$R_1 \le R_2 \le ... \le R_N .$$

In our example, the modified AD test statistic is computed as:

$$A^2* = A^2\left(1+\frac{4}{N}-\frac{25}{N^2}\right)=1.1412,$$

which exceeds the critical value of 0.787 at 5% confidence level. Thus, the normality of the log returns is rejected in this sample.

We use the MLE to estimate the daily probability distribution of the log return under the statistical probability measure $\mathbb{P}$ in the case of the MJD model. MLE is an estimation of the prior model parameter vector $\theta_{\mathbb{P}}$ to maximize the likelihood of observing the particular series of observations. The optimization problem in terms of the log likelihood function is:

$$\max_{\theta_{\mathbb{P}}} l(\theta_{\mathbb{P}}) = \sum_{t=1}^{N} \ln \mathbb{P}(R_t;\theta_{\mathbb{P}}),$$

where $(R_{t=1,2,...,N})$ is the sample log return series and $\mathbb{P}(R_t;\theta_{\mathbb{P}})$ is the log return probability density given by the equation (4-12) for the Merton jump diffusion model. Note because these are risk neutral densities, they need to be converted to the statistical density by replacing the instantaneous risk free interest rate r by the instantaneous return on the asset m and the convexity correction $\varpi_{\mathbb{P}}$ is under the statistical probability measure $\mathbb{P}$.

For the estimation of MJD model, small time approximation for the likelihood of increments in the time interval $[t, t+\Delta]$ is employed following Cont and Tankov (2004a):

$$\mathbb{P}(R_\Delta;\theta) \simeq \lambda\Delta\mathbb{P}(R_\Delta \,|\, j=1;\theta) + (1-\lambda\Delta)\mathbb{P}(R_\Delta \,|\, j=0;\theta) \qquad (4\text{-}29),$$

where $\mathbb{P}(R_\Delta \mid j=1;\theta)$ is the conditional probability density of log return R in a small time Δ given that one jump has occurred and $\mathbb{P}(R_\Delta \mid j=0;\theta)$ is the conditional probability density of log return R in a small time Δ given that jump did not occur.

The results are reported in Table 4.1 and 4.2. Figure 4.5 gives the estimated log return probability density plot and Figure 4.6 provides the plot of the log probability density to better illustrate the tail behavior. As expected, the true log return distribution is characterized by the higher peak, the heavier lower tail, and the thinner upper tail than the BS density. These important features of the true log return probability density are captured better by the MJD model. Similar to Honoré (1998) and Cont and Tankov (2004a), the large estimated value of the intensity parameter in the MJD model $\lambda_{\mathbb{P}} = 60$ (this means that on average the process jumps 60 times per year) casts a doubt to the legitimacy of modeling jumps as a rare event under the statistical measure $\mathbb{P}$. This large estimated value of the intensity parameter in the MJD model suggests that the log return process actually moves by frequent jumps instead of diffusion process and therefore this is an indication to resort to the Lévy models of infinite activity.

Next, Figure 4.7 plots the statistically estimated MJD Lévy measure of the equation (4-5) using the reported values in Table 4.1 for the range of $[\mu_{\mathbb{P}} - 10\times10^{-8}$, $\mu_{\mathbb{P}} + 10\times10^{-8}]$. The reason that this MJD Lévy measure is a spike at $\mu_{\mathbb{P}}$ is due to the extremely small value of the estimated standard deviation parameter of the log return jump size $\delta_{\mathbb{P}}$.

For a notational clarity, let $\mathbb{P}^S$ be the statistical prior probability measure and $\theta_{\mathbb{P}}^S$ be the statistical prior parameter vector which is:

$$\theta_{\mathbb{P}}^S = (\sigma_{\mathbb{P}}^S = 0.08085, \lambda_{\mathbb{P}}^S = 60, \mu_{\mathbb{P}}^S = -0.010476, \delta_{\mathbb{P}}^S = 1.600779 \times 10^{-9}) \quad \text{(3-30)},$$

where the superscript 'S' means 'statistical'.

4.5.2 Computation of the Risk-Neutral Prior with Option Data

Our data consist of daily settlement call option prices on S&P 500 futures with March 2005 maturity obtained from Chicago Mercantile Exchange (CME) Daily Bulletin for the sample period from March 24, 2004, through March 16, 2005 for the total of 248 trading days[19]. These American style option prices are converted to European style option prices using Barone-Adesi and Whaley (1987) quadratic approximation method to adjust for the early exercise premium. After eliminating call prices less than 0.125 due to reliability issues, the data used consist of a total of 6567 call prices. Daily series of three month Treasury Bill rate are used as appropriate risk-free interest rates.

Using the above data, simply running an unregularized calibration of the equation (4-15) on each day for the sample period produces the successively updated risk-neutral prior parameter solutions. Let $(\mathbb{P}_t^{RN})$ be the series (i.e. $t = 1, 2, ..., 248$) of successively updated risk-neutral prior probability measures and $(\theta_{\mathbb{P},t}^{RN})$ be the series of successively updated risk-neutral prior parameter vectors which are illustrated in Figure 4.8.

[19] March 16, 2005 is one day before the last trading day.

Taking the average of the series of successively updated risk-neutral prior parameter vectors $(\theta_{\mathbb{P},t}^{RN})$ yields the (fixed) risk-neutral prior parameter vector $\theta_{\mathbb{P}}^{RN}$ and let $\mathbb{P}^{RN}$ be the (fixed) risk-neutral prior probability measure. Table 4.3 reports the MJD risk-neutral prior parameter vector as follows:

$$\theta_{\mathbb{P}}^{RN} = (\sigma_{\mathbb{P}}^{RN} = 0.09544, \lambda_{\mathbb{P}}^{RN} = 0.77742, \mu_{\mathbb{P}}^{RN} = -0.14899, \delta_{\mathbb{P}}^{RN} = 0.09411) \quad (3\text{-}31).$$

Figure 4.9 gives the plot of the risk-neutral log return probability density with the maturity $T = 0.25$ years and Figure 4.10 provides the plot of the log of Figure 4.9 to better illustrate the tail behavior. As expected, the MJD model captures the negative skewness and the excess kurtosis of the log return probability density under the risk-neutral probability measure.

Next, Figure 4.11 plots the risk-neutral MJD Lévy measure of the equation (4-5) using the reported values in Table 4.3. It is symmetric around the mean log return jump size $\mu_{\mathbb{P}}^{RN}$ and its total mass is equal to the intensity $\lambda_{\mathbb{P}}^{RN}$. When Figures 4.7 and 4.11 are compared, we notice the remarkable difference of the Lévy measure under the statistical probability measure and the risk-neutral probability measure.

4.5.3 Comparison between the Statistical Prior and the Risk-Neutral Prior

Figure 4.12 illustrates the difference between the statistical prior probability measure $\mathbb{P}^{S}$ and the risk-neutral prior probability measure $\mathbb{P}^{RN}$ for four different maturities $T = 0.25,\ 0.5,\ 0.75,$ and 1. Table 4.5 reports the difference in the standardized moments. We observe that the volatility, the negative skewness and the excess kurtosis all become more pronounced under risk-neutral measure $\mathbb{P}^{RN}$ due to

the heavier lower tails. This is a well-documented fact which is due to the market participants' fear for the market crush.

Notice also the remarkable difference between the statistical prior parameter vector $\theta_{\mathbb{P}}^{S}$ of the equation (4.30) and the risk-neutral prior parameter vector $\theta_{\mathbb{P}}^{RN}$ of the equation (4.31). Under the risk-neutral probability measure, the asset price jumps only less than once (i.e. 0.77742 times) compared to 60 jumps under the statistical measure. This indicates that modeling jumps as rare events is legitimate under the pricing measure, but not under the statistical measure. The risk-neutral mean log return jump size $\mu_{\mathbb{P}}^{RN}$ is -14.899% which is far larger than the statistical counterpart $\mu_{\mathbb{P}}^{S}$ of -1.0476%. The uncertainty regarding the log return jump size is also much larger under the risk-neutral measure with $\delta_{\mathbb{P}}^{RN} = 0.09411$ compared to the statistical counterpart of $\delta_{\mathbb{P}}^{S} = 1.600779 \times 10^{-9}$.

4.6 Empirical Result of Regularized Calibration with Relative Entropy

In this section, we apply Cont and Tankov's (2004 a, b) regularization method with the relative entropy described briefly in the previous section to the index options parametrically with the MJD model. We compare the result of the calibration under the following three different methods.

Method 1 (M1): Unregularized calibration of the equation (4-15).

Method 2 (M2): Regularized calibration of the equation (4-24) with the statistical prior of the equation (4-30).

Method 3 (M3): Regularized calibration of the equation (4-24) with the (fixed) risk-neutral prior.

4.6.2 Empirical Results

We describe the implemented regularization procedures step by step using the data on March 30, 2004 (245 days to maturity).

Method 1 (M1)

This is the simplest and fastest in computational time and the calibrated risk-neutral parameter vector is shown in Table 4.6. The sum of the squared pricing errors (the value of the objective function) is 0.132312.

Method 2 (M2)

The first step is to run the unregularized NLS which is the Method 1. We obtain the unregularized risk-neutral diffusion parameter estimate:

$$\hat{\sigma}_{\alpha=0} = 0.108775 .$$

The second step is to compute the model intrinsic *a priori* quadratic pricing error $\hat{e}_{\alpha=0}{}^2$ by running the unregularized NLS (i.e. Method 1) with $\hat{\sigma}_{\alpha=0} = 0.108775$. We obtain:

$$\hat{e}_{\alpha=0}{}^2 = 0.132311.$$

The third step is to estimate the regularization parameter α by trading off the precision of the calibration for the numerical stability and the uniqueness of the calibration solution within the error bounds of the model intrinsic *a priori* quadratic pricing error $\hat{e}_{\alpha=0}{}^2$. In other words, the third step is to numerically solve the following equation with gradient-descent algorithms:

$$e^2(\alpha,\theta_\alpha)=\min_{\theta_\alpha^{\mathbb{Q}}}\sum_{i=1}^{N}\left|C_i^{\text{model}}(\theta_\alpha^{\mathbb{Q}})-C_i^{\text{market}}\right|^2+\alpha\mathcal{E}_T(\mathbb{Q}|\mathbb{P})\approx 1.2\times 0.132311 \qquad (4\text{-}32).$$

But, the problem is that the equation (4-32) has no solution on $\mathbb{R}$. The cause is the remarkable difference between the statistical prior parameter vector $\theta_{\mathbb{P}}^S$ of the equation (4-30) and the unregularized risk-neutral solution. Suppose that the solution of the regularized calibration problem is close to the unregularized risk-neutral solution. Using the statistical prior $\theta_{\mathbb{P}}^S$ of the equation (4-30) and the result of Method 1 as the solution of the regularized calibration problem, the relative entropy turns out to be too large because the statistical measure and the risk-neutral measure is too different:

$$\mathcal{E}_T(\mathbb{Q}^{M1}|\mathbb{P}^S)=4.99078\times 10^{15}.$$

In this case, the equation (4-32) becomes:

$$\min_{\theta_\alpha^{\mathbb{Q}}}\sum_{i=1}^{N}\left|C_i^{\text{model}}(\theta_\alpha^{\mathbb{Q}})-C_i^{\text{market}}\right|^2+\alpha\times 4.99078\times 10^{15}\approx 1.2\times 0.132311 \qquad (4\text{-}33).$$

The solution of the regularization parameter α in the equation (4-33) is approximately zero. Therefore, there is no meaning to regularize the calibration problem. But, this is because of the assumption that the solution of the regularized calibration problem is close to the unregularized risk-neutral solution. Next, consider the opposite case in which the solution of the regularized calibration problem is close to the statistical prior $\theta_{\mathbb{P}}^S$ of the equation (4-30). Suppose that the solution of the regularized calibration problem is equal to the statistical prior $\theta_{\mathbb{P}}^S$ of the equation (4-30) for simplicity. The relative entropy is zero by definition:

$$\mathcal{E}_T(\mathbb{P}^S \,|\, \mathbb{P}^S) = 0 .$$

Thus, the value of the regularization parameter α does not really matter and the sum of the squared pricing errors is calculated as:

$$\sum_{i=1}^{N} \left| C_i^{\text{model}}(\mathbb{P}^S) - C_i^{\text{market}} \right|^2 = 2834.1 .$$

Again, this is because of the significant difference between the statistical measure and the pricing measure. This example illustrates the fact that using the statistical prior $\theta_{\mathbb{P}}^S$ of the equation (4-30), the regularized calibration solution cannot be close to the unregularized risk-neutral solution nor the statistical prior $\theta_{\mathbb{P}}^S$. In addition, there is no regularized calibration solution which is in-between the unregularized risk-neutral solution and the statistical prior $\theta_{\mathbb{P}}^S$ because of the remarkable difference between these two. Numerically speaking, this is equivalent to stating that using the statistical prior $\theta_{\mathbb{P}}^S$, the equation (4-32) has no real-valued solution with constraints:

$$\alpha, \sigma, \lambda, \text{ and } \delta > 0 .$$

Thus, the regularized calibration with the statistical prior is not implementable.

Method 3 (M3)

The first two steps are the same as Method 2. The third step is to numerically solve the equation (4-32) which is solvable this time because of the proximity between the risk-neutral prior $\mathbb{P}^{RN}$ and the regularized calibration solution (these are both risk-neutral measures). Numerically solving the equation (4-32) yields the estimate for the regularization parameter as:

$$\hat{\alpha} = 0.043162 .$$

Finally, regularized calibration result $\theta^{\mathbb{Q}}$ shown in Table 4.6 can be obtained by numerically solving the following optimization problem with $\hat{\alpha}$ using gradient-descent algorithms:

$$\theta^{\mathbb{Q}} = \arg\min_{\theta^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + 0.043162 \times \mathcal{E}_T(\mathbb{Q}|\mathbb{P}) .$$

The sum of the squared pricing errors (the value of the objective function) is 0.156149:

$$\sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + 0.043162 \times \mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = 0.156149 ,$$

where $\sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 = 0.132777$ and $\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = 0.541481$.

Table 4.6 reports the results of the calibration, and Figure 4.13 through 4.18 compare the calibrated log return probability density and the calibrated Lévy measure for each method on six different maturity dates. We observe that the calibrated risk-neutral parameters are quite similar with or without regularization. This can be confirmed by observing the almost same calibrated log return probability densities and the Lévy measures between Method 1 and Method 3. It seems that Lévy measures are more sensitive to the regularization (i.e. the small difference in the calibrated parameters) than log return probability measures which is shown in Panel C of Figure 4.18. We also find that the difference in calibrated parameters between the regularized and the unregularized become more pronounced especially for near maturity options shown in Figure 4.18.

We believe that the reason of this no significant difference between the unregularized calibration and the regularized calibration is the use of $c = 1.2$ in the equation (4-25) and (4-32). This constant c controls the degree of trading off the precision for the numerical stability and uniqueness of the calibration. The larger c indicates more willingness to sacrifice the precision and $c = 1$ means that the calibration is unregularized (i.e. $\alpha = 0$). It is our opinion that the use of $c = 1.2$ does not sacrifice the precision too much for the numerical stability and uniqueness which means that the regularized calibration result will by design be close to the unregularized calibration result. But, we are not suggesting the use of larger c (for example 2) because it can be a too much of sacrifice of the precision. So, what is the optimal value for c? Cont and Tankov (2004a, b) recommend the use of $1.1 < c < 1.5$. But, in the end, the choice of c is entirely up to the user's discretion. One thing to remind is that the larger c means the introduction of larger bias toward the prior and the regularized calibration result will show more significant difference from the result of the unregularized calibration. But, the use of larger c means the user's stronger belief in the prior information than the today's information. Such situation is very difficult to imagine.

4.7 Conclusion

In this paper, we investigate the effect of the regularization with the relative entropy in the framework of the parametric calibration of the Merton jump-diffusion model to the index options. It is important to realize that using the relative entropy

and using the prior mean the introduction of the bias of the calibration solution toward the prior to gain the numerical stability and the uniqueness by making the objective function more convex. This means that the user has some belief in the use of the prior (i.e. otherwise, why do you bother?). Another important point is that the regularization is not the method to locate the global solution, it is the method to enhance the uniqueness and the stability of the calibration solution by sacrificing its precision. In terms of the choice of the prior, the only implementable prior is the risk-neutral prior. The regularized calibration with the statistical prior is not implementable because the statistical prior is a statistical measure and it is too much away from the risk-neutral (i.e. pricing) measure.

The result shows that with or without regularization, calibrated risk-neutral parameters, calibrated log return probability densities, and calibrated Lévy measures are not significantly different. It seems that Lévy measures are more sensitive to the regularization than log return probability measures. Notice also that the difference in calibrated parameters between the regularized and the unregularized become more pronounced especially for near maturity options.

From our empirical result that the regularization with the relative entropy using the risk-neutral prior does not make any significant difference in the solution and considering the extra computation time necessary for the regularization procedures, we prefer the calibration without any regularization which uses only today's information and yields the unbiased solution although the unregularized calibration solution may not be stable and unique.

Chapter 5: Other Approaches to Obtain Option Price Implied Risk-Neutral Probability Density Function of Asset Prices

5.1 Introduction

Market participants' expectations of the future underlying asset price moves can be recovered from option prices because option prices depend on the market participants' expectations of the future underlying asset price moves. This option implied dynamics of risk-neutral probability density functions (RNDs) of future underlying asset price provide valuable information for risk managers in making investment decisions which is very different from the information provided by Benchmark Black-Scholes lognormal dynamics of RNDs. Previous studies show that Black-Scholes assumption of constant volatility of the underling's return fails to capture the true dynamics of volatility and this leads to the failure of the Black-Scholes to capture the true dynamics of the expectation of future asset price move.[20] For example, negatively sloped Black-Scholes implied volatilities across exercise prices on any trading day indicate the negatively skewed RND of underlying asset price. This means that option prices reveal that the likelihood of an extreme downward price move of the underlying is greater than that of an extreme upward move. Also the smile or sneer pattern of the Black-Scholes implied volatilities across exercise prices indicates that option prices reveal the likelihood of extreme move of the future underling's price is much greater than that allowed by the Black-Scholes. Thus obtaining the evolution in time of RNDs of financial asset prices from option prices can provide the dynamic behavior of market's assessment of risks.

[20] For example, Rubinstein (1994), Dumas, Fleming, and Whaley (1998), and Cont and da Fonseca (2002).

Among numerous techniques to estimate the RND of underlying asset price on the option's maturity date using option prices, two popular but different classes of methods have been developed.

The first class of method is based on the relationship that the RND of the underlying asset price at the option's maturity date can be obtained by twice differentiating the call price function with respect to the strike price discovered by Breeden and Litzenberger (1978). In order to take the second derivative of the call price function, a continuous call price function is necessary. Obviously, no options are traded at continuous strike prices, rather they are traded at very limited number of strike prices. Thus, this method basically comes down to a method of interpolation and extrapolation. Bates (1991) employs a cubic spline (a peace-wise third order polynomials) to interpolate the observed call option prices subject to constraints. This approach requires a relatively large number of degrees of freedom because of the complex form of the call price function. Ait-Sahalia and Lo (1998) use a non-parametric kernel regression to estimate the call price function. This approach is not practical to implement because of its data-intensive nature. We need to make assumptions simply to reduce the dimensionality of the problem since this approach involves a large number of regressors. Instead of directly interpolating call option prices, Shimko (1993) first interpolates Black-Scholes implied volatilities by fitting a quadratic function across strike prices. Then this volatility smile is inverted to obtain a continuous call pricing function as a function of strike price through the Black-Scholes formula. Campa, Chang, and Reider (1998) interpolates Black-Scholes implied volatilities by fitting a cubic spline.

This type of semiparametric method imposes no specific dynamics on the underlying asset price and makes no assumptions about the parametric form of the RND. But the shortcoming is that the different methods of interpolation produce very different results in the form of RNDs.

The second class of method starts from imposing a particular parametric form on the RND and then recovers its parameters by the minimization between the observed option prices and model prices generated by the assumed parametric form.[21] Melick and Thomas (1997) use a mixture of three lognormals. Bahra (1997) uses a mixture of two lognormals. This is a more restrictive method than the first class.

Contrast to option implied methods to recover RND, the traditional methods start from imposing strong assumptions on the dynamics of underlying financial asset price. These strong assumptions enable RND to be estimated in closed form. Typically, geometric Brownian motion of the underlying asset price is assumed[22] which results in the benchmark Black-Scholes lognormal distribution of terminal asset price. This is the most restrictive case.

The goal of this paper is to assess the difference of the dynamics of RNDs between option implied methods and the benchmark Black-Scholes. None of the previous studies examined the difference in the dynamics of RNDs throughout the

[21] Melick and Thomas (1997) points out that this approach is more general than specifying the dynamics of the underlying asset price. Because a particular dynamics gives a unique RND but a particular RND is consistent with many different dynamics of underlying asset price.

[22] Together with the constant volatility of asset price return and the constant risk-free interest rate.

life of an option from the start till the end between the option implied and the traditional Black-Scholes.

A time series of S&P 500 futures option prices with December 2003 maturity traded on the Chicago Mercantile Exchange (CME) for the period of approximately one year between December 23 2002 and December 18 2003 is used. For each day in the sample, RND of underlying S&P 500 index on the option's expiration date December 19 2003 is estimated. The resulting dynamics of RNDs for the entire life of the option is plotted.

We find that the option implied dynamics of RNDs is characterized by the well-documented features of negative skewness and excess kurtosis of the underlying asset price relative to Black-Scholes lognormal RNDs[23]. This feature becomes more pronounced as the maturity of the option approaches but may become less pronounced at very near maturity.

We examine the goodness of fit of the benchmark Black-Scholes lognormal dynamics to the option implied dynamics. Kolmogorov-Smirnov test (KS-test) concludes at the 5% significance level that the Black-Scholes RNDs do not come from the option-implied RNDs for approximately one-third of the sample period. The author attributes the observed pattern of the increasing deviation between the option-implied RNDs and the Black-Scholes RNDs when the maturity of the option is far and near to the use of volatility surface by the option-implied RNDs. This highlights the shortcoming of the Black-Scholes assumption of the constant volatility.

---

[23] Rubinstein (1994) and Dumas, Fleming, and Whaley (1998).

The remainder of the paper is organized as follows. Section 5.2 describes the methodology for estimating RNDs of financial asset price. Section 5.3 explains the data set of S&P 500 futures option prices with December 2003 maturity obtained from CME and presents the estimated dynamics of RNDs. We document the time series behavior of mean, standard deviation, skewness, and kurtosis for each of three different methods. Section 5.4 examines the goodness of fit of the Black-Scholes lognormal dynamics to the option implied dynamics of RNDs. Section 4.5 concludes.

We should be aware of the fact that the estimated RND is a risk-neutral distribution, not the actual distribution throughout this literature.

5.2 Option Implied Method and Traditional Method to Estimate Risk Neutral Density (RND) of Financial Asset Price

5.2.1 Volatility Interpolation (VI) Method of Shimko (1993)

Under Black-Scholes assumptions, the price of European call and put options at date t maturing at date $T \equiv t + \tau$, written on a stock with the price S_t at date t, and the strike price K can be written as the present value of the expected future payoffs following Cox, Ross, Rubinstein (1979):

$$\text{call}\left(K, \tau = T - t\right) = e^{-r\tau} \int_{K}^{\infty} \left(S_T - K\right) f\left(S_T\right) dS_T \qquad (5\text{-}1)$$

$$\text{put}\left(K, \tau = T - t\right) = e^{-r\tau} \int_{0}^{K} \left(K - S_T\right) f\left(S_T\right) dS_T \qquad (5\text{-}2),$$

where the present value is calculated with respect to the risk-free interest rate r. The expectation is calculated with respect to $f(S_T)$ which is the risk-neutral density (RND) of the stock at date T.

The explicit expression of RND can be obtained from option prices following Breeden and Litzenberger (1978). Rearrange the equation (5-1) as:

$$\text{call}(K,\tau) = e^{-r\tau}[\int_K^\infty S_T f(S_T) dS_T - K\int_K^\infty f(S_T) dS_T] \quad (5\text{-}3).$$

The partial derivative of call option price function (5-3) with respect to the strike price K becomes:

$$\begin{aligned}\frac{\partial \,\text{call}(K,\tau)}{\partial K} &= e^{-r\tau}\left[-\int_K^\infty f(S_T) dS_T\right] \\ &= -e^{-r\tau}\left[1 - F(S_T = K)\right] \\ &= -e^{-r\tau}\int_K^\infty f(S_T) dS_T \quad (5\text{-}4),\end{aligned}$$

where $F(S_T)$ is the risk-neutral cumulative density function of the stock price at date T. $\int_K^\infty f(S_T) dS_T = 1 - F(S_T = K)$ is the probability that the stock price exceeds the strike price at the maturity (i.e. call option finishes in the money). The partial derivative of the equation (5-4) with respect to the strike price K gives the RND of stock price at date T:

$$\frac{\partial^2 \,\text{call}(K,\tau)}{\partial K^2} = e^{-r\tau} f(S_T = K) \quad (5\text{-}5)$$

Taking derivatives requires continuous call option pricing functions. Instead of directly interpolating call option prices,[24] Shimko (1993) first interpolates Black-Scholes implied volatilities using a quadratic function across strike prices.[25] Then this volatility smile is inverted to obtain a continuous call pricing function as a function of strike price through the Black-Scholes formula.[26]

A method of normalization is necessary for the recovered RND to behave nicely. This is for the RND to have an integral of one, for the tails of the RND to decline monotonically and decay quickly. For this purpose, lognormal probability density is grafted for strike prices outside the observed range by matching the density and cumulative density of the option implied density function with a lognormal probability density in both lower and upper tail.[27]

5.2.2 Mixture of Two Lognormals Method (2LN) of Bahra (1997)

This method begins by specifying the parametric form of RND directly as a mixture of two lognormal distributions. Bahra shows that the negative skewness and the excess kurtosis of the data are well captured by this parametric form and this parametric form ensures for the tails of the RND to decline monotonically and decay quickly. The RND is then estimated by minimizing the distance of the model-generated prices to the market option prices.

---

[24] Directly interpolating call prices can produce inaccurate RND because small price errors can be transformed into large errors in the estimated RND, especially in the tails.

[25] Campa, Chang, and Reider (1998) interpolates the implied volatilities using cubic splines.

[26] Black-Scholes formula is merely used as a means of mapping option prices in terms of implied volatilities.

[27] This means that the implied volatilities are constant outside the range of traded strike prices.

The RND is given a mixture of two-lognormal distributions:

$$f(S_T) = \pi L(\mu_1, \sigma_1; S_T) + (1-\pi) L(\mu_2, \sigma_2; S_T) \quad (5\text{-}6)$$

where $L(\mu_i, \sigma_i; S_T)$ is the probability density function of lognormal distribution meaning that whose logarithm is normal with mean μ_i and standard deviation σ_i. Parameter π is the weight assigned to each lognormal distribution. The mean and standard deviation parameters for each lognormal distribution μ_i and σ_i together with the weight assigned to each π determines the overall shape of the RND.

By substituting (5-6) into (5-1) and (5-2), now the call and put option pricing functions (5-1) and (5-2) can be written as:

$$\text{call}(K, \tau = T-t)$$

$$= e^{-r\tau} \int_K^\infty \left[\pi L(\mu_1, \sigma_1; S_T) + (1-\pi) L(\mu_2, \sigma_2; S_T) \right] (S_T - K) dS_T \quad (5\text{-}7),$$

$$\text{put}(K, \tau = T-t)$$

$$= e^{-r\tau} \int_0^K \left[\pi L(\mu_1, \sigma_1; S_T) + (1-\pi) L(\mu_2, \sigma_2; S_T) \right] (K - S_T) dS_T \quad (5\text{-}8).$$

Next substitute the density functions of lognormal distributions:

$$L(\mu_1, \sigma_1; S_T) = \frac{1}{S_T \sigma_1 \sqrt{2\pi}} \exp[-\frac{(\ln S_T - \mu_1)^2}{2\sigma_1^2}] \quad (5\text{-}9),$$

$$L(\mu_2, \sigma_2; S_T) = \frac{1}{S_T \sigma_2 \sqrt{2\pi}} \exp[-\frac{(\ln S_T - \mu_2)^2}{2\sigma_2^2}] \quad (5\text{-}10).$$

Bahra (1997) demonstrated that using change of variables twice allows a transformation from lognormal distributions to normal distributions. Equations (5-7) and (5-8) have closed form solutions:

$$\text{call}(K,\tau) = e^{-r\tau}[\pi\{\exp(\mu_1 + \frac{1}{2}\sigma_1^2)N(d_1) - N(d_2)\}$$

$$+(1-\pi)\{\exp(\mu_2 + \frac{1}{2}\sigma_2^2)N(d_3) - KN(d_4)\}] \quad (5\text{-}11),$$

$$\text{put}(K,\tau) = e^{-r\tau}[\pi\{-\exp(\mu_1 + \frac{1}{2}\sigma_1^2)N(-d_1) + KN(-d_2)\}$$

$$+(1-\pi)\{-\exp(\mu_2 + \frac{1}{2}\sigma_2^2)N(-d_3) + KN(-d_4)\}] \quad (5\text{-}12).$$

where $d_1 = \dfrac{-\ln S_T + \mu_1 + \sigma_1^2}{\sigma_1}$, $d_2 = d_1 - \sigma_1$, $d_3 = \dfrac{-\ln S_T + \mu_2 + \sigma_2^2}{\sigma_2}$, and $d_4 = d_3 - \sigma_2$.

The RND is then estimated by minimizing the sum of squared errors between the fitted option prices by the model and market option prices across all exercise prices. The minimization problem is:

$$\underset{\mu_1,\mu_2,\sigma_1,\sigma_2,\pi}{\text{Min}} \quad \sum_{i=1}^{n}\left[\text{call}(K_i,\tau) - \text{call}^{\text{market}}{}_i\right]^2 + \sum_{i=1}^{n}\left[\text{put}(K_i,\tau) - \text{put}^{\text{market}}{}_i\right]^2 + \left[\pi\exp[\mu_1 + \frac{1}{2}\sigma_1^2] + (1-\pi)\exp[\mu_2 + \frac{1}{2}\sigma_2^2] - e^{r\tau}S_t\right] \quad (5\text{-}13)$$

subject to $\sigma_1, \sigma_2 > 0$ and $0 \le \pi \le 1$, over the range of observed strike prices $K_1, K_2, ..., K_n$. The mean of the RND should equal the forward price of the stock $F_t = S_t e^{r\tau}$.

5.2.3 Traditional Method

Traditional method is presented using general notations first and then it is applied to this paper's purpose.

Suppose that the value of variable x follows the general stochastic differential equation

$$dx = A(x,t)dt + B(x,t)dw \qquad (5\text{-}14),$$

where the drift coefficient $A(x,t)$ and the diffusion coefficient $B(x,t)$ are functions of the variable x and date t and dw is a Wiener process. Probabilistic behavior of the variable x is represented by its conditional probability $p(x_T, T|x_t, t)$ in which x_t is the current value of variable x and t is the current date while x_T being the value of the variable x at some future date T. Fokker-Planck (FP) equation [28] [29]

$$\frac{\partial p(x_T, T|x_t, t)}{T}$$

$$= -\frac{\partial (A(x_T, T)\, p(x_T, T|x_t, t))}{\partial x_T} + \frac{1}{2}\frac{\partial^2 (B(x_T, T)^2\, p(x_T, T|x_t, t))}{\partial {x_T}^2} \qquad (5\text{-}15),$$

gives the evolution of the probability density function of the variable x in time. FP equation is an equation of motion for the distribution function of stochastic variable x. It is parabolic in the sense that it involves a second derivative with

[28] Also known as forward Kolmogorov equation.

[29] See Mathematical Appendix for its derivation.

respect to one variable x_T, and a first derivative with respect to the other variable T.[30]

Solving this linear second-order partial differential equation of parabolic type requires an initial condition and boundary conditions.

Consider a futures contract on a stock. The relationship between the futures price and the spot price is:

$$F_t = S_t e^{(r-q)(T-t)} \quad (5\text{-}16).$$

where F_t is the futures price at date t, S_t is the spot price at t, q is the rate of dividend yield, and T is the maturity date of the futures contract. Assume that the risk-neutral process for the spot price is given by geometric Brownian motion:

$$dS_t = (r-q)S_t dt + \sigma S_t dw \quad (5\text{-}17).$$

Applying Ito's lemma produces the futures price dynamics:

$$dF_t = \sigma F_t dw \quad (5\text{-}18)$$

This means that in a risk-neutral world the futures price follows geometric Brownian motion with zero drift coefficient and diffusion coefficient of σF_t.

Then, the Fokker-Planck equation which describes the dynamics of the RND becomes

$$\frac{\partial p(F_T, T|F_t, t)}{T} = \frac{1}{2}\frac{\partial^2(\sigma^2 F_T{}^2 p(F_T, T|F_t, t))}{\partial F_T{}^2} \quad (5\text{-}19).$$

Note that at the maturity of the futures the futures price is equal to the asset's spot price $F_T = S_T$.

[30] Equations of this type are known as heat or diffusion equations.

Utilizing the initial condition,

$$p(F_T, t \mid F_t, t) = \delta(F_T - F_t) \qquad (5\text{-}20)$$

the equation (5-19) can be solved as:

$$p(F_T, T \mid F_t, t) = \frac{1}{\sigma F_T \sqrt{2\pi(T-t)}} \exp[-\frac{[\ln(F_t / F_T) - \frac{1}{2}\sigma^2(T-t)]^2}{2\sigma^2(T-t)}] \qquad (5\text{-}21).$$

This is the benchmark Black-Scholes lognormal RND.

5.3 Estimating the Dynamics of RNDs from S&P 500 Futures Options Data

To compare difference in the dynamics of RNDs obtained from three different estimators, an application to the data set of futures option prices on S&P 500 index is presented.

5.3.1 Data Set

Our data consist of daily settlement prices of futures options on the S&P 500 index with December 2003 maturity obtained from Chicago Mercantile Exchange (CME) Daily Bulletin for the period December 23, 2002, through December 18, 2003. The time to maturity ranges from 0.9918 years (362 days) to 0.0027 years (1 day). Two filters are applied. Options with prices less than 0.125 are eliminated. In the money options are eliminated because they are illiquid. Thus all options used are out of money options: for moneyness = strike price / futures price greater than 1 call prices are used and for moneyness less than 1 put prices are used. LIBOR in US dollars with nearest maturity is used as an appropriate risk-free interest rate.

Since these are American-style options, we use Barone-Adesi and Whaley (1987) quadratic approximation method to adjust for the early exercise premium.[31]

5.3.2 Dynamics of RND Estimates

To implement volatility interpolation (VI) method Black-Scholes implied volatilities are calculated and interpolated by a quadratic function across the range of strike prices on each day. The resulting implied volatility surface is shown in Figure 5.1 as a function of the moneyness (= strike price / futures price) of the option and time to maturity. Figure 5.1 illustrates the well-documented feature of the implied volatility's dependence on moneyness and time-to-maturity[32]. On any single trading day the implied volatility is not constant over a range of moneyness and the shape of implied volatility curve changes as the time to maturity changes. As the maturity of the option approaches, the curve becomes steeper. This implied volatility surface is inverted to continuous call option prices through Black-Scholes formula. Twice differentiating this continuous call price function yields the RND. The resulting dynamics of RNDs obtained from volatility interpolation method is depicted in Figure 5.2. It captures the evolution in time of RNDs of S&P 500 index on the option's expiration date December 19 2003. As expected, RND becomes less dispersed as the option's expiration date approaches since the outcome becomes more certain at near maturity.

Mixture of two lognormals method (2LN) recovers the option implied RNDs by minimizing the distance between the model prices calculated by equations (5-11)

---

[31] See Mathematical Appendix for this procedure.
[32] Rubinstein (1994), Dumas, Fleming, and Whaley (1998), and Cont and da Fonseca (2002).

and (5-12) and the observed option prices according to the criterion function equation (5-13). The resulting dynamics of RNDs is illustrated in Figure 5.4.

Solution of the Fokker-Planck (FP) equation assuming the geometric Brownian motion on the underlying asset price dynamics yields the benchmark Black-Scholes lognormal RND of equation (5-21). The annualized historical volatility is calculated in a usual fashion:

$$\sigma = \sqrt{252} \times \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(R_i - \overline{R})^2}$$

where $R_i = \ln(F_t / F_{t-1})$.

Using spot futures price F_t and the historical volatility $\sigma = 0.167615335$, the evolution of RND in time is depicted in Figure 5.6.

5.3.3 Instantaneous Profile and the dynamics of Moments

Since it is very difficult to notice the difference among Figures 5.2, 5.4, and 5.6, the instantaneous profiles of the entire dynamics of RNDs are taken on the following four dates, March 18, June 18, September 18, and November 18 2003. These correspond to approximately nine-month, six-month, three-month, and one-month to maturity. These are reported in Figures 5.8 through 5.11 together with their moments on Tables 5.1 through 5.4. These instantaneous profiles show that the option implied RNDs possess heavier lower tails than the Black-Scholes RNDs suggesting that the market assesses the greater likelihood of extreme downward move of underlying asset price than allowed by the Black-Scholes.

The term structures of mean, standard deviation, skewness, and kurtosis of the RNDs are plotted in Figure 5.12.[33] Note the several differences between the option implied dynamics and the Black-Scholes dynamics. Although all three methods have same means and similar patterns of standard deviations, their skewness and kurtosis are remarkably different. The option implied dynamics is characterized by the negative skewness and excess kurtosis for asset prices documented by Ait-Sahalia and Lo (1998) which generally increase as the maturity of the option approaches. At very near maturity of within 15 days to expiration, the volatility interpolation method produces decreasing excess kurtosis of the underlying asset price. In contrast, the kurtosis recovered from mixture of two lognormals method continues to increase. Obviously, benchmark Black-Scholes lognormal dynamics of RNDs completely fails to capture these dynamic behaviors of skewness and kurtosis.

5.4 Goodness of Fit Test

This paper uses Kolmogorov-Smirnov test (KS-test) to examine the difference between the option implied dynamics of RNDs and the Black-Scholes dynamics of RNDs. KS-test is one of the best known and widely used goodness of fit test for continuous distributions to determine if the empirical distribution function $F_n'(x)$ obtained from a sample of observations $x_1, x_2, ..., x_n$ comes from the hypothesized known population distribution function $F(x)$. We hypothesize that the option implied RND describes the true population distribution $F(x)$ and test if Black-Scholes RND which is the empirical distribution $F_n'(x)$ calculated using the

[33] Moments are directly computed from the each RND.

historical volatility from 249 observations of futures option price comes from the $F(x)$. The KS-test statistic D is the maximum absolute deviation between $F_n'(x)$ and $F(x)$ across the range of the stochastic variable x:

$$D = \underset{x}{Max}\left|F_n'(x) - F(x)\right|.$$

If the test statistic D is greater than the critical value[34], the null hypothesis is rejected and we conclude that the empirical distribution does not come from the hypothesized population distribution.

Figure 5.13 reports the result of KS-test of the Black-Scholes RNDs to the option implied RNDs obtained from volatility interpolation method at 5% significance level. Null hypothesis is rejected on 80 trading days out of 249 (approximately 32%). Figure 5.14 reports the result of KS-test of the Black-Scholes RNDs to the option implied RNDs obtained from mixture of two lognormals approach and rejects the null on 92 trading days out of 249 (approximately 37%). Note that the difference becomes significant when the maturity of the option is far and near. In this paper, Black-Scholes dynamics of RNDs is obtained using the historical volatility for the sample period. On the other hand, option implied dynamics incorporate the volatility surface portrayed in Figure 5.1. When the maturity of the option is far (near), at the money implied volatility is sizably larger (smaller) than the historical volatility as reported in Figure 5.16. The time series of absolute deviation between the two is plotted in the Figure 5.17 which resembles

[34] Critical value is calculated by $1.36/\sqrt{n}$ where n is the sample size.

Figures 5.13 and 5.14 in that when the maturity of the option is far and near the difference becomes large.

The similarity of the two option implied dynamics is reported in Figure 5.15. The difference is very stable at approximately 2% except for a very short period of time at near maturity.

In summary, the difference between the option implied dynamics and the Black-Scholes dynamics of RNDs is attributed to the difference between the use of constant volatility by the Black-Scholes and the use of volatility surface by the option implied. Volatility surface captures the dependence of volatility on moneyness and the evolution of this volatility curve as the time to maturity of the option changes.

5.5 Conclusion

This paper investigates the dynamics of risk-neutral probability density functions (RNDs) of S&P 500 index using a time series of S&P 500 futures options prices. Two different methods, volatility interpolation method and mixture of two lognormals method, for estimating RNDs from option prices are used. The difference in the dynamic behavior of RNDs between the option implied method and the benchmark Black-Scholes lognormal method was then examined.

We find that the option implied dynamics of RNDs is characterized by the persistent negative skewness and excess kurtosis which become more pronounced as the maturity of the option approaches but may become less pronounced at very near

maturity of within 15 days to expiration. In contrast, the benchmark Black-Scholes lognormal dynamics yields completely different dynamics of skewness and kurtosis, persistent positive skewness and time-decreasing kurtosis.

Next, we performed a goodness of fit test of the benchmark Black-Scholes lognormal RNDs to the option implied RNDs. The results from Kolmogorov-Smirnov test suggest that at 5% significance level Black-Scholes lognormal RNDs do not come from option-implied RNDs on approximately one-third of the life of the option. The author attributes the observation that the difference between Black-Scholes RNDs and the option implied RNDs becomes significant when the maturity of the option is far and near to the use of volatility surface by the option implied RNDs.

Overall, our results indicate that option prices do not support the Black-Scholes world of constant volatility and the lognormal distribution of asset prices at option's expiration. The use of information provided by volatility surface is a key element in option pricing.

Table 1.1: Martingale Gambling Strategy

Trial	0	1	2	3	4	5	6	7
Result		Loss	Loss	Loss	Loss	Loss	Loss	Win
Bet		$2	$4	$8	$16	$32	$64	$128
Net Gain		-$2	-$4	-$8	-$16	-$32	-$64	+$128
Wealth	$200	$198	$194	$186	$170	$138	$74	$202

Table 1.2: One-to-one correspondence between an infinitely divisible distribution and a Lévy process

Infinitely divisible probability measure $\mathbb{P}$ $\rightleftarrows$	Lévy process
Normal distribution	Brownian motion (with drift)
Poisson distribution	Poisson process
Compound Poisson distribution	Compound Poisson process
Cauchy distribution	Cauchy process
Exponential distribution	Gamma process

Table 1.3: Poisson process

Lévy process	Gaussian variance	Lévy measure	drift	variation	sample path
Poisson process	$A = 0$	$\ell = \lambda\delta(x-1)$	$\gamma = 0$	finite	rcll step functions of step size 1

Table 1.4: Compound Poisson process

Lévy process	Gaussian variance	Lévy measure	drift	variation	sample path
Compound Poisson Process	$A = 0$	$\ell = \lambda f(x)$	$\gamma = 0$	finite	rcll step functions

Table 2.1: Relationship between frequency f and angular frequency ω

Frequency f (cycles/second)	Angular Frequency ω (radians/second)
1 Hz	$2\pi = 360°$ Hz
10 Hz	$20\pi = 360° \times 10$ Hz
100 Hz	$200\pi = 360° \times 100$ Hz

Table 2.2: Summary of Important Properties of Fourier Transform in Angular Frequency Domain ω Hz (radians/second)

Property $\mathcal{F}[y(t)](\omega)$	Time Domain Function $y(t)$	Fourier Transform
Linearity	$af(t)+bg(t)$	$a\mathcal{F}(\omega)+b\mathcal{G}(\omega)$
Even Function	$f(t)$ is even	$\mathcal{F}(\omega)\in\mathbb{R}$
Odd Function	$f(t)$ is odd	$\mathcal{F}(\omega)\in\mathbb{I}$
Symmetry	$\mathcal{F}(t)$	$2\pi f(-\omega)$
Differentiation	$\frac{df(t)}{dt}$	$-i\omega\mathcal{F}(\omega)$
	$\frac{d^k f(t)}{dt^k}$	$(-i\omega)^k\mathcal{F}(\omega)$
Time Scaling	$f(at)$	$\frac{1}{\lvert a\rvert}\mathcal{F}(\frac{\omega}{a})$
Time Shifting	$f(t-t_0)$	$e^{i\omega t_0}\mathcal{F}(\omega)$
Convolution	$f*g\equiv\int_{-\infty}^{\infty}f(\tau)g(t-\tau)d\tau$	$\mathcal{F}(\omega)\mathcal{G}(\omega)$
Multiplication	$f(t)g(t)$	$\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}(\varpi)\mathcal{G}(\omega-\varpi)d\varpi$
Modulation (Frequency Shifting)	$e^{-i\omega_0 t}f(t)$	$\mathcal{F}(\omega-\omega_0)$

Table 2.3: Summary of Important Properties of Fourier Transform in Frequency Domain f Hz (cycles/second)

Property $\mathcal{F}\left[y(t)\right](f)$	Time Domain Function $y(t)$	Fourier Transform
Linearity	$af(t)+bg(t)$	$a\mathcal{F}(f)+b\mathcal{G}(f)$
Even Function	$g(t)$ is even	$\mathcal{G}(f)\in\mathbb{R}$
Odd Function	$g(t)$ is odd	$\mathcal{G}(f)\in\mathbb{I}$
Symmetry	$\mathcal{G}(t)$	$g(-f)$
Differentiation	$\frac{dg(t)}{dt}$	$2\pi if\mathcal{G}(f)$
	$\frac{d^k g(t)}{dt^k}$	$(2\pi if)^k\mathcal{G}(f)$
Time Scaling	$g(at)$	$\frac{1}{\lvert a\rvert}\mathcal{G}(\frac{f}{a})$
Time Shifting	$g(t-t_0)$	$e^{-2\pi ift_0}\mathcal{G}(f)$
Convolution	$f*g\equiv\int_{-\infty}^{\infty}f(\tau)g(t-\tau)d\tau$	$\mathcal{F}(f)\mathcal{G}(f)$
Multiplication	$f(t)g(t)$	$\mathcal{F}*\mathcal{G}$
Modulation (Frequency Shifting)	$e^{2\pi if_0 t}g(t)$	$\mathcal{G}(f-f_0)$

Table 2.4: How to Calculate Standardized Moments from Characteristic Function and Moment Generating function

	Moments	n -th cumulant	n -th raw moment
Mean	$E[X]$	cum_1	r_1
Variance	$E\{X-E[X]\}^2$	cum_2	$r_2-r_1^2$
Skewness	$\frac{E\{X-E[X]\}^3}{\left(\sqrt{E\{X-E[X]\}^2}\right)^3}$	$\frac{cum_3}{cum_2^{3/2}}$	$\frac{2r_1^3-3r_1r_2+r_3}{(r_2-r_1^2)^{3/2}}$
Excess Kurtosis	$\frac{E\{X-E[X]\}^4}{\left(\sqrt{E\{X-E[X]\}^2}\right)^4}-3$	$\frac{cum_4}{cum_2^2}$	$\frac{-6r_1^4+12r_1^2r_2-3r_2^2-4r_1r_3+r_4}{(r_2-r_1^2)^2}$

Note:

- 2 -nd central moment of $X = E\{X-E[X]\}^2 = E[X^2]-E[X]^2 = Var[X]$ $= cum_2 = r_2 - r_1^2$.
- 3 -rd central moment of $X = E\{X-E[X]\}^3 = cum_3 = 2r_1^3 - 3r_1r_2 + r_3$.
- Skewness of $X = Skewness[X] = \frac{E\{X-E[X]\}^3}{\left(\sqrt{E\{X-E[X]\}^2}\right)^3} = \frac{cum_3}{cum_2^{3/2}}$ $= \frac{2r_1^3-3r_1r_2+r_3}{(r_2-r_1^2)^{3/2}}$.
- 4 -th central moment of $X = E\{X-E[X]\}^4 = cum_4 + 3\{E\{X-E[X]\}^2\}^2$ $= cum_4 + 3cum_2^2 = (-6r_1^4+12r_1^2r_2-3r_2^2-4r_1r_3+r_4)+3(r_2-r_1^2)^2$.
- Excess Kurtosiss of $X = \frac{E\{X-E[X]\}^4}{\left(\sqrt{E\{X-E[X]\}^2}\right)^4}-3 = \frac{cum_4+3cum_2^2}{cum_2^2}-3$ $= \frac{cum_4}{cum_2^2} = \frac{-6r_1^4+12r_1^2r_2-3r_2^2-4r_1r_3+r_4}{(r_2-r_1^2)^2}$.

Table 2.5: Examples of Characteristic Functions

Distribution	$\mathbb{P}(x)$	$\phi(\omega)$
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(X-\mu)^2}{2\sigma^2}\right\}$	$\exp(i\mu\omega-\frac{\sigma^2\omega^2}{2})$
Exponential	ae^{-ax}	$\frac{a}{a-i\omega}$
Gamma	$\frac{b^{-a}e^{-x/b}x^{a-1}}{\Gamma(a)}$	$(1-ib\omega)^{-a}$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	$\exp\left[\lambda(e^{ix}-1)\right]$

Table 3.1 Role of parameter Y

Y	Lévy measure $\ell_{CGMY,X}(x)$		Variation of the process
	monotonically decreasing	total mass $\int_{-\infty}^{\infty} \ell_{CGMY,X}(x)dx$	
$Y < -1$	no	finite	finite
$-1 < Y < 0$	yes	finite	finite
$0 < Y < 1$	yes	infinite	finite
$1 < Y < 2$	yes	infinite	infinite

Note: This table is based on the Table 1 of CGMY (2002).

Table 3.2: Model Comparison

Model	Sample Paths Property	Total mass of Lévy measure $\int_{-\infty}^{\infty} \ell_X(x)dx$	Variation of the process
BS	continuous	0	infinite
MJD	continuous with occasional Jumps	finite and small	infinite
VG	purely discontinuous	∞	finite
NIG	purely discontinuous	∞	infinite
CGMY	purely discontinuous	$0 < \leq \infty$	finite or infinite

Table 3.3: Calibrated Parameters

Parameters		Daily Average	Standard Error	Minimum	Maximum
BS	σ	0.14437	0.02253	0.09714	0.19339
MJD	σ	0.09544	0.00669	0.07470	0.11761
	λ	0.77742	0.59858	0.29551	5.66366
	μ	-0.14899	0.06361	-0.29654	-0.00985
	δ	0.09411	0.02968	0.02681	0.14922
VG	θ	-0.16798	0.03133	-0.36617	-0.10567
	σ	0.13503	0.01758	0.09363	0.16591
	υ	0.39608	0.23908	0.00802	0.89766
NIG	α	20.7408	12.7483	9.60631	89.6984
	β	-11.7309	3.71075	-30.0897	-6.87204
	δ	0.24832	0.11783	0.15429	1.01836
CGMY	C	0.06675	0.05032	0.00259	0.33387
	G	2.35246	0.96472	0.34914	7.63829
	M	262.805	990.01	18.4187	7314.23
	Y	1.19952	0.29492	0.56393	1.77157

Note: The daily average, the standard error, the minimum, and the maximum of the calibrated parameters for the sample period are presented. Note that we set NIG's μ equal to zero because this parameter is redundant.

Table 3.4: In-Sample Fit

Model	Daily Average	Standard Error	Minimum	Maximum
BS	327.314	170.428	0.77675	626.491
MJD	1.98586	2.42562	0.07417	10.1348
VG	4.38282	2.81334	0.13971	14.1986
NIG	1.62785	2.12555	0.06271	8.9489
CGMY	0.82602	1.24139	0.00435	5.5291

Note: The daily average, the standard error, the minimum, and the maximum of the sum of squared errors (SSE) for the sample period is presented.

Table 3.5: Out-of-sample absolute pricing errors per call option

		Moneyness K / F_0				
Days to Expiration	Model	< 0.94	0.94 – 0.98	0.98 – 1.02	1.02 – 1.06	> 1.06
≥ 180	BS	\$11.14	\$5.568	\$2.493	\$2.527	\$5.861
	MJD	0.802	1.016	1.056	0.958	0.626
	VG	0.881	1.065	1.173	1.006	0.696
	NIG	0.788	0.965	1.064	0.967	0.609
	CGMY	0.799	0.938	1.047	0.966	0.609
60 – 180	BS	5.278	3.598	1.365	2.047	2.774
	MJD	0.575	0.514	0.602	0.586	0.347
	VG	0.694	0.613	0.685	0.649	0.436
	NIG	0.504	0.537	0.612	0.589	0.348
	CGMY	0.420	0.477	0.598	0.579	0.337
< 60	BS	1.258	1.703	0.794	1.264	0.615
	MJD	0.417	0.399	0.507	0.282	0.140
	VG	0.493	0.386	0.516	0.300	0.165
	NIG	0.400	0.379	0.506	0.271	0.120
	CGMY	0.265	0.374	0.503	0.284	0.188

Note: Today's model price is computed using the yesterday's implied parameters for each call option. We define the absolute pricing error per option as the sample average of the absolute difference between the market price and the model price. Note that the entire sample is divided according to its maturity and its moneyness. The long term is defined as more than or equal to 180 days to maturity and the short term is less than 60 days to maturity. We define the deep ITM as the moneyness less than 0.94, the moneyness between 0.94 and 0.98 is ITM, the moneyness between 0.98 and 1.02 is ATM, the moneyness between 1.02 and 1.06 is OTM, and the moneyness greater than 1.06 is deep OTM.

Table 3.6: Orthogonality test results of out-of-sample dollar pricing errors

Independent variable	BS	MJD	VG	NIG	CGMY
β_0	-15.227	-10.864	-11.726	-10.353	-6.602
(1)	(-6.744)*	(-18.929)*	(-19.050)*	(-18.262)*	(-11.819)*
β_1	-21.898	17.529	17.700	16.672	10.959
(K_i / F_0)	(-5.852)*	(18.427)*	(17.345)*	(17.744)*	(11.837)*
β_2	23.439	-8.351	-7.879	-7.863	-5.496
$((K_i / F_0)^2)$	(12.481)*	(-17.492)*	(-15.388)*	(-16.673)*	(-11.827)*
β_3	7.415	1.436	1.481	1.295	1.022
(T_i)	(7.445)*	(5.671)*	(5.453)*	(5.176)*	(4.147)*
β_4	522.964	63.646	72.040	58.544	43.140
(r_i)	(10.462)*	(5.009)*	(5.286)*	(4.664)*	(3.488)*
Adjusted R^2	0.472	0.069	0.111	0.069	0.024
Observations	6567	6567	6567	6567	6567
F statistic$_{(4, 6567)}$	1464.14	122.477	204.886	122.179	41.776

Note: Out-of-sample dollar pricing errors ε_i are regressed on the moneyness K_i / F_0, the moneyness squared $(K_i / F_0)^2$, the time to maturity T_i, and the interest rate r_i:

$$\varepsilon_i = \beta_0 + \beta_1 (K_i / F_0) + \beta_2 (K_i / F_0)^2 + \beta_3 T_i + \beta_4 r_i + e_i .$$

t-statistics are in parentheses. * indicates the significance at 5% level.

Table 4.1: Statistically estimated parameters of each model

Model	Log likelihood	Model parameter vector $\theta_{\mathbb{P}}$		
BS	883.466	$\sigma_{\mathbb{P}} = 0.107419$		
MJD	886.568	$\sigma_{\mathbb{P}} = 0.080850$	$\lambda_{\mathbb{P}} = 60.00000$	$\mu_{\mathbb{P}} = -0.010476$
		$\delta_{\mathbb{P}} = 1.600779 \times 10^{-9}$		

Table 4.2: Annualized standardized moments of statistical log return probability density

Model	Standard Deviation	Skewness	Excess Kurtosis
True	0.107640	-0.016440	-0.000347
BS	0.107419	0	0
MJD	0.114548	-0.045892	0.004197

Table 4.3: Risk-neutral prior parameters

Parameters		Average	Standard Error	Minimum	Maximum
MJD	σ	0.09544	0.00669	0.07470	0.11761
	λ	0.77742	0.59858	0.29551	5.66366
	μ	-0.14899	0.06361	-0.29654	-0.00985
	δ	0.09411	0.02968	0.02681	0.14922
BS	σ	0.14437	0.02253	0.09714	0.19339

Note: The average, the standard error, the minimum, and the maximum of the successively updated risk-neutral parameters for the sample period are presented.

Table 4.4: Annualized standardized moments of risk-neutral log return probability density

Model	Standard Deviation	Skewness	Excess Kurtosis
BS	0.14437	0	0
MJD	0.18235	-0.931611	1.34135

Table 4.5: Annualized standardized moments of statistical vs. risk-neutral prior log return probability density

Model	Standard Deviation	Skewness	Excess Kurtosis
Statistical	0.114548	-0.045892	0.004197
Risk-neutral	0.18235	-0.931611	1.34135

Table 4.6: Calibration Results

Methods	$\sigma^{\mathbb{Q}}$	$\lambda^{\mathbb{Q}}$	$\mu^{\mathbb{Q}}$	$\delta^{\mathbb{Q}}$
March 30, 2004 (245 days to maturity)				
M1	0.108775	0.307107	-0.267493	0.140028
M2				
M3	0.108995	0.300333	-0.272398	0.136964
July 1, 2004 (180 days to maturity)				
M1	0.093676	0.485600	-0.197655	0.111576
M2				
M3	0.092968	0.509876	-0.190205	0.114493
September 27, 2004 (120 days to maturity)				
M1	0.095187	0.732404	-0.144411	0.082790
M2				
M3	0.095463	0.714316	-0.147262	0.081582
November 22, 2004 (80 days to maturity)				
M1	0.094648	0.662536	-0.107062	0.080695
M2				
M3	0.095101	0.630734	-0.111636	0.079383
January 20, 2005 (40 days to maturity)				
M1	0.091795	1.296250	-0.070378	0.056700
M2				
M3	0.091314	1.259500	-0.075242	0.052736
March 4, 2005 (10 days to maturity)				
M1	0.086929	1.009830	-0.054321	0.043944
M2				
M3	0.088270	0.736562	-0.067053	0.046267

Table 5.1: Moments of RND estimates with 9-month to maturity

RND Estimator	Mean	Standard Deviation	Skewness	Kurtosis
VI	864.779	183.395	-0.1901	3.0095
2LN	863.474	185.570	-0.1834	2.9894
BS	863.400	127.000	0.4445	3.3533

Note: Moments of RND recovered from volatility interpolation (VI), mixture of two lognormals (2LN), and the benchmark Black-Scholes lognormal (BS) methods. This table quantifies the difference of RND estimates among three methods with respect to their first four moments. The estimates are based on CME settlement prices of futures options on S&P 500 index with December 2003 maturity on March 18, 2003.

Table 5.2: Moments of RND estimates with 6-month to maturity

RND Estimator	Mean	Standard Deviation	Skewness	Kurtosis
VI	1006.465	137.243	-0.5682	3.5095
2LN	1006.305	136.252	-0.4972	3.5004
BS	1006.400	120.654	0.3614	3.2331

Note: Moments of RND recovered from volatility interpolation (VI), mixture of two lognormals (2LN), and the benchmark Black-Scholes lognormal (BS) methods. This table quantifies the difference of RND estimates among three methods with respect to their first four moments. The estimates are based on CME settlement prices of futures options on S&P 500 index with December 2003 maturity on June 18, 2003.

Table 5.3: Moments of RND estimates with 3-month to maturity

RND Estimator	Mean	Standard Deviation	Skewness	Kurtosis
VI	1037.345	100.405	-0.7738	5.2039
2LN	1037.014	98.800	-0.7555	4.4533
BS	1037.300	88.463	0.2565	3.1172

Note: Moments of RND recovered from volatility interpolation (VI), mixture of two lognormals (2LN), and the benchmark Black-Scholes lognormal (BS) methods. This table quantifies the difference of RND estimates among three methods with respect to their first four moments. The estimates are based on CME settlement prices of futures options on S&P 500 index with December 2003 maturity on September 18, 2003.

Table 5.4: Moments of RND estimates with 1-month to maturity

RND Estimator	Mean	Standard Deviation	Skewness	Kurtosis
VI	1033.045	56.118	-0.7328	5.0895
2LN	1032.409	55.536	-1.0367	6.4456
BS	1032.800	51.181	0.1488	3.0394

Note: Moments of RND recovered from volatility interpolation (VI), mixture of two lognormals (2LN), and the benchmark Black-Scholes lognormal (BS) methods. This table quantifies the difference of RND estimates among three methods with respect to their first four moments. The estimates are based on CME settlement prices of futures options on S&P 500 index with December 2003 maturity on November 18, 2003.

Figure 1.1: Examples of a function f: a → f(a)

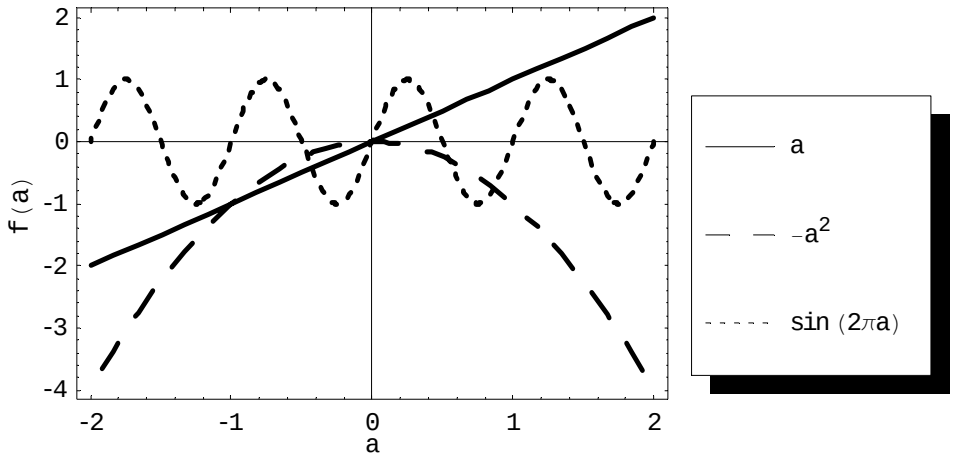


Figure 1.2: Plot of the probability density function of an exponential random variable with λ = 0.01

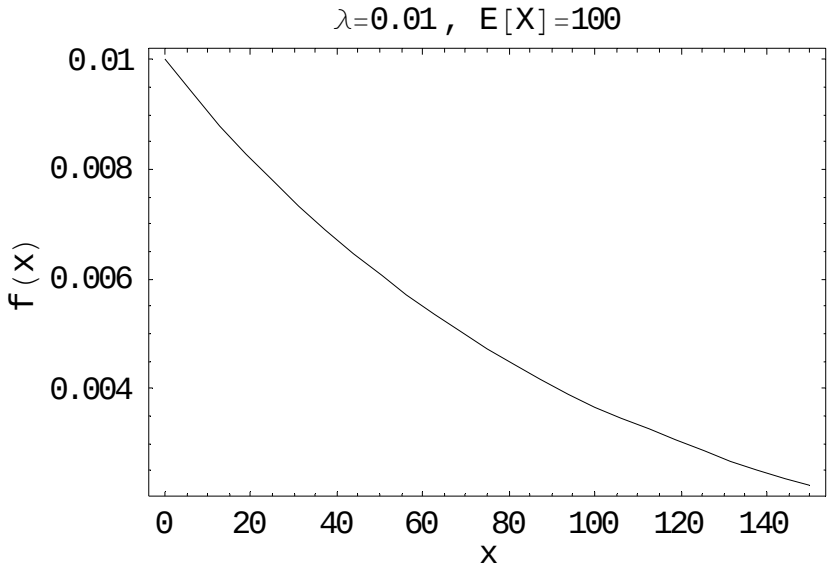


Figure 1.3: Zero arrival of an event in [0, t]

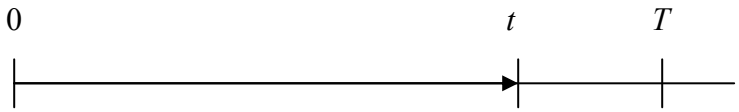


Figure 1.4: One arrival of an event in [0, t]

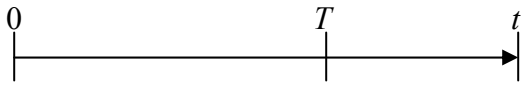


Figure 1.5: k arrivals of an event in [0, t]

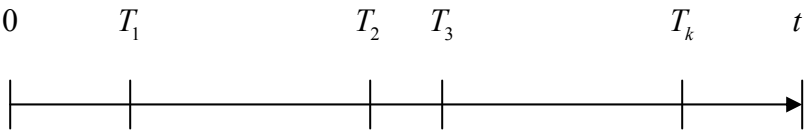


Figure 1.6: An event arrivals k times in the interval [0, t]

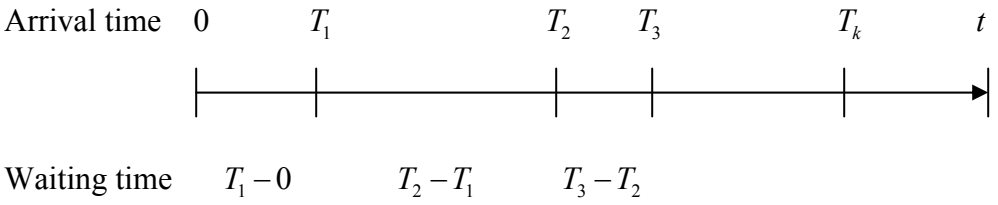


Figure 1.7: Simulated sample paths of Poisson processes with the intensity $\lambda = 5$ and $0 \leq t \leq 2$

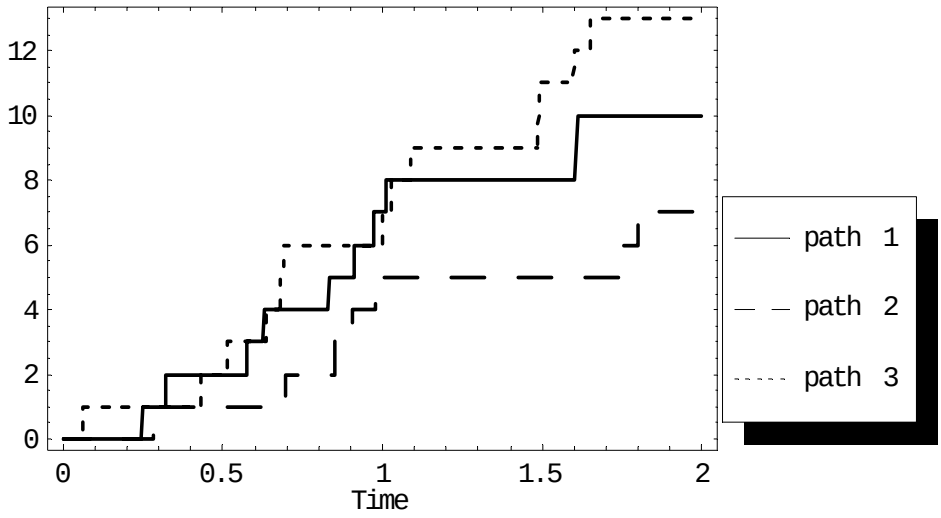


Figure 1.8: Simulated sample paths of a compound Poisson process with the intensity $\lambda = 5$ and $0 \leq t \leq 2$

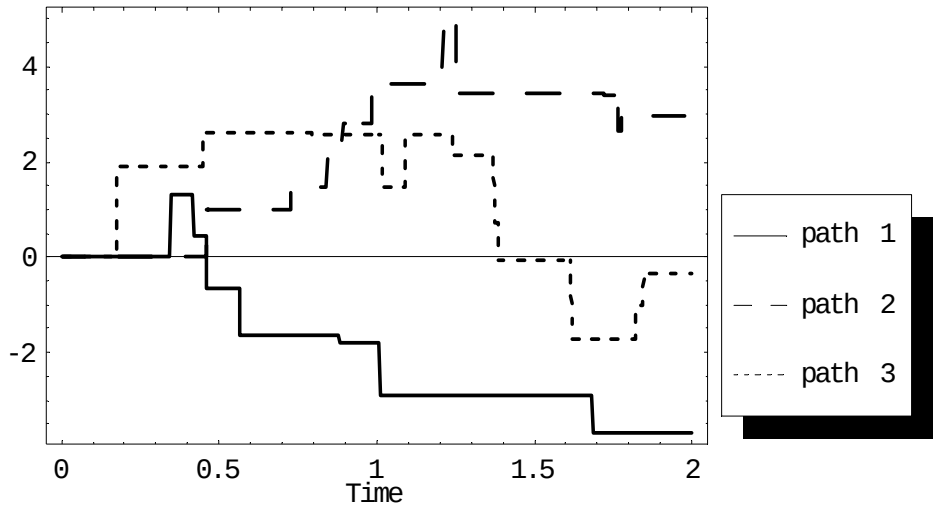


Figure 2.1: Illustration of wavelength λ

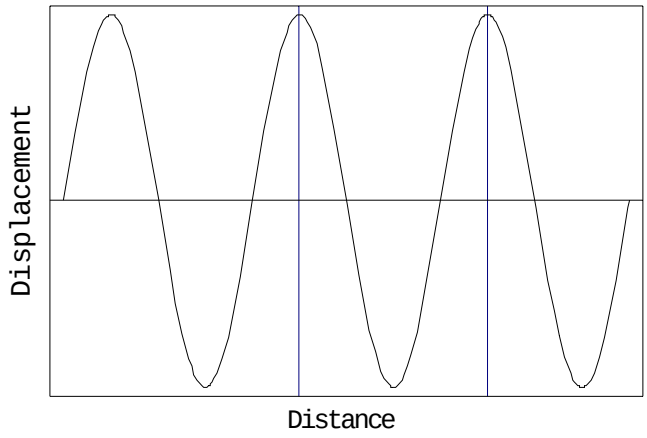


Figure 2.2: Plot of 5 Hz sine wave g(t) = sin(2π(5)t)

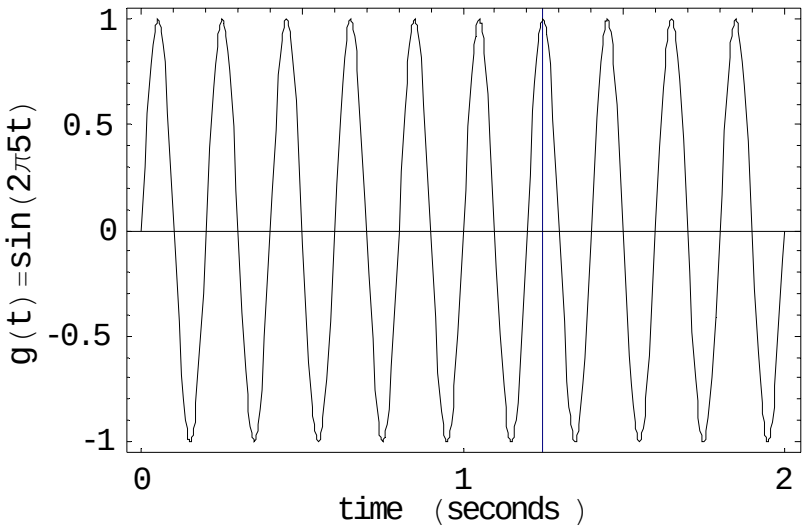


Figure 2.3: The definition of sine and cosine function with unit circle

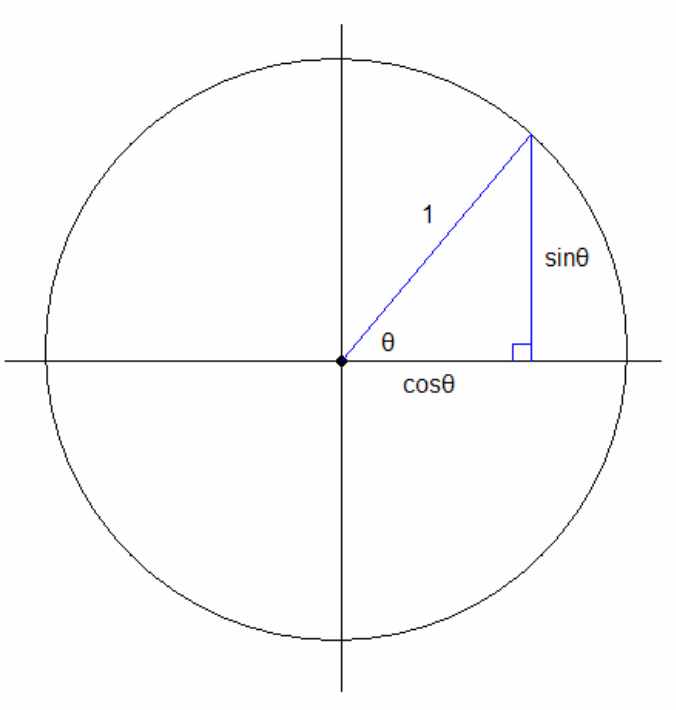


Figure 2.4: Sine and cosine function sin(θ) and cos(θ)

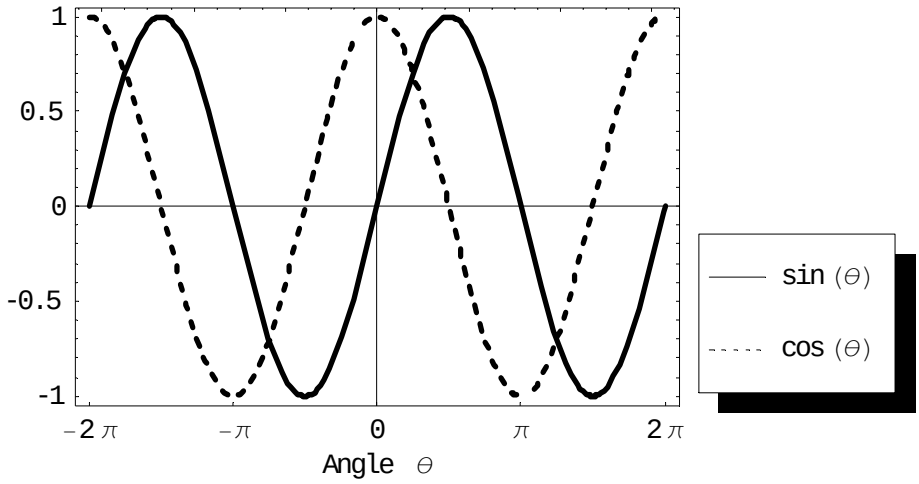


Figure 2.5: Plot of a sine wave g(t) = sin(2πft) with different fundamental frequency f_0

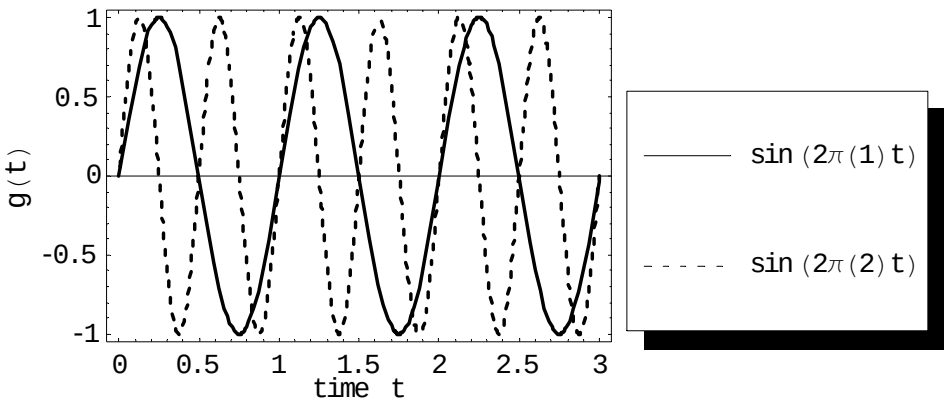


Figure 2.6: Plot of a 1 Hz sine wave g(t) =a sin(2πt) with different amplitudes a = 1/2, 1, and 2

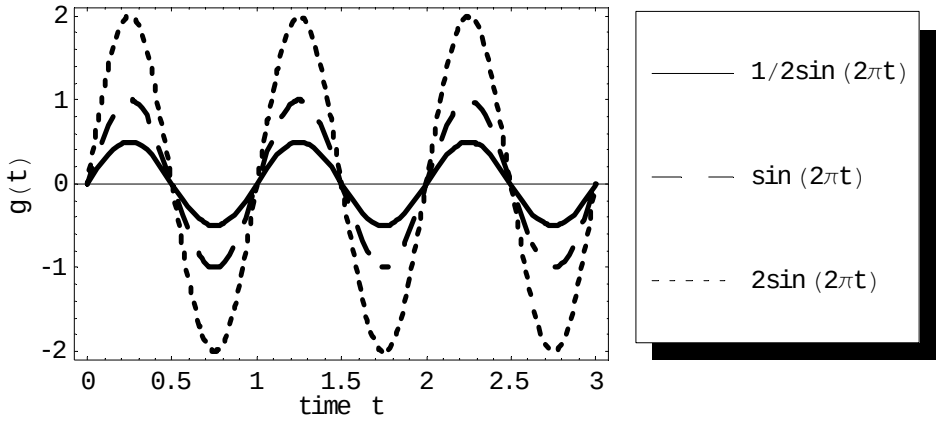


Figure 2.7: Plot of a 1 Hz Sine Wave g(t) = sin(2πt + b) with Different Phase b = 0, π/2, and - π/2

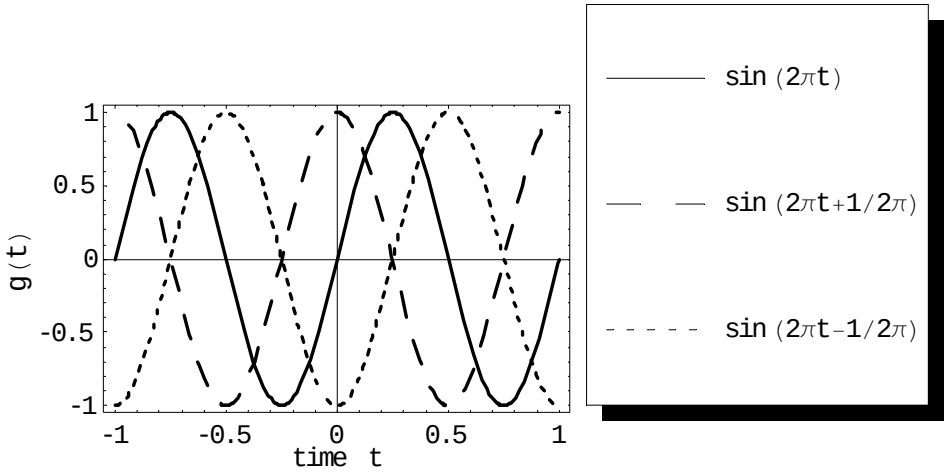
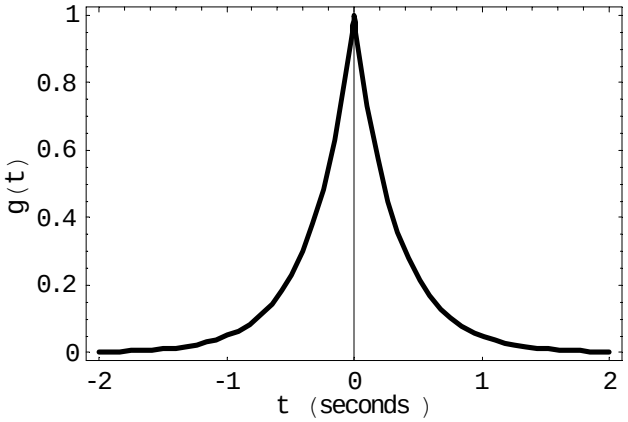
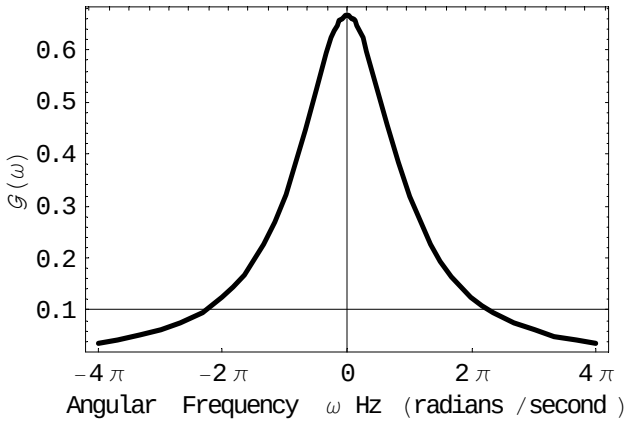


Figure 2.8: Plot of double-sided exponential function and its Fourier transforms

A) Plot of a double-sided exponential function $g(t) = e^{-3|t|}$.



B) Plot of FT of $g(t) = e^{-3|t|}$ in Angular Frequency Domain, $\mathcal{G}(\omega)$.



C) Plot of FT of $g(t) = e^{-3|t|}$ in Frequency Domain, $\mathcal{G}(f)$.

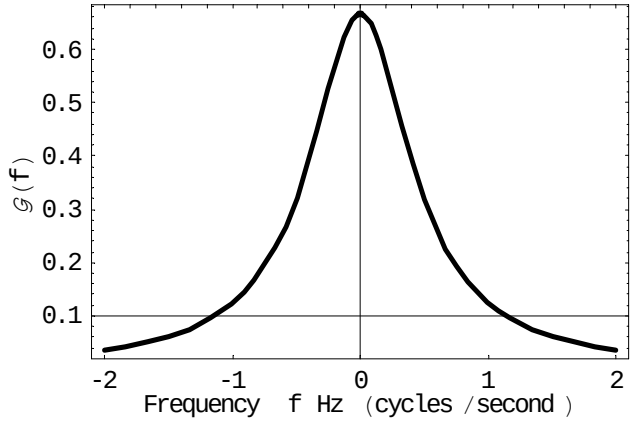
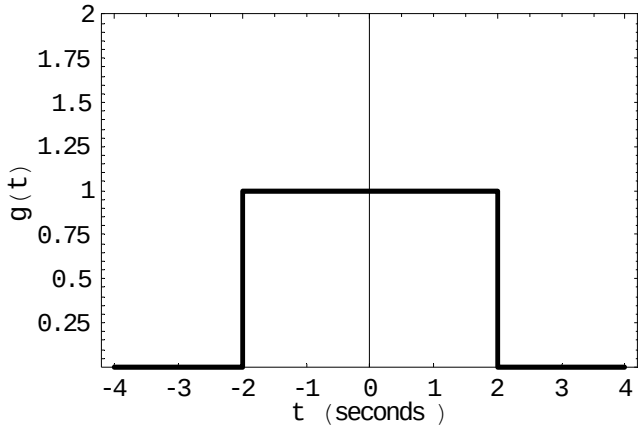
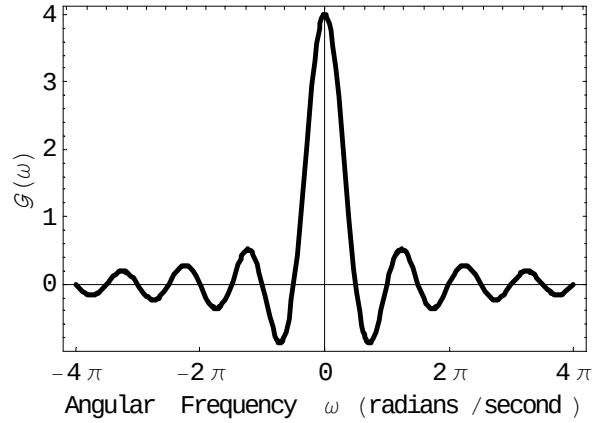


Figure 2.9: Plot of rectangular pulse and its Fourier transforms

A) Plot of a rectangular pulse $g(t)$.



B) Plot of the FT of $g(t)$ in Angular Frequency Domain, $\mathcal{G}(\omega)$.



C) Plot of the FT of $g(t)$ in Frequency Domain, $\mathcal{G}(f)$.

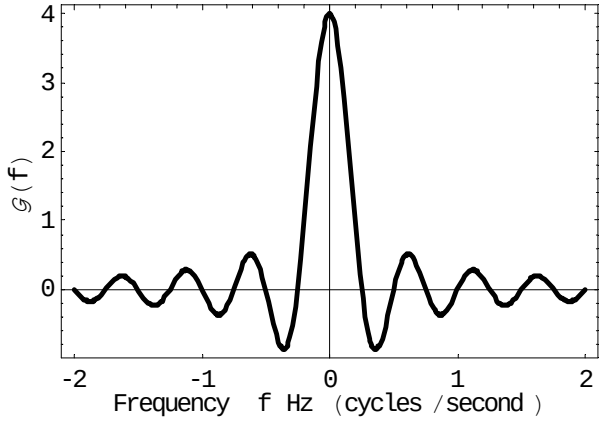
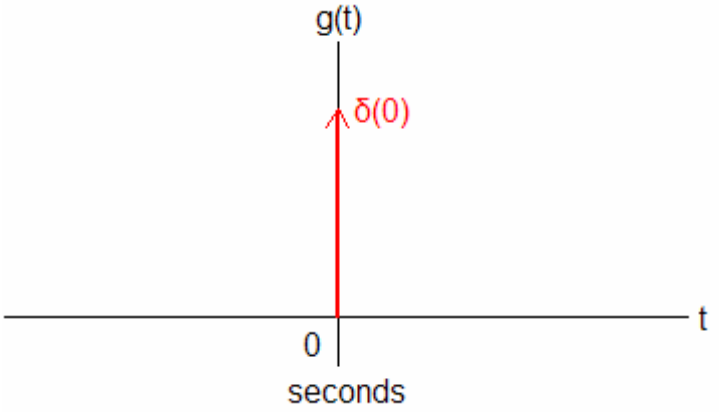
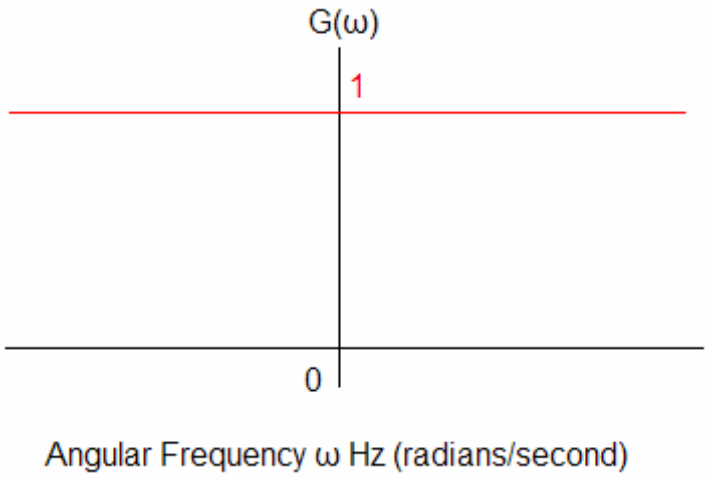


Figure 2.10: Plot of Dirac's delta function and its Fourier transforms

A) Plot of $g(t) = \delta(t)$.



B) Plot of FT of $g(t)$ in Angular Frequency Domain, $\mathcal{G}(\omega) = 1$.



C) Plot of FT of $g(t)$ in Frequency Domain, $\mathcal{G}(f) = 1$.

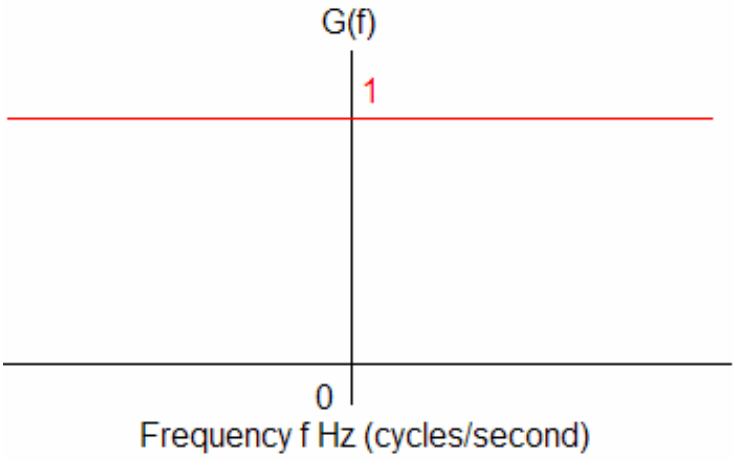
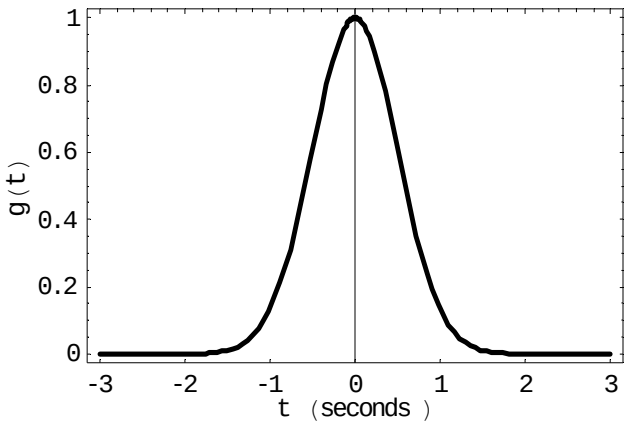
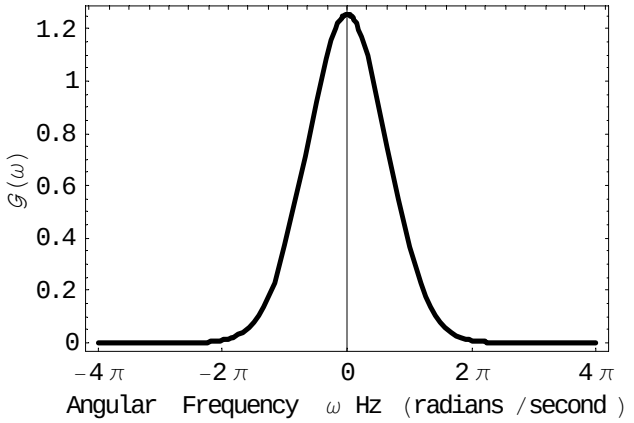


Figure 2.11: Plot of Gaussian function and Fourier transforms

A) Plot of Gaussian function $g(t) = e^{-2t^2}$.



B) Plot of FT of $g(t)$ in Angular Frequency Domain, $\mathcal{G}(\omega)$.



C) Plot of FT of $g(t)$ in Frequency Domain, $\mathcal{G}(f)$.

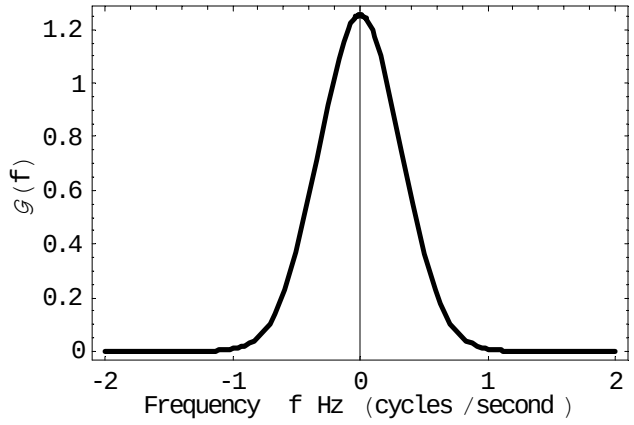
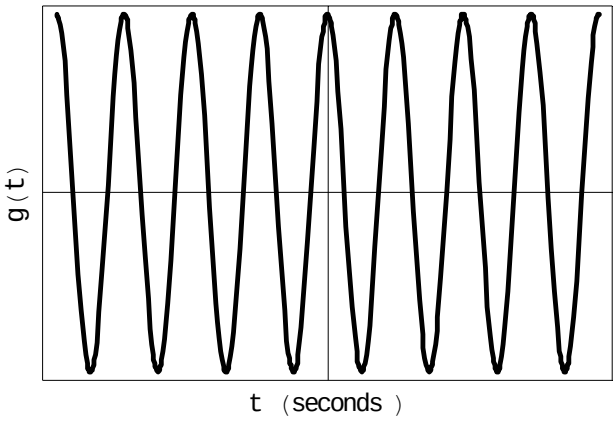
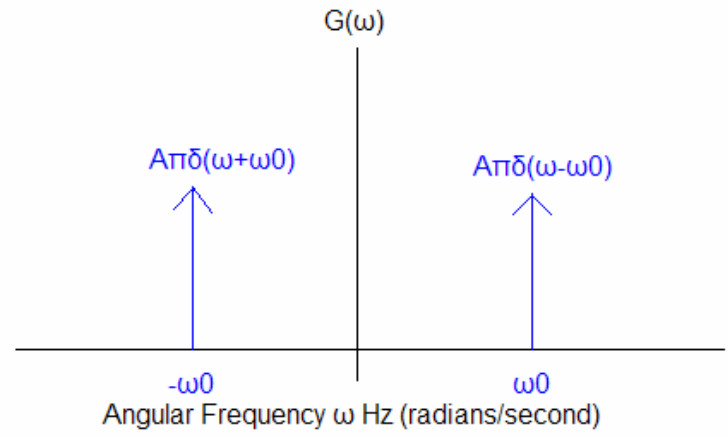


Figure 2.12: Plot of a cosine wave and Fourier transforms

A) Plot of a Cosine Wave $g(t) = A\cos(2\pi f_0 t) = A\cos(\omega_0 t)$. Amplitude of the wave is given by A.



B) Plot of FT of $g(t)$ in Angular Frequency Domain, $\mathcal{G}(\omega)$.



C) Plot of FT of $g(t)$ in Frequency Domain, $\mathcal{G}(f)$.

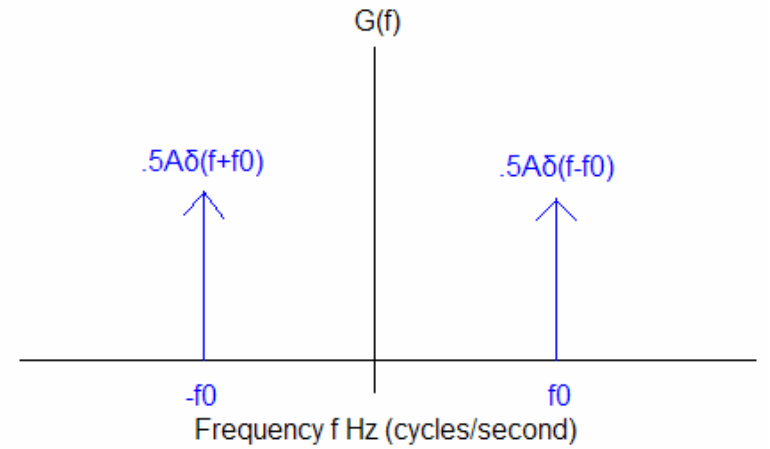
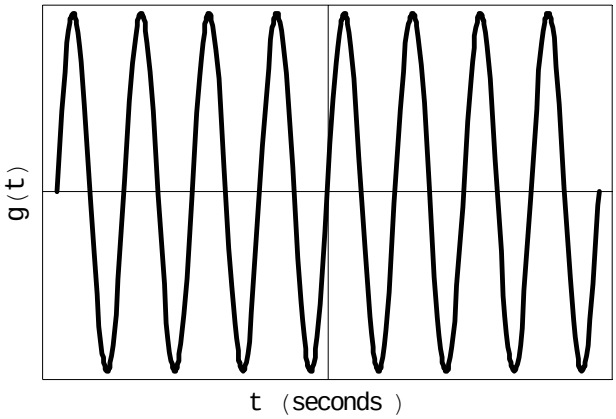
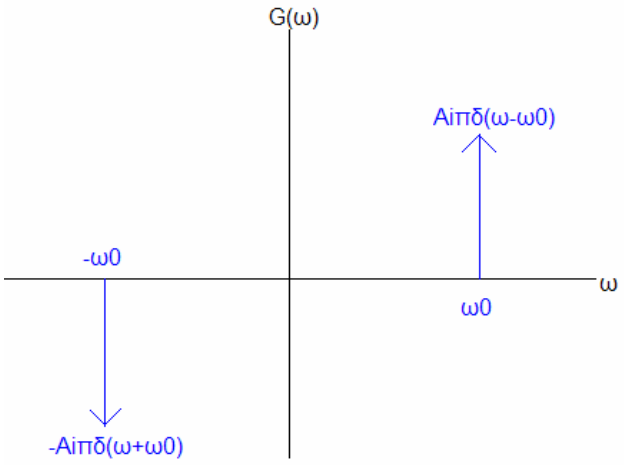


Figure 2.13: Plot of a sine wave and Fourier transforms

A) Plot of a Sine Wave $g(t) = A\sin(2\pi f_0 t) = A\sin(\omega_0 t)$.



B) Plot of FT of $g(t)$ in Angular Frequency Domain, $\mathcal{G}(\omega)$.



C) Plot of FT of $g(t)$ in Frequency Domain, $\mathcal{G}(f)$.

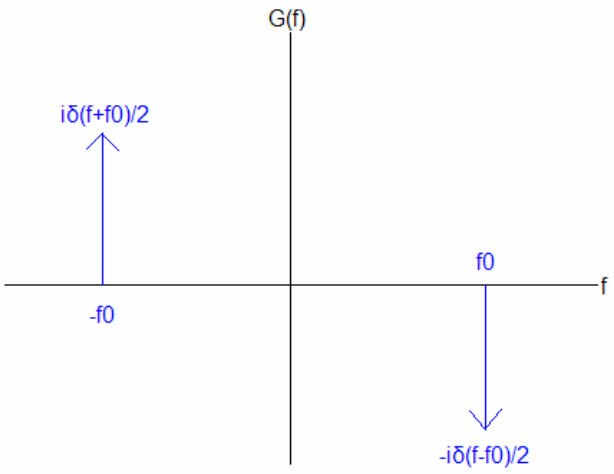


Figure 2.14: Plot of a function h(x) for n = 1, n = $10^{1/2}$, and n = 10

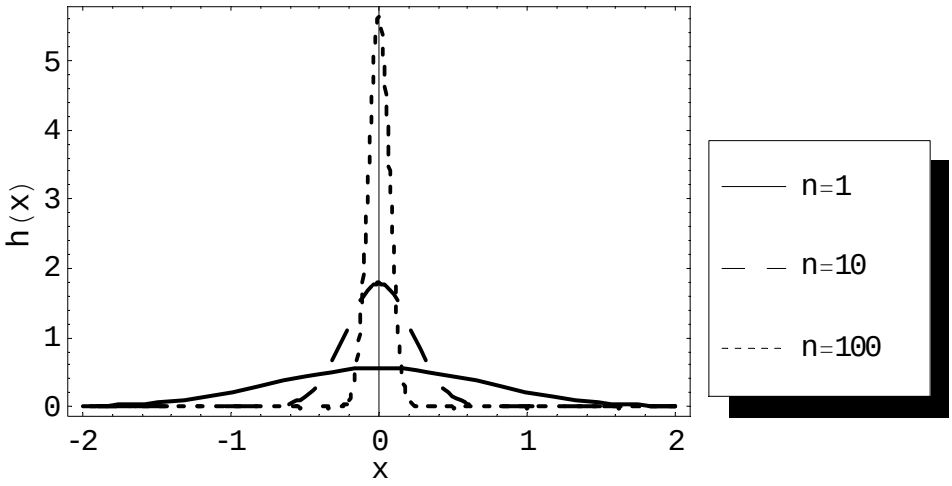
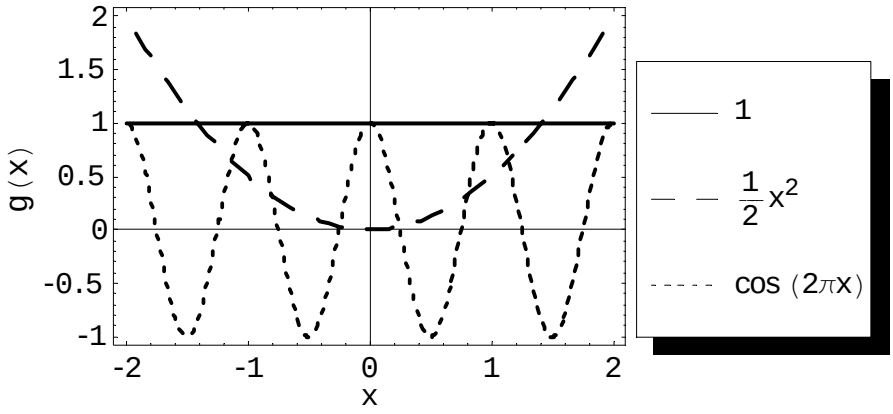


Figure 2.15: Plot of even and odd functions

A) Even Functions: *g(x) = 1, g(x) = 0.5x^2, and g(x) = cos(2πx)*.



B) Odd Functions: *g(x) = x, g(x) = 0.5x^3, and g(x) = sin(2πx)*.

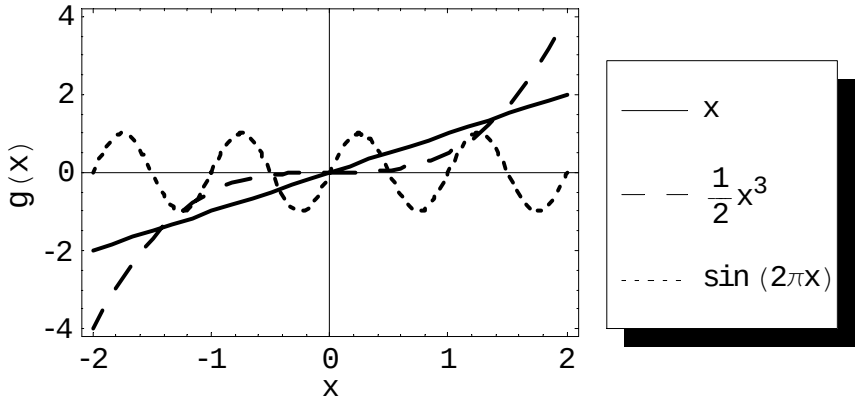


Figure 2.16: Plot of two Gaussian functions f and g and their convolution f*g

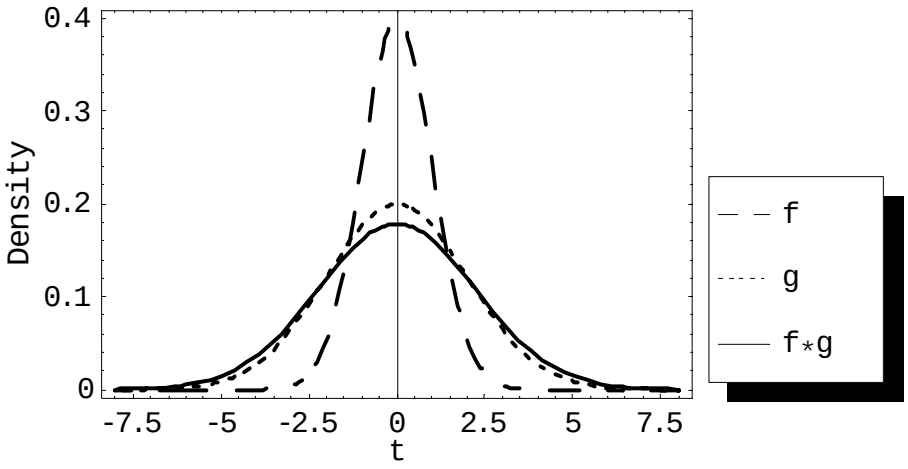


Figure 2.17: Relationship between a strike price K and a log-strike price k ≡ lnK

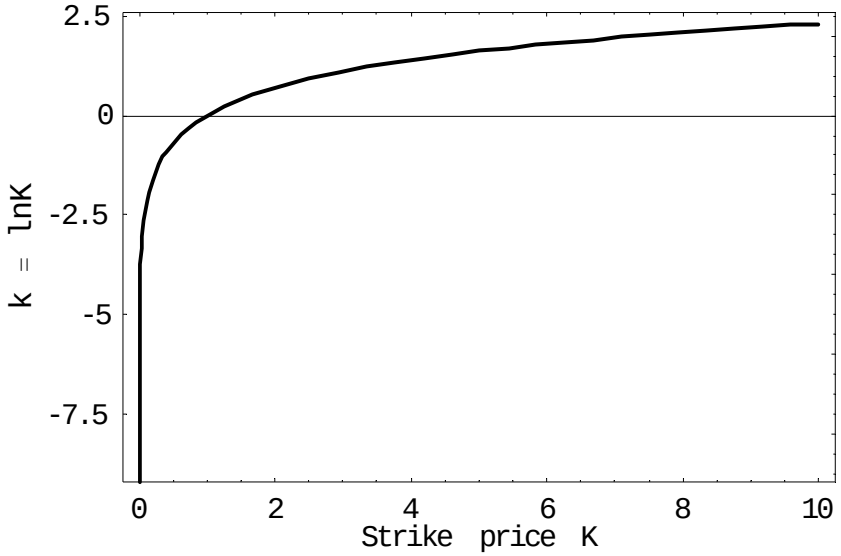


Figure 2.18: Plot of an exponential function $e^{0.5k}$ with respect to log-strike price $k \equiv \ln K$

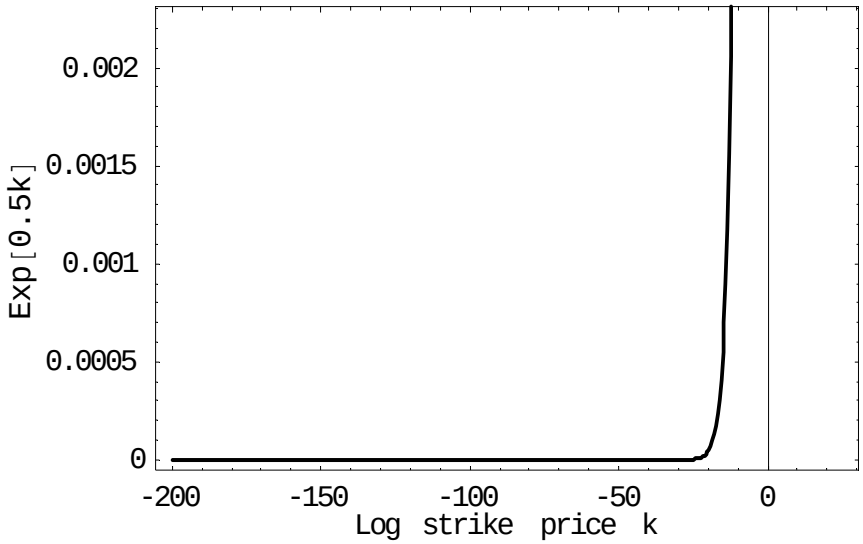
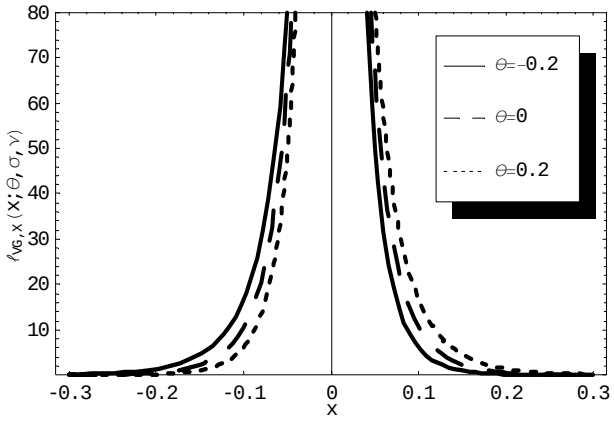
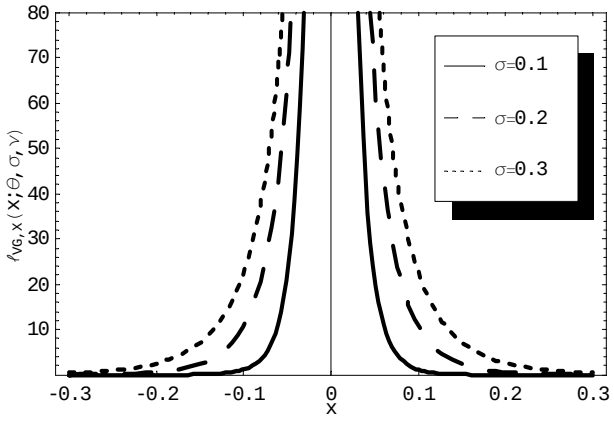


Figure 3.1: Plot of Lévy measure $l_{VG, X}(x; \theta,\sigma,\upsilon)$ of VG process

A) Lévy measure of VG process with various values for θ. Parameters fixed are $\sigma = 0.2$ and $\upsilon = 0.1$.



B) Lévy measure of VG process with various values for σ. Parameters fixed are $\theta = 0$ and $\upsilon = 0.1$.



C) Lévy measure of VG process with various values for υ. Parameters fixed are $\theta = 0$ and $\sigma = 0.2$.

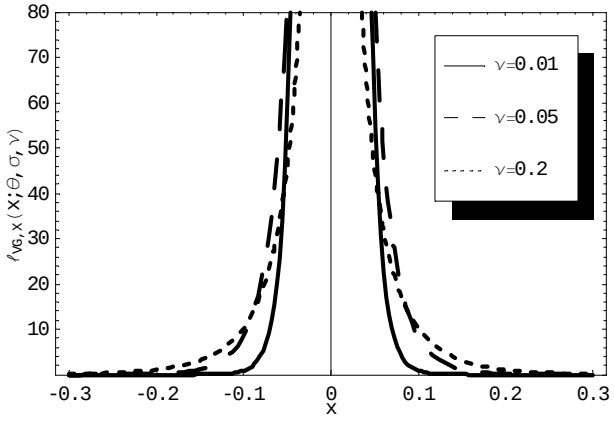
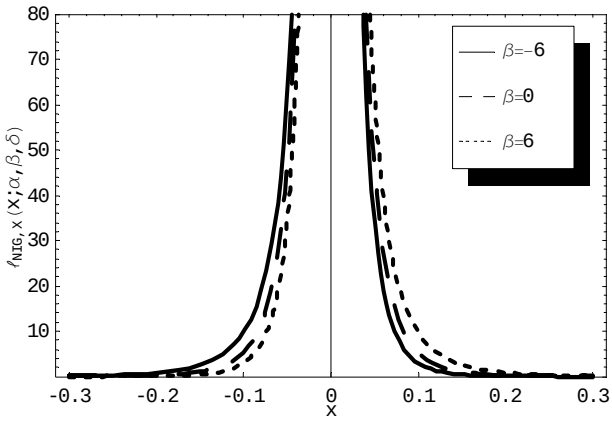
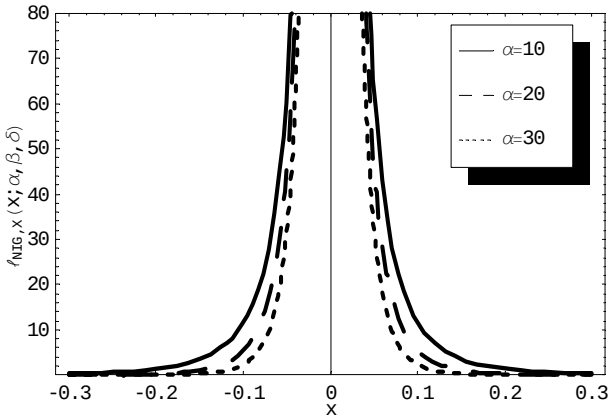


Figure 3.2: Plot of Lévy measure $l_{NIG,X}(x; \alpha,\beta,\delta)$ of NIG process

A) Lévy measure of NIG process with various values for β. Parameters fixed are $\alpha = 20$ and $\delta = 0.6$.



B) Lévy measure of NIG process with various values for α. Parameters fixed are $\beta = 0$ and $\delta = 0.6$.



C) Lévy measure of NIG process with various values for δ. Parameters fixed are $\alpha = 20$ and $\beta = 0$.

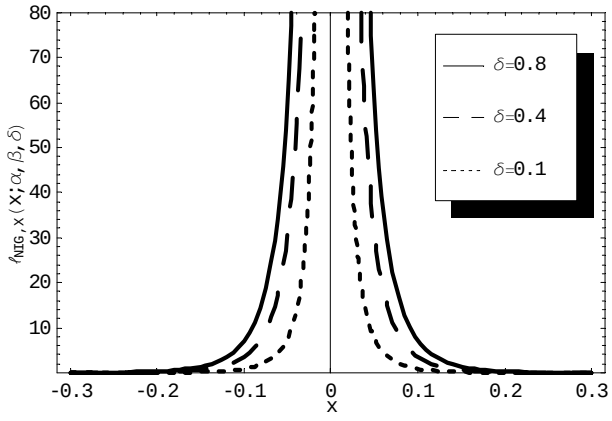
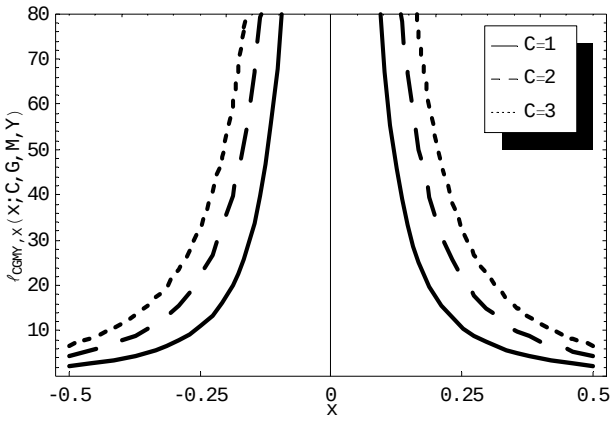


Figure 3.3: Plot of Lévy measure $l_{CGMY,X}(x; C, G, M, Y)$ of CGMY process

A) Lévy measure of CGMY process with various values for *C.* Parameters fixed are *G = 1, M = 1,* and *Y = 0.9*.



B) Lévy measure of CGMY process with various values for *G* and *M*. Parameters fixed are *C = 1* and *Y = 0.9*.

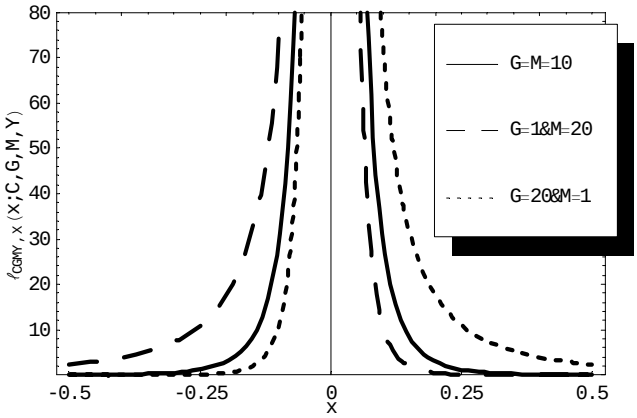
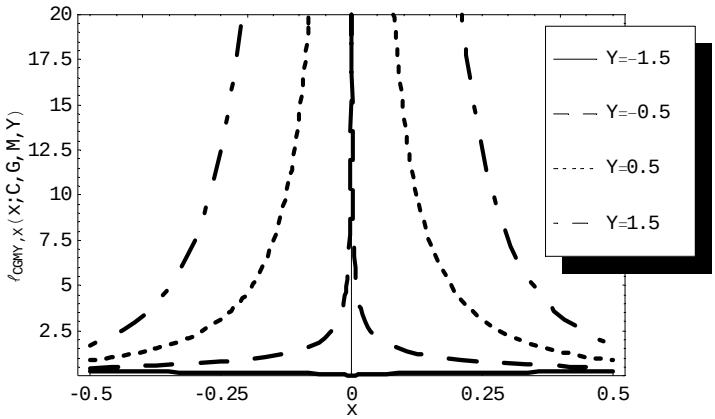


Figure 3.4: Plot of Lévy measure $l_{CGMY,X}(x; C, G, M, Y)$ of CGMY process

C) Lévy measure of CGMY process with various values for *Y*. Parameters fixed are *C = 0.5, G = 1,* and *M = 1*.



D) Lévy measure of CGMY process with $Y = -1.5$ Parameters fixed are $C = 0.5$, $G = 1$, and $M = 1$.

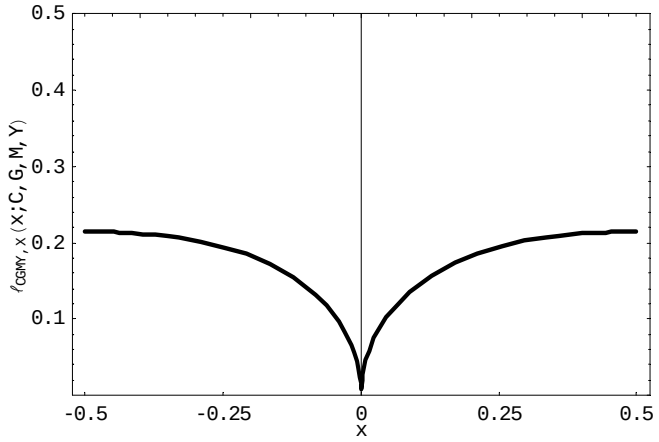
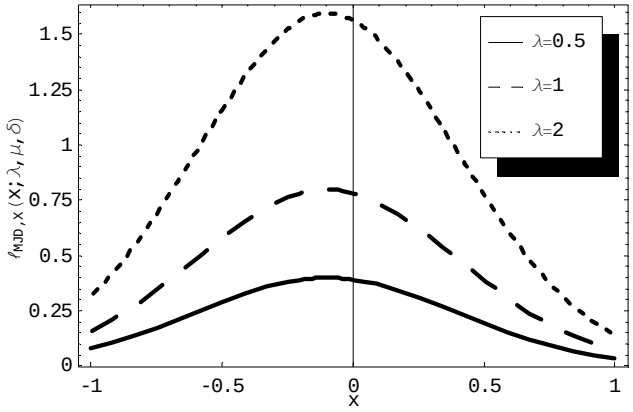
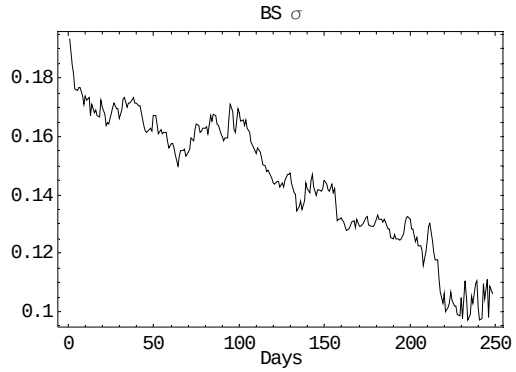


Figure 3.5: Plot of Lévy measure $l_{MJD,X}(x; \lambda, \mu, \delta)$ of MJD process with various values for λ



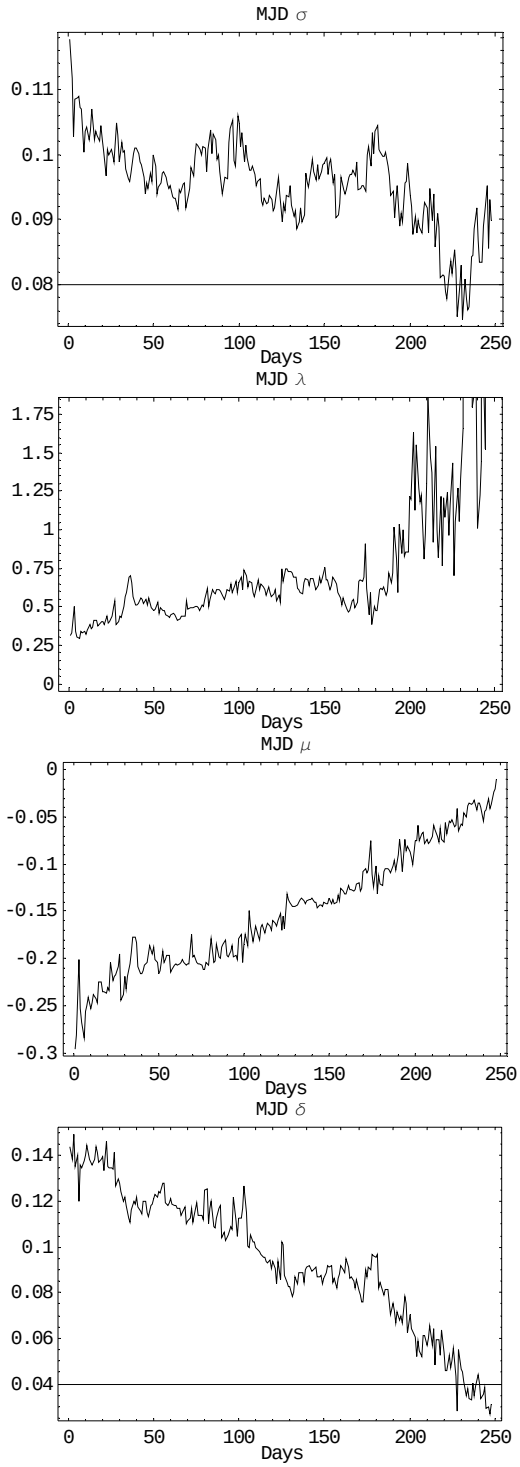
Note: Parameters fixed are $\mu = -0.1$ and $\delta = 0.5$.

Figure 3.6: Dynamics of calibrated parameters: σ of Black-Scholes model



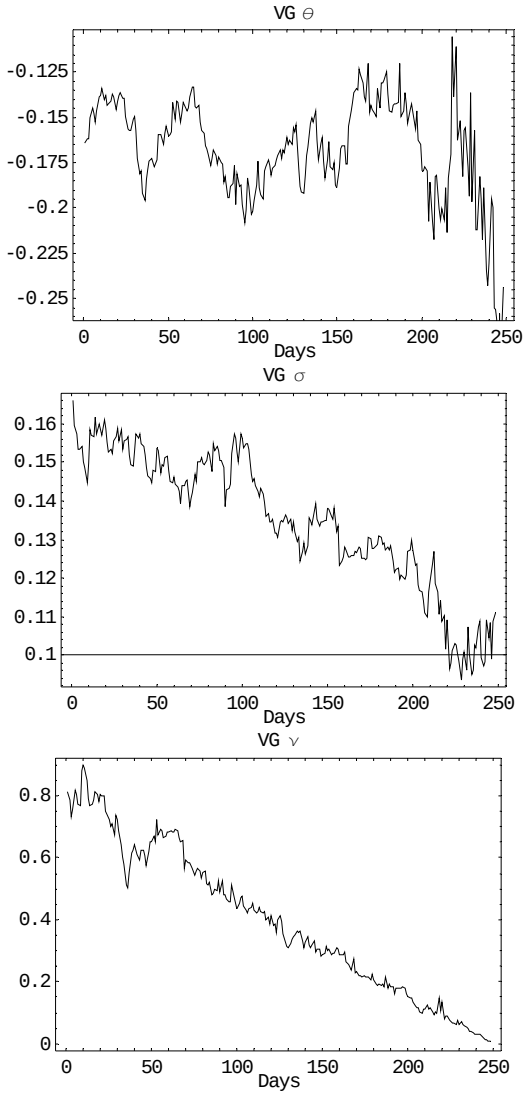
Note: Time series of calibrated parameters for BS model are plotted from the day 1 which is March 24, 2004, through day 248 which is March 16, 2005.

Figure 3.7: Dynamics of calibrated parameters: σ, λ, μ, δ of Merton jump-diffusion model



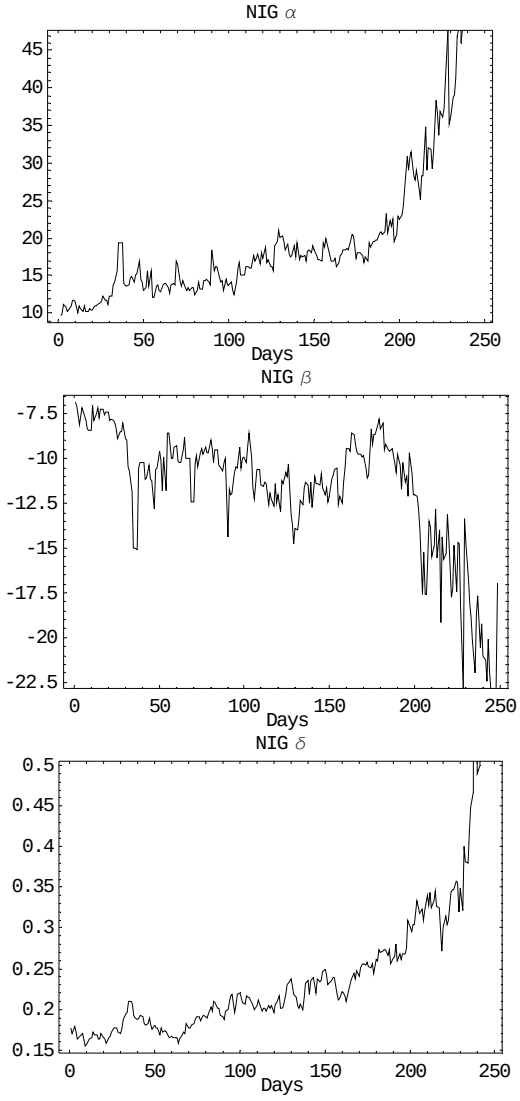
Note: Time series of calibrated parameters for MJD model are plotted from the day 1 which is March 24, 2004, through day 248 which is March 16, 2005.

Figure 3.8: Dynamics of calibrated parameters: θ, σ, υ of Variance Gamma model



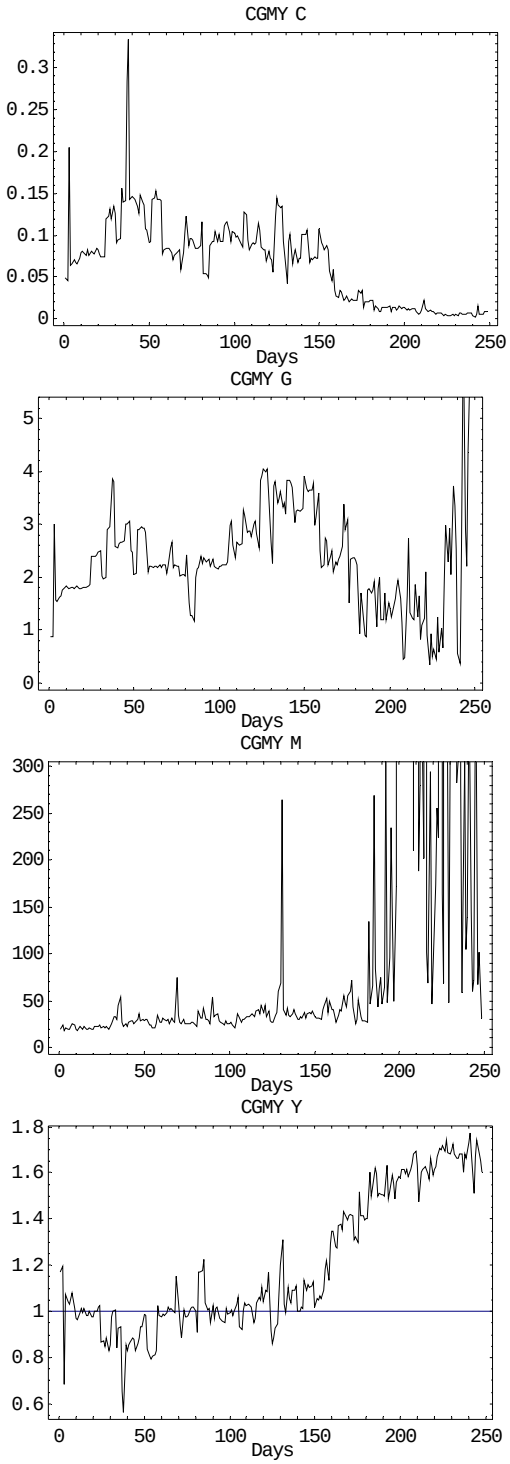
Note: Time series of calibrated parameters for VG model are plotted from the day 1 which is March 24, 2004, through day 248 which is March 16, 2005.

Figure 3.9: Dynamics of calibrated parameters: α, β, δ of Normal Inverse Gaussian model



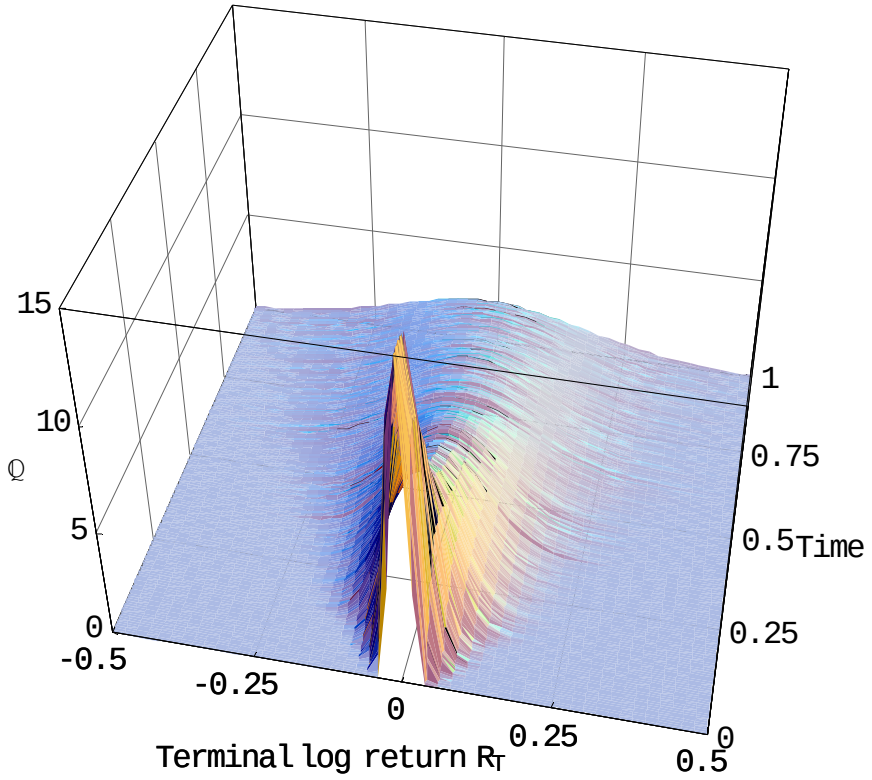
Note: Time series of calibrated parameters for NIG model are plotted from the day 1 which is March 24, 2004, through day 248 which is March 16, 2005.

Figure 3.10: Dynamics of calibrated parameters: C, G, M, Y of CGMY model



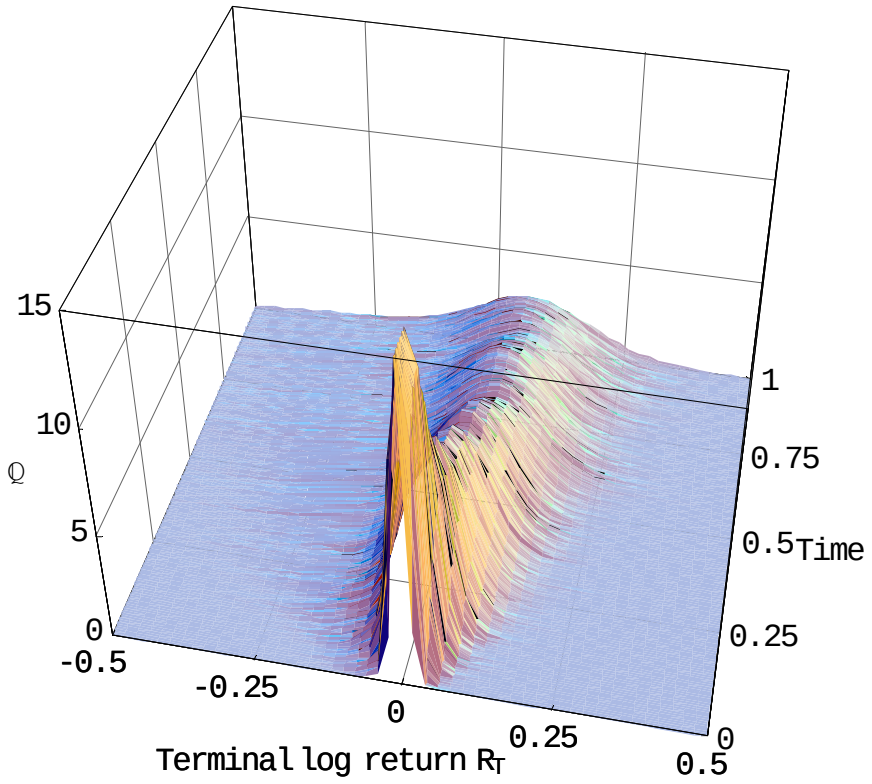
Note: Time series of calibrated parameters for CGMY model are plotted from the day 1 which is March 24, 2004, through day 248 which is March 16, 2005.

Figure 3.11: Black-Scholes model implied dynamics of log-return probability density



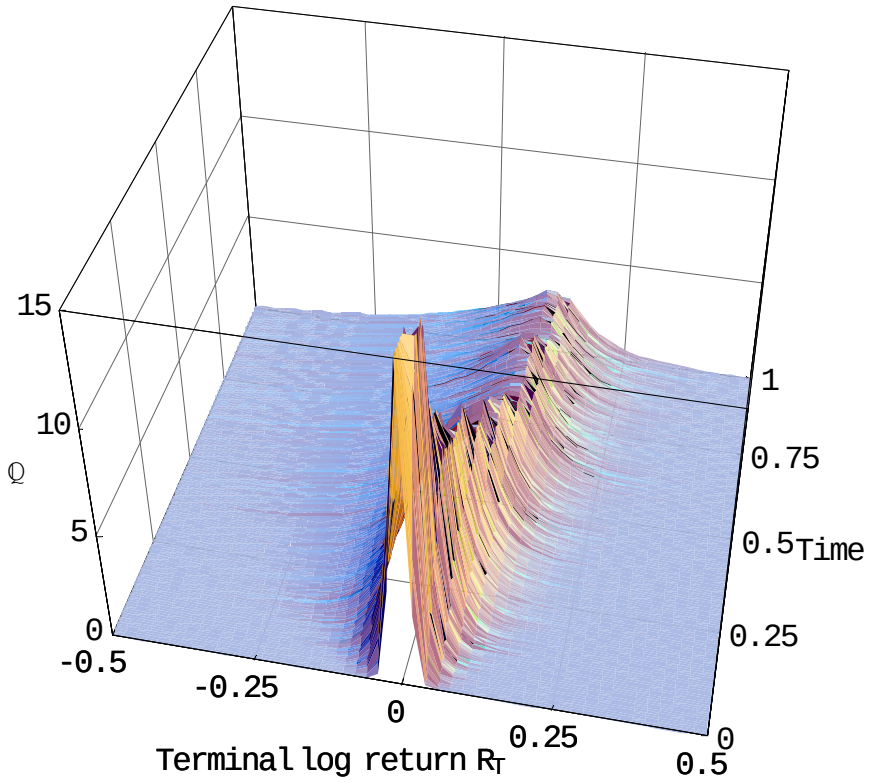
Note: Calibrated probability density functions of log returns in BS model on one day before last trading day March 16, 2005 are plotted as the time to maturity nears from 1 to 0. Note that the above figures illustrate density functions only for the range of probability density between 0 and 15.

Figure 3.12: MJD model implied dynamics of log-return probability density



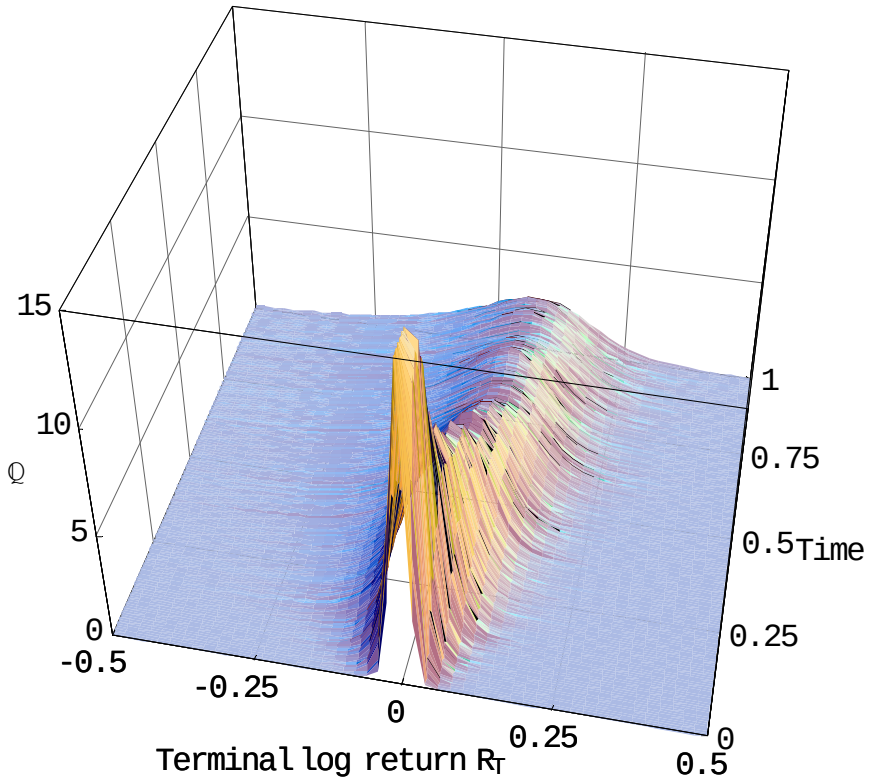
Note: Calibrated probability density functions of log returns in MJD model on one day before last trading day March 16, 2005 are plotted as the time to maturity nears from 1 to 0. Note that the above figures illustrate density functions only for the range of probability density between 0 and 15.

Figure 3.13: VG model implied dynamics of log-return probability density



Note: Calibrated probability density functions of log returns in VG model on one day before last trading day March 16, 2005 are plotted as the time to maturity nears from 1 to 0. Note that the above figures illustrate density functions only for the range of probability density between 0 and 15.

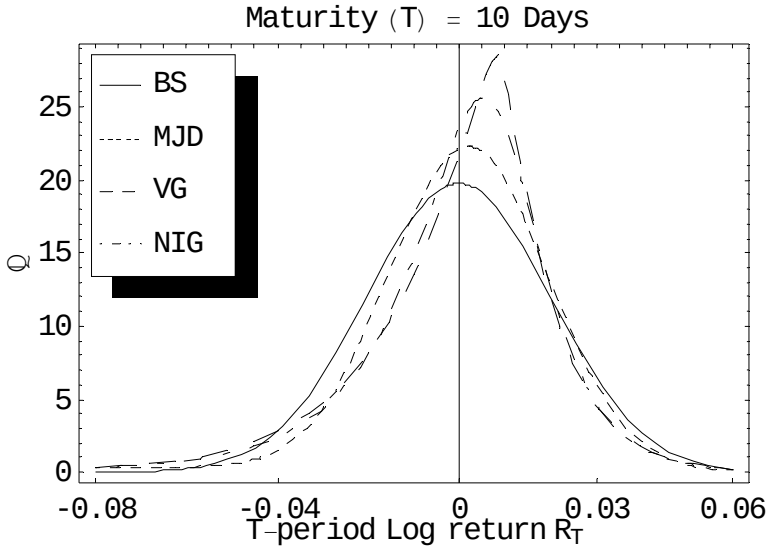
Figure 3.14: NIG model implied dynamics of log-return probability density



Note: Calibrated probability density functions of log returns in NIG model on one day before last trading day March 16, 2005 are plotted as the time to maturity nears from 1 to 0. Note that the above figures illustrate density functions only for the range of probability density between 0 and 15.

Figure 3.15: Plot of implied log-return probability density

A) March 4, 2005. 10 days to maturity.



B) January 20, 2005. 40 days to maturity.

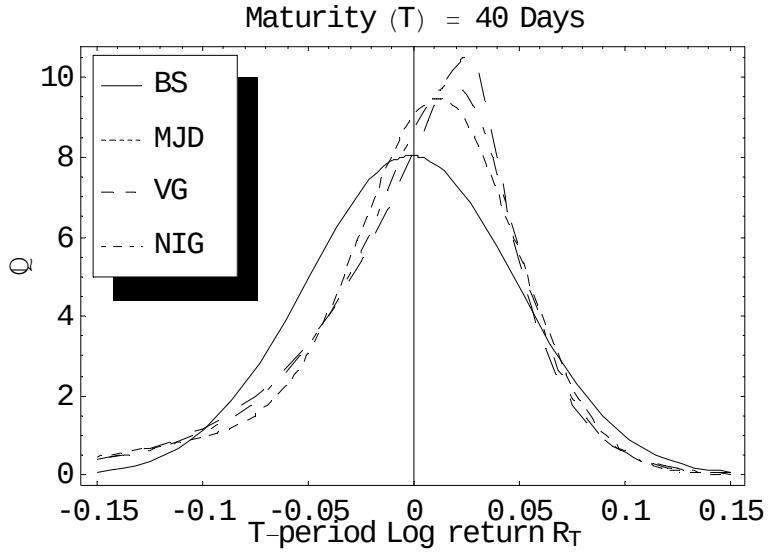
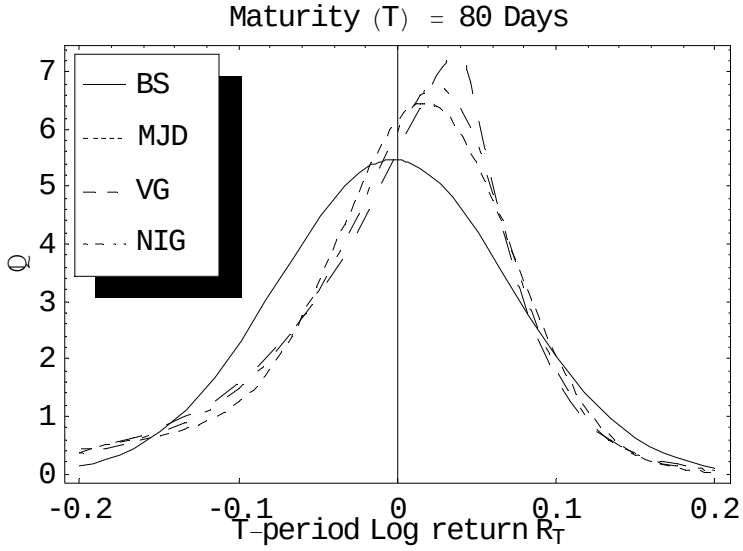


Figure 3.15: Plot of implied log-return probability density (Continued)

C) November 22, 2004. 80 days to maturity.



D) September 27, 2004. 120 days to maturity.

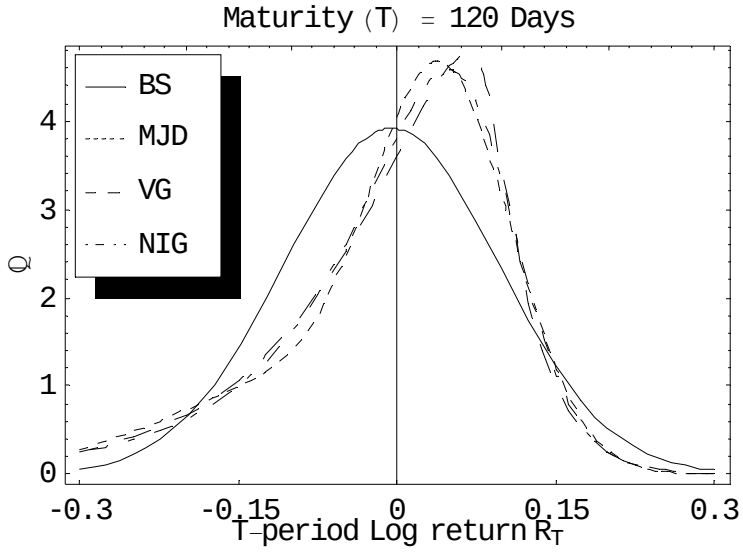
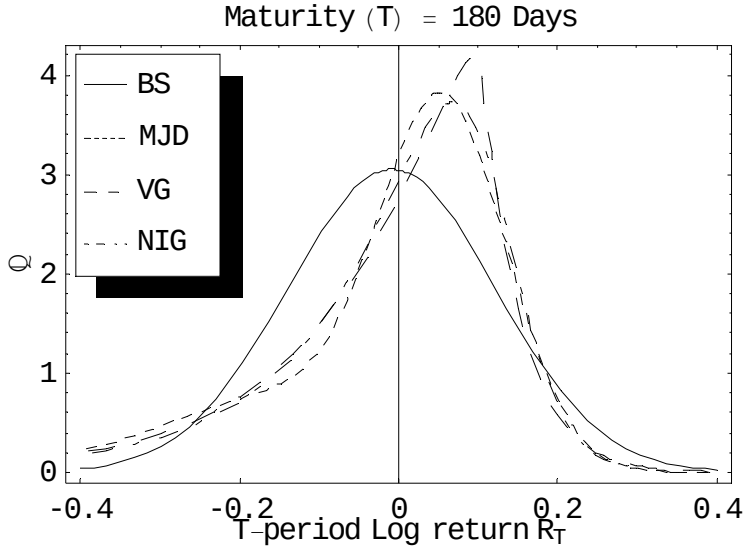


Figure 3.15: Plot of implied log-return probability density (Continued).

E) July 1, 2004. 180 days to maturity.



F) March 30, 2004. 245 days to maturity.

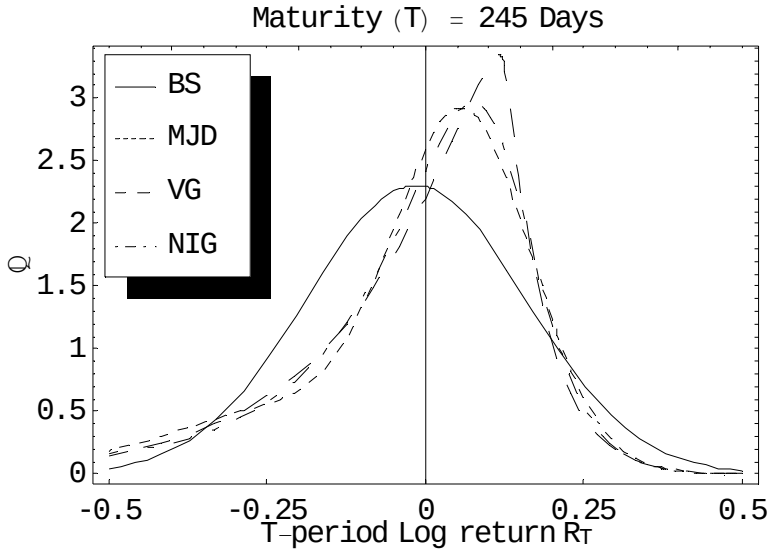
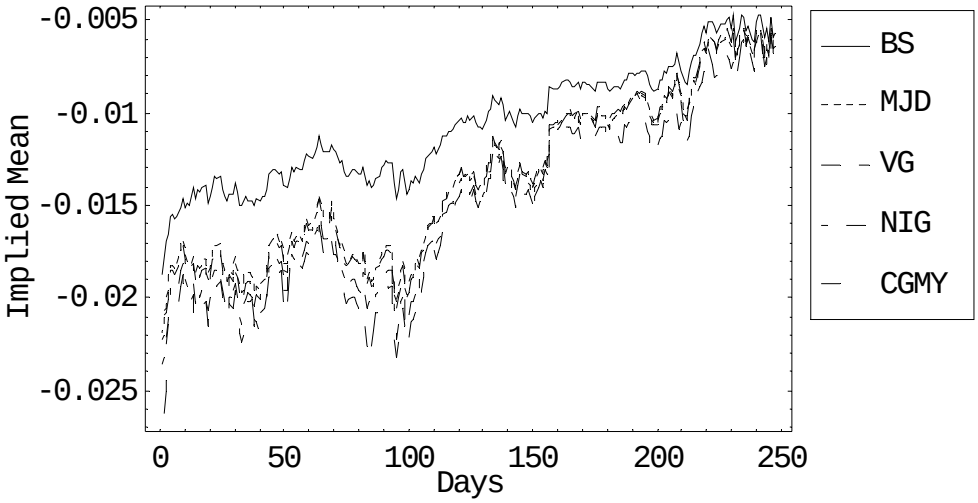
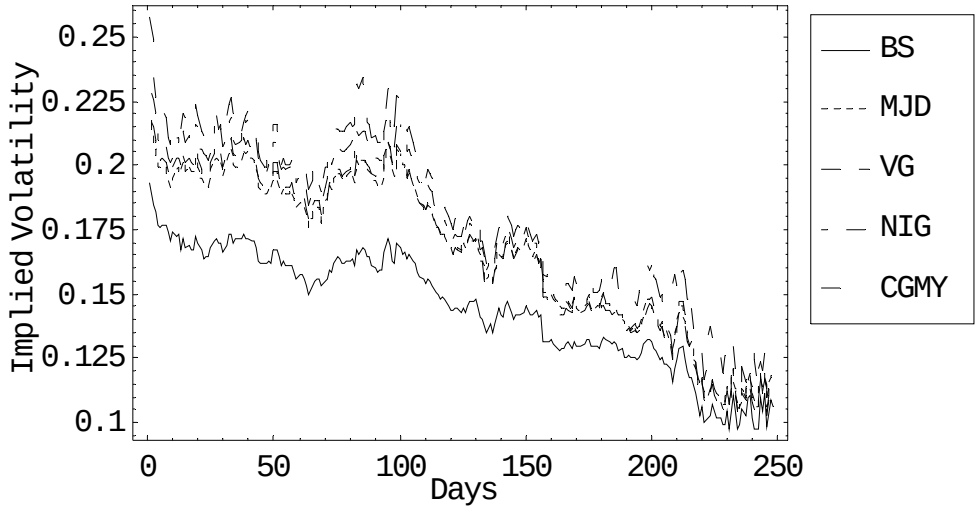


Figure 3.16: Dynamics of implied moments of log-return probability density

A) Implied Mean of Log Returns



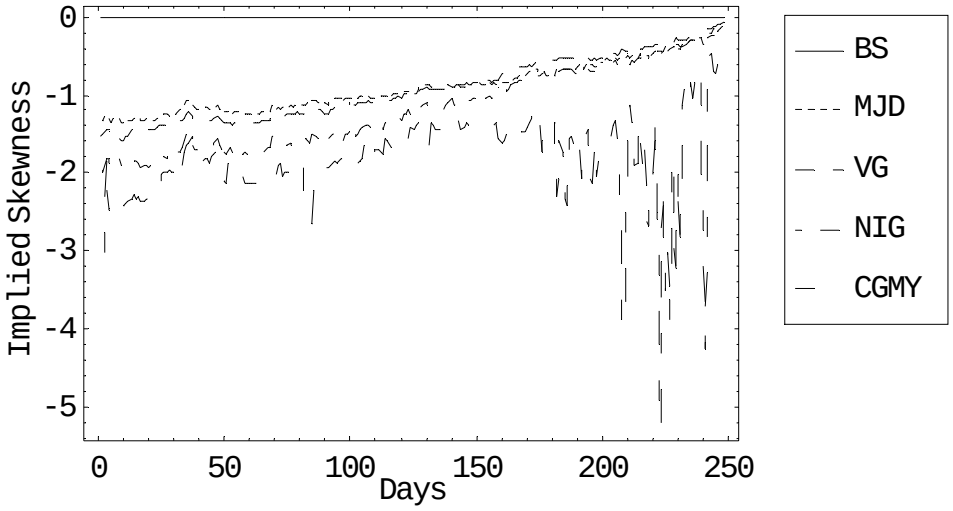
B) Implied Volatility of Log Returns



Note: Time series of implied moments of log return probability density functions on one day before last trading day March 16, 2005 are plotted for each model. Note that the day 1 is March 24, 2004 and the day 248 is March 16, 2005.

Figure 3.16: Dynamics of implied moments of log-return probability density (Continued)

C) Implied Skewness of Log Returns



D) Implied Excess Kurtosis of Log Returns

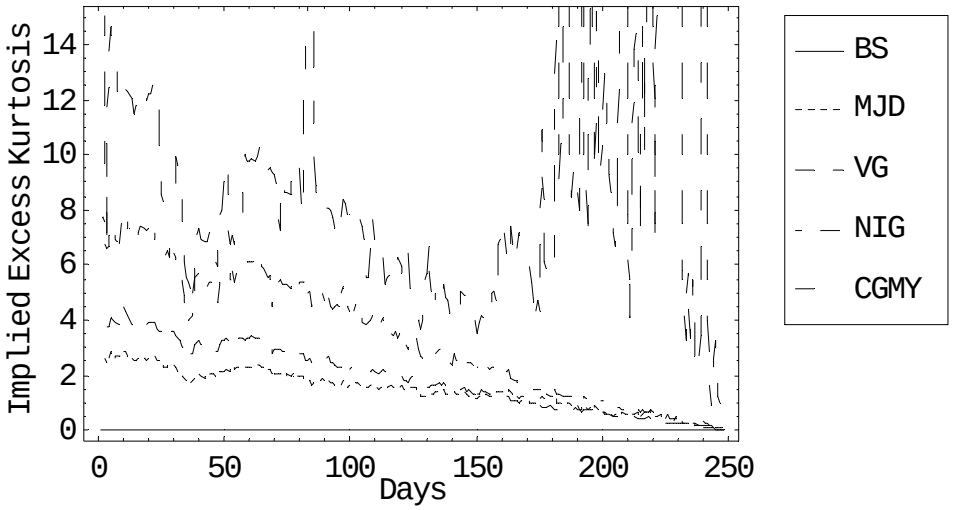


Figure 3.17: Implied dynamics of Lévy density: MJD model

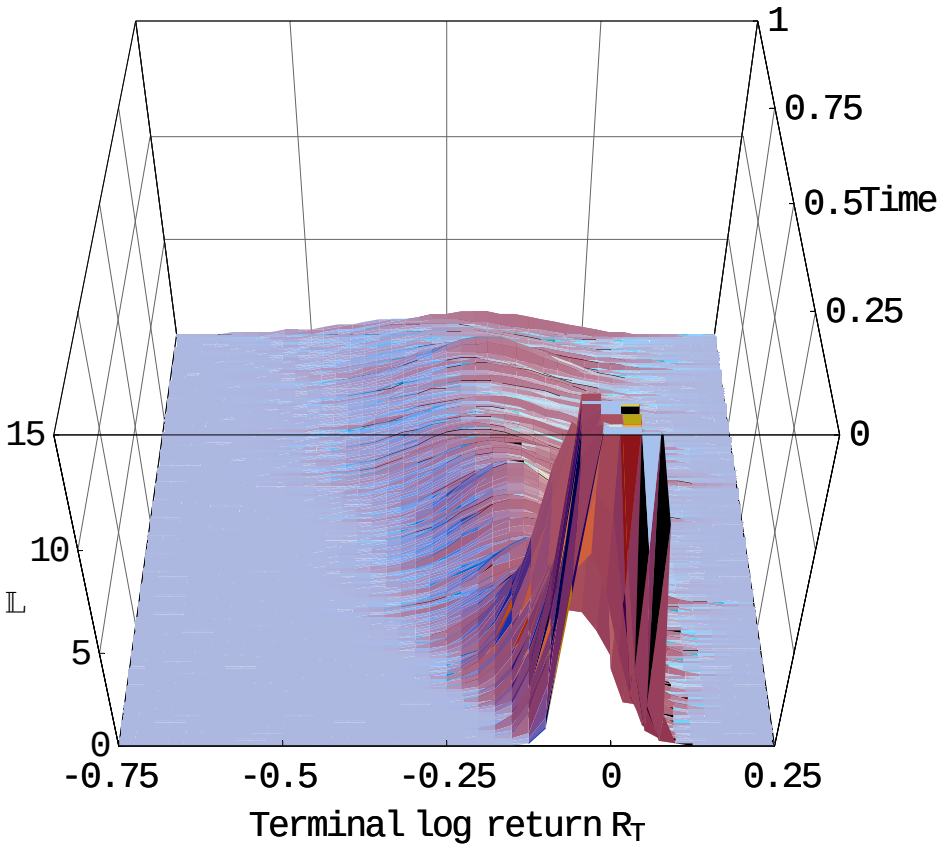


Figure 3.18: Implied dynamics of Lévy density: VG model

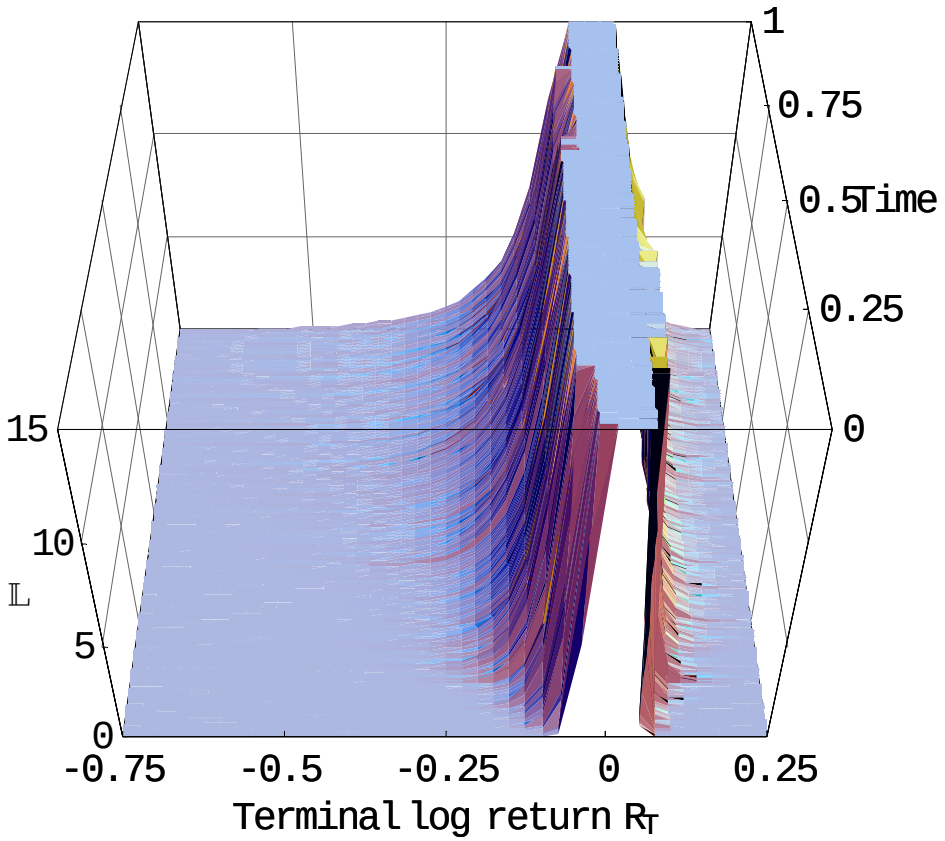


Figure 3.19: Implied dynamics of Lévy density: NIG model

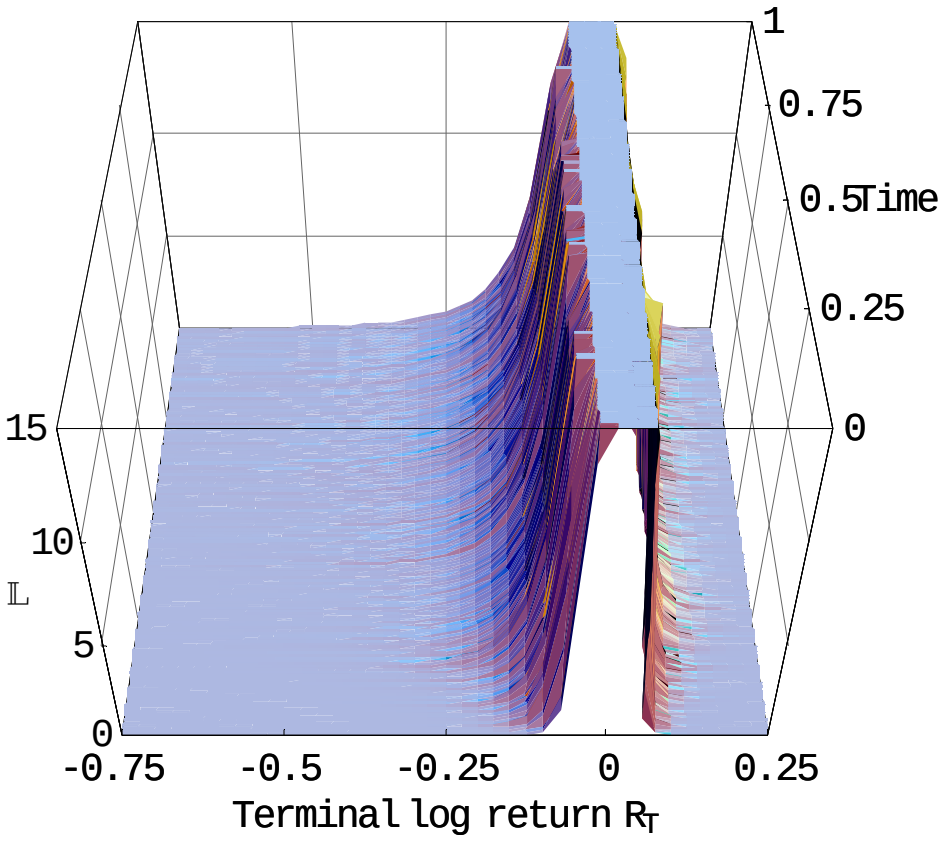


Figure 3.20: Implied dynamics of Lévy density: CGMY model

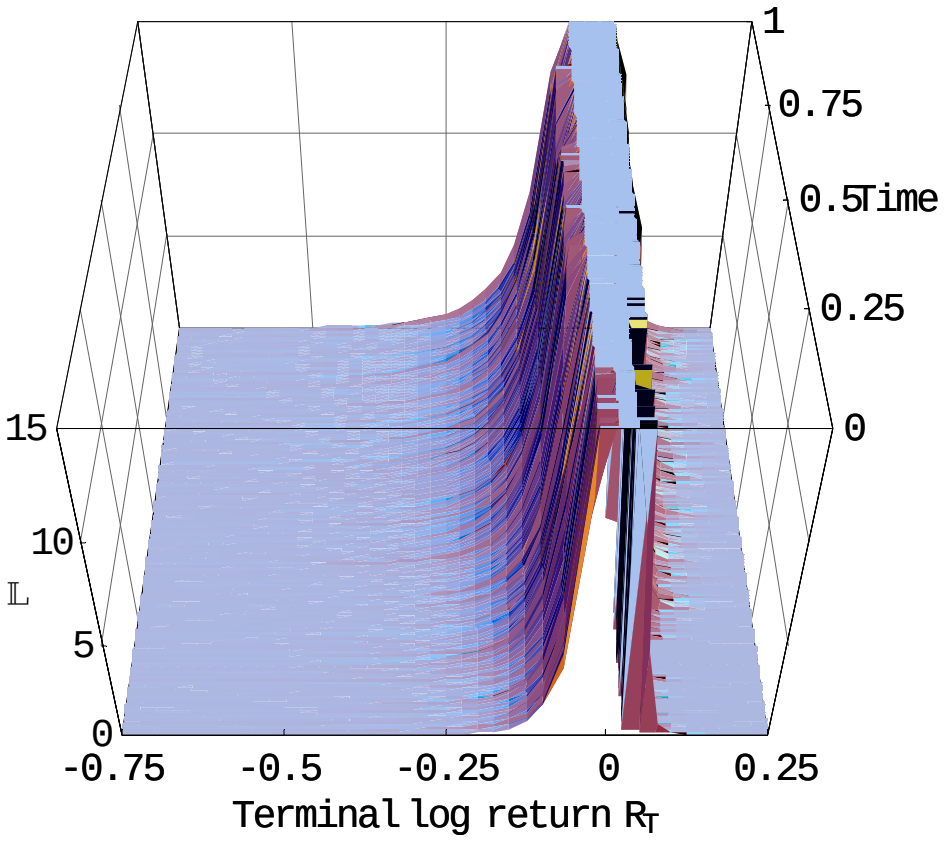
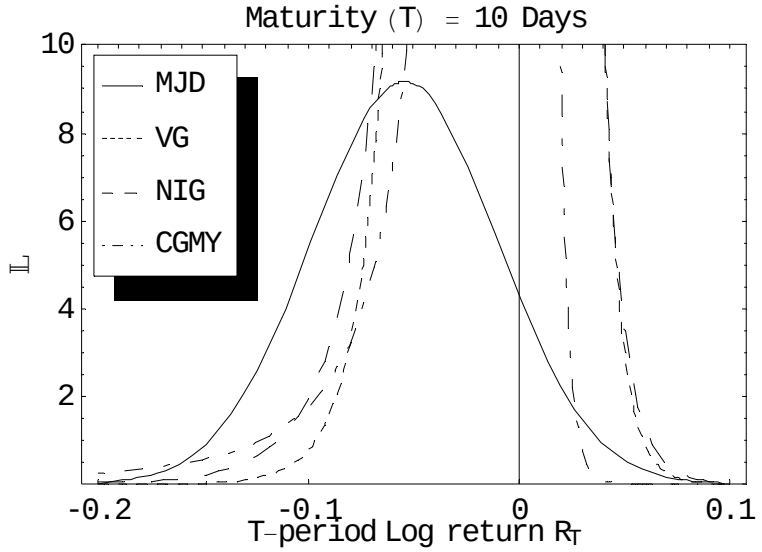


Figure 3.21: Plot of implied Lévy density of log returns

A) March 4, 2005. 10 days to maturity.



B) January 20, 2005. 40 days to maturity.

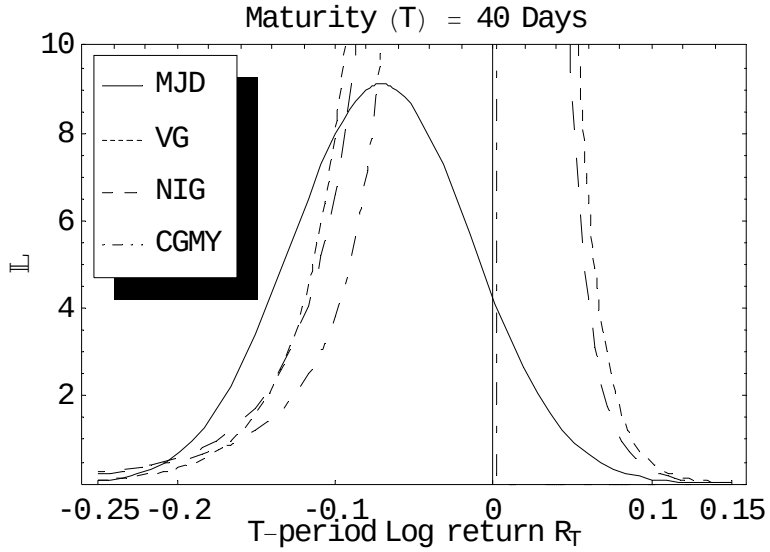
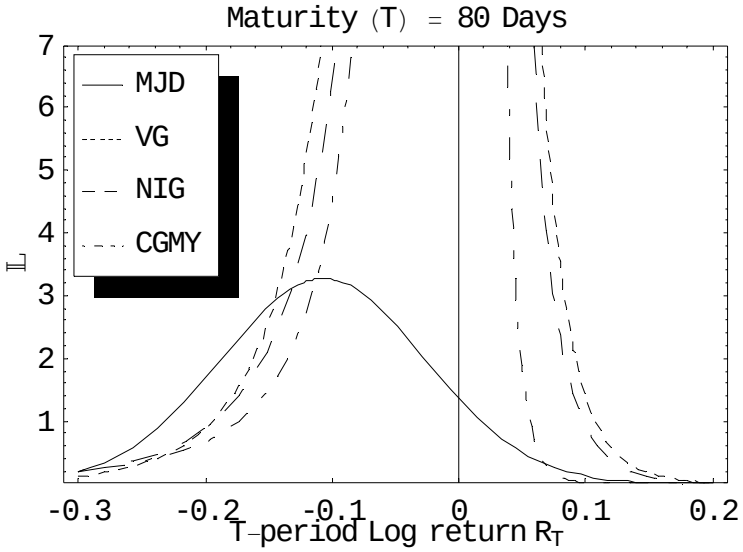


Figure 3.21: Plot of implied Lévy density of log returns (Continued)

C) November 22, 2004. 80 days to maturity.



D) September 27, 2004. 120 days to maturity.

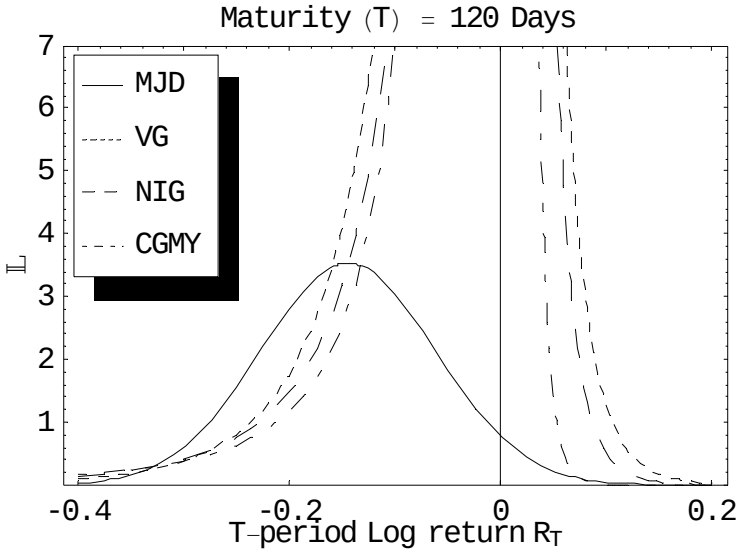
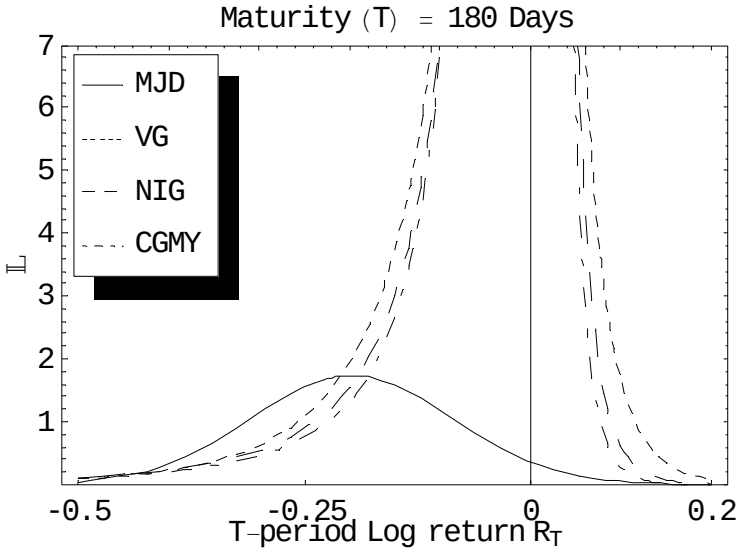


Figure 3.21: Plot of implied Lévy density of log returns (Continued)

E) July 1, 2004. 180 days to maturity.



F) March 30, 2004. 245 days to maturity.

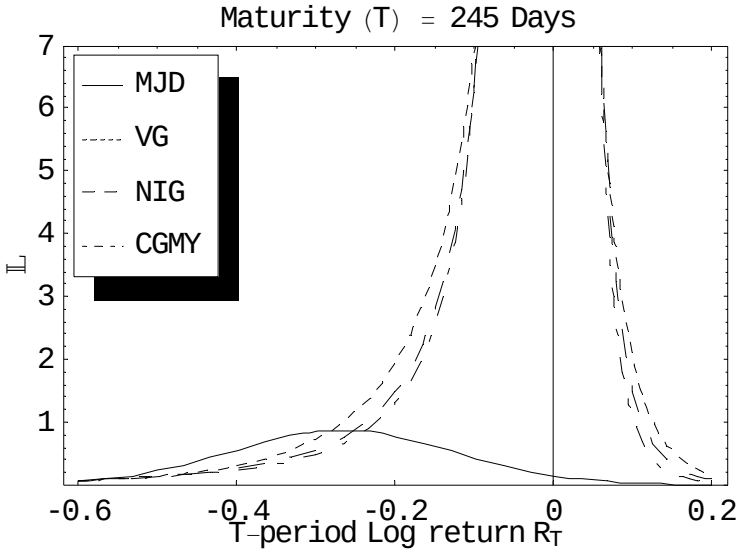
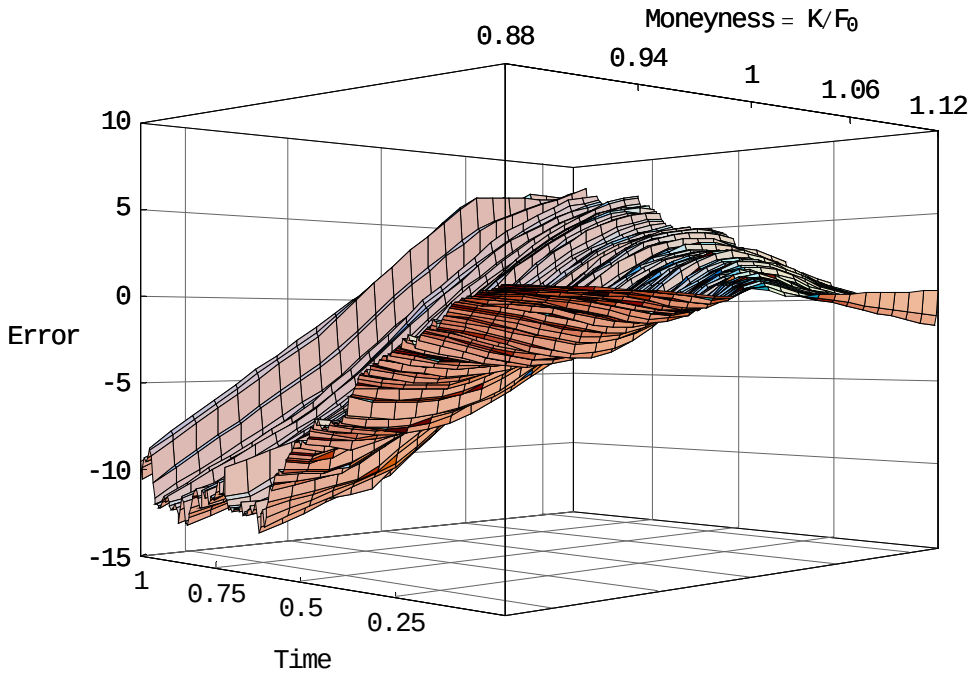
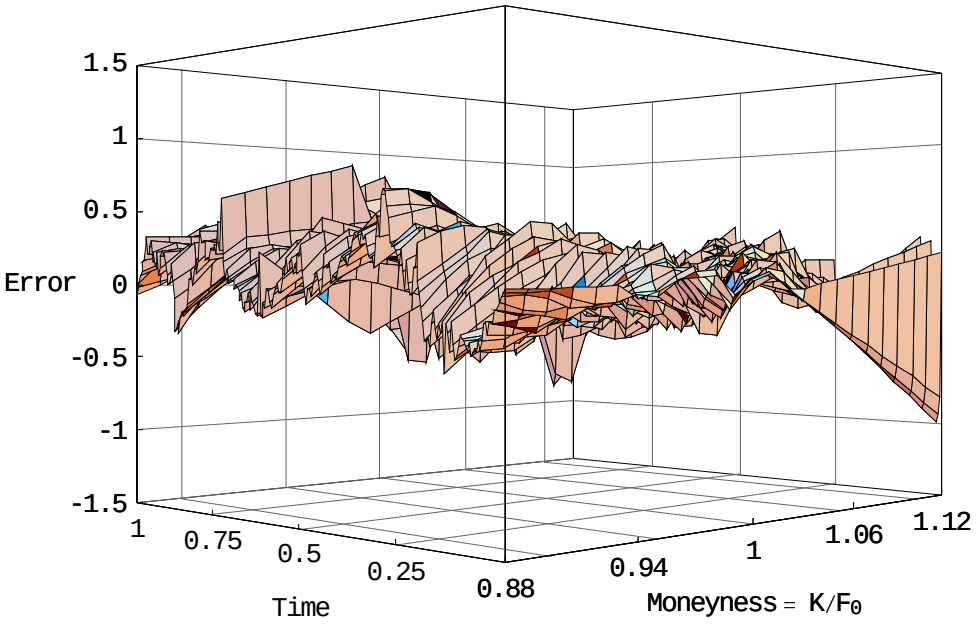


Figure 3.22: Dynamics of in-sample dollar pricing errors: BS model



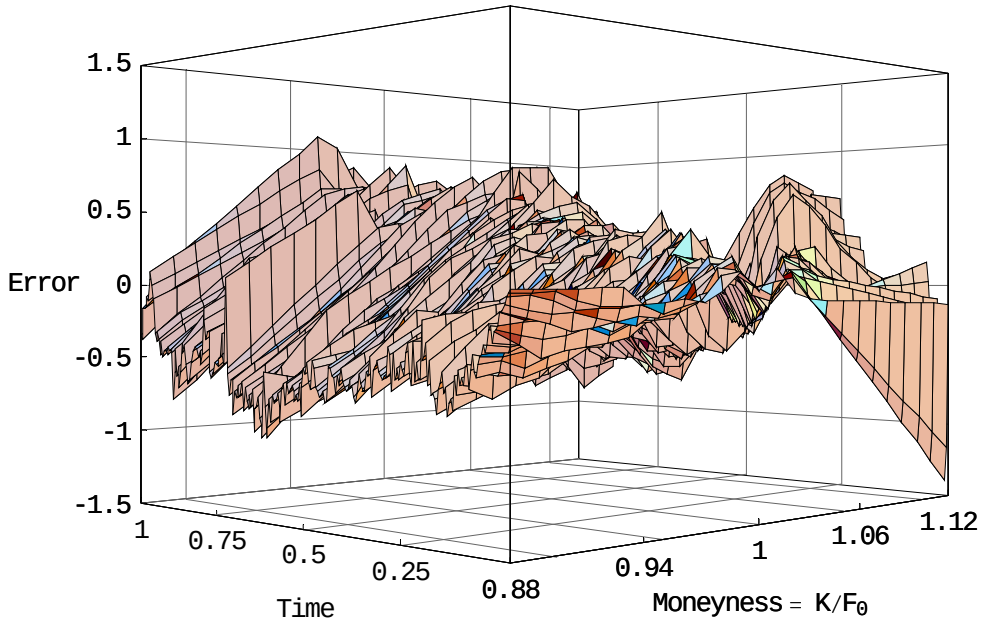
Note: In-sample dollar pricing error defined as $C_i^{\text{model}}(S_0, K_i, T; \eta^{\mathbb{Q}}{}_0) - C_i^{\text{market}}$ is calculated on each day and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 3.23: Dynamics of in-sample dollar pricing errors: MJD model



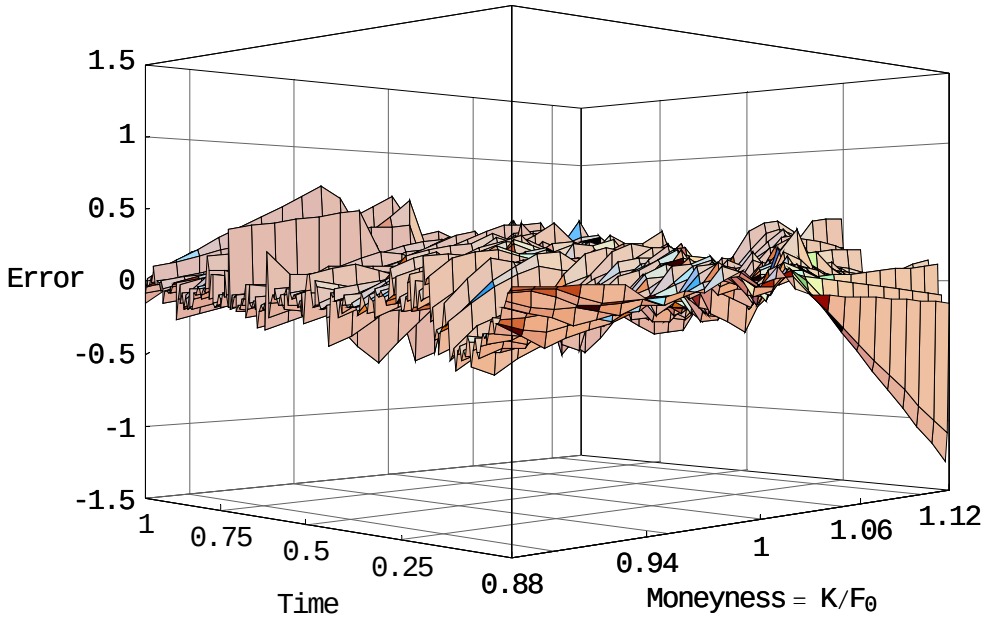
Note: In-sample dollar pricing error defined as $C_i^{\text{model}}(S_0, K_i, T; \eta^{\mathbb{Q}}{}_0) - C_i^{\text{market}}$ is calculated on each day and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 3.24: Dynamics of in-sample dollar pricing errors: VG model



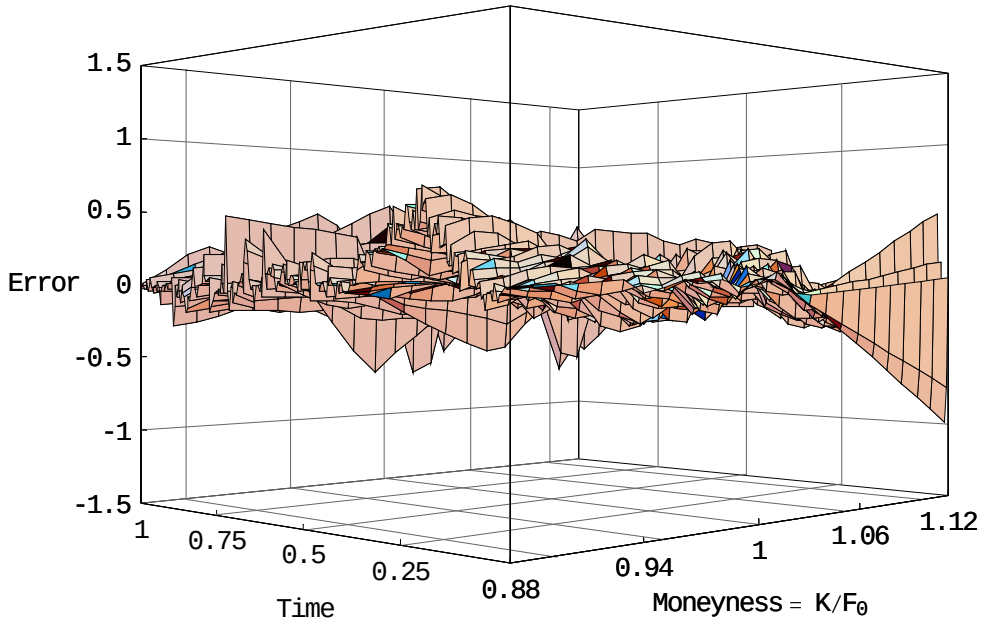
Note: In-sample dollar pricing error defined as $C_i^{\text{model}}(S_0, K_i, T; \eta^{\mathbb{Q}}{}_0) - C_i^{\text{market}}$ is calculated on each day and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 3.25: Dynamics of in-sample dollar pricing errors: NIG model



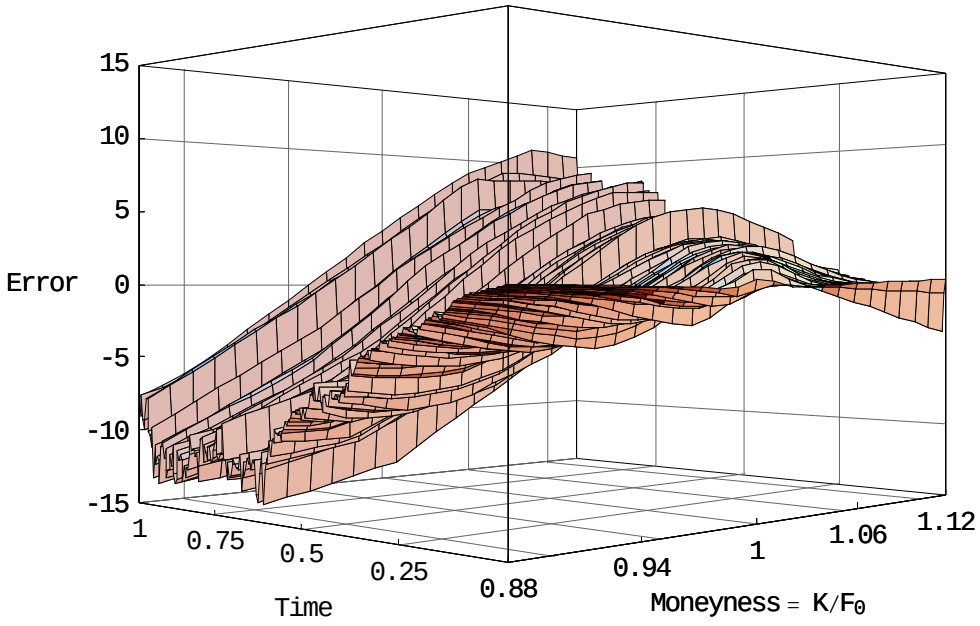
Note: In-sample dollar pricing error defined as $C_i^{\text{model}}(S_0, K_i, T; \eta^{\mathbb{Q}}{}_0) - C_i^{\text{market}}$ is calculated on each day and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 3.26: Dynamics of in-sample dollar pricing errors: CGMY model



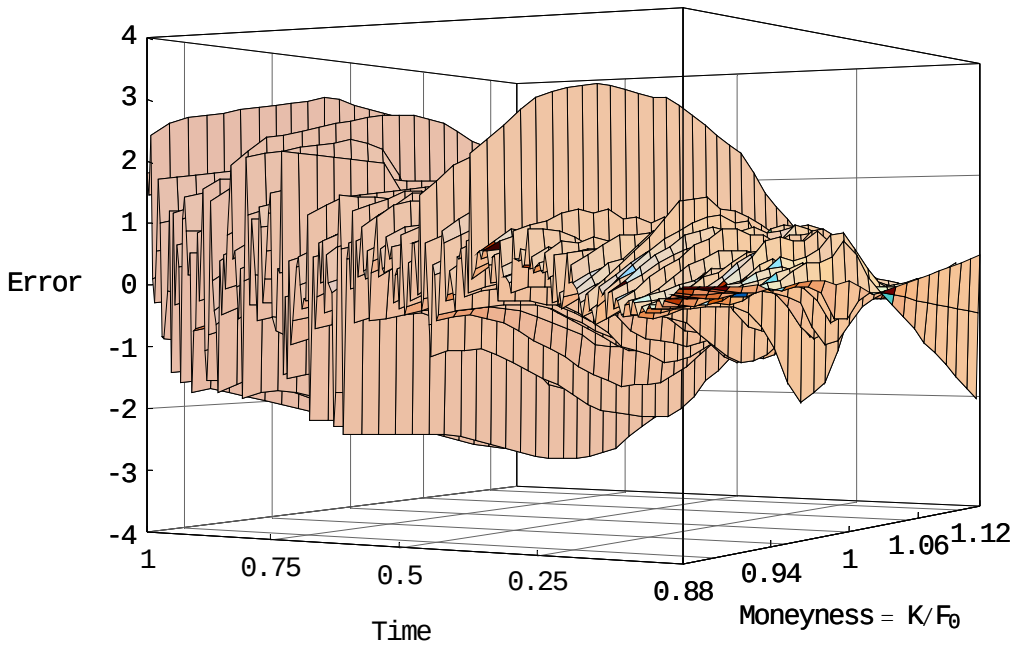
Note: In-sample dollar pricing error defined as $C_i^{\text{model}}(S_0, K_i, T; \eta^{\mathbb{Q}}{}_0) - C_i^{\text{market}}$ is calculated on each day and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 3.27: Dynamics of out-of-sample dollar pricing errors: BS model



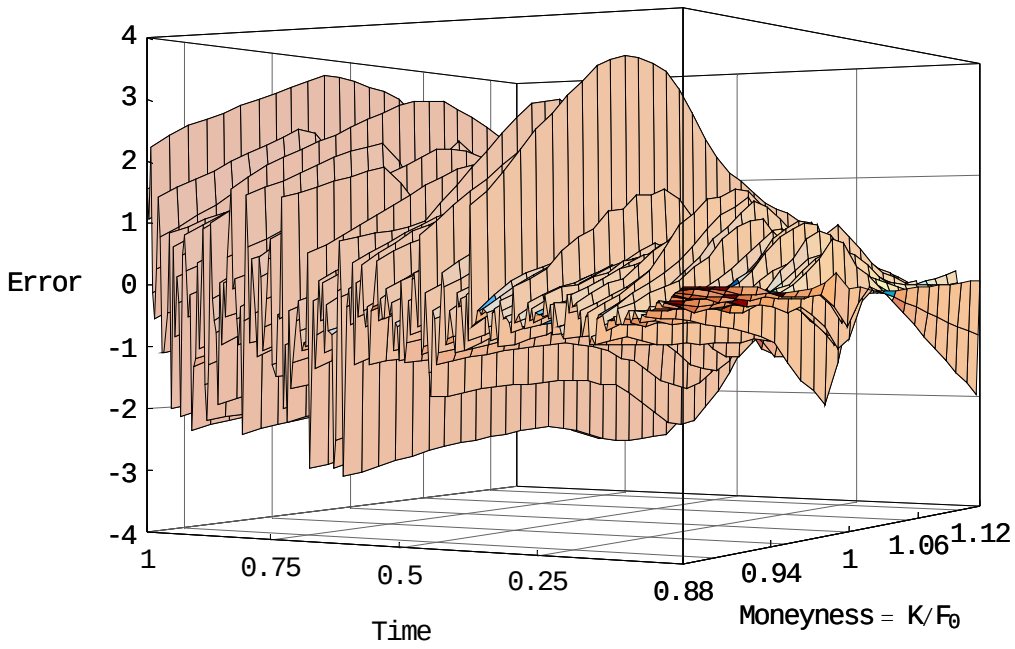
Note: Out-of-sample dollar pricing error defined as $C_t^{\text{model}}(F_t, K, T-t; \eta^{\mathbb{Q}}{}_{t-1})$ $-C_t^{\text{market}}$ is calculated on each day across different strike prices and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 3.28: Dynamics of out-of-sample dollar pricing errors: MJD model



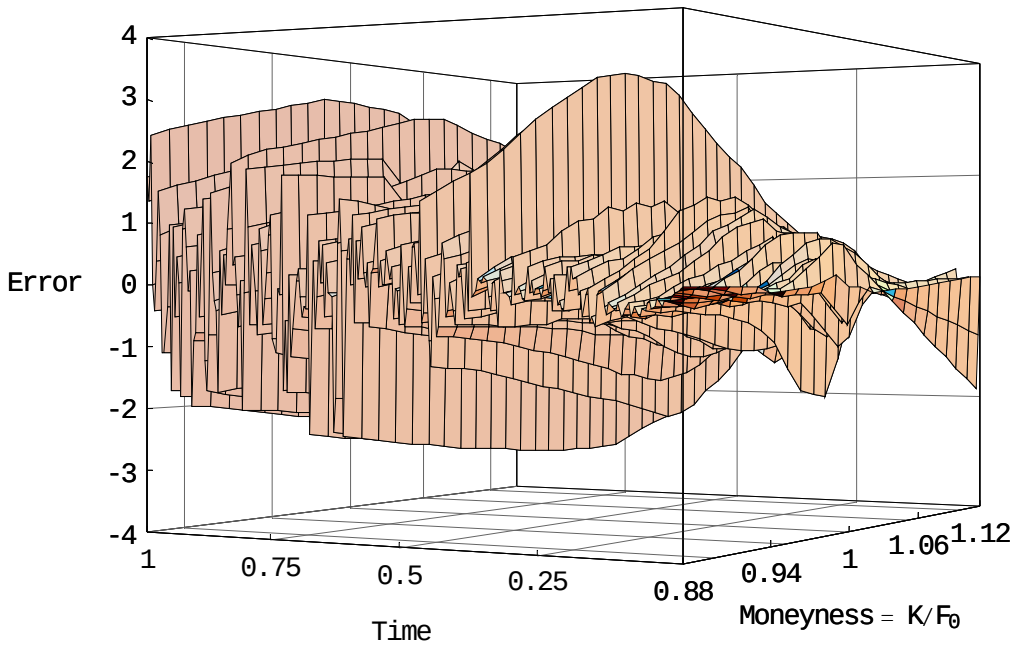
Note: Out-of-sample dollar pricing error defined as $C_t^{\text{model}}(F_t, K, T-t; \eta^{\mathbb{Q}}{}_{t-1})$ $-C_t^{\text{market}}$ is calculated on each day across different strike prices and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 3.29: Dynamics of out-of-sample dollar pricing errors: VG model



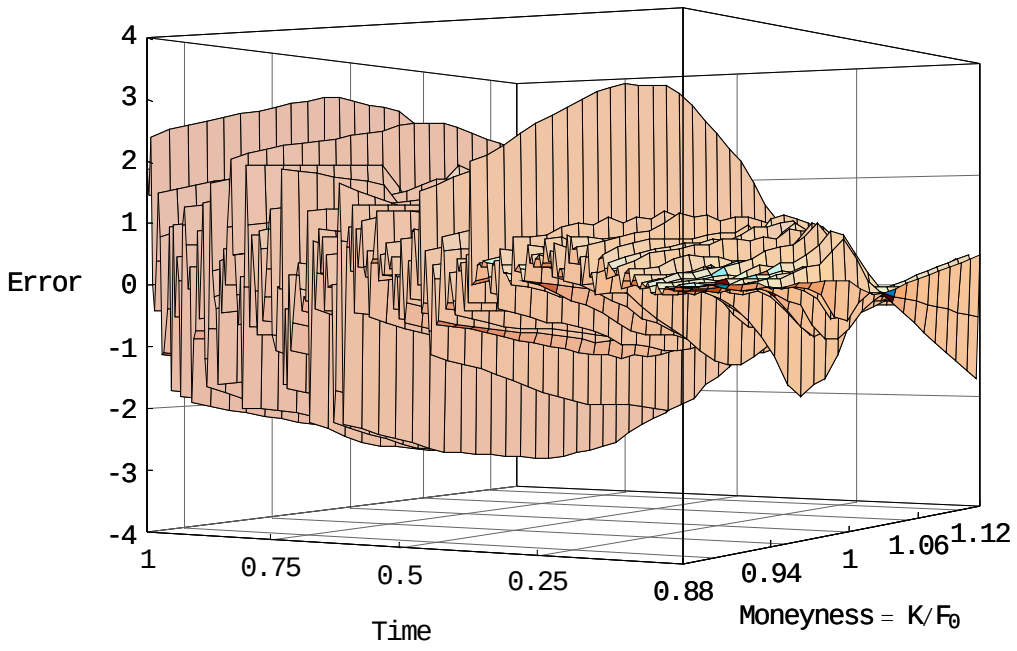
Note: Out-of-sample dollar pricing error defined as $C_t^{\text{model}}(F_t, K, T-t; \eta^{\mathbb{Q}}{}_{t-1})$ $-C_t^{\text{market}}$ is calculated on each day across different strike prices and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 3.30: Dynamics of out-of-sample dollar pricing errors: NIG model



Note: Out-of-sample dollar pricing error defined as $C_t^{\text{model}}(F_t, K, T-t; \eta^{\mathbb{Q}}{}_{t-1})$ $-C_t^{\text{market}}$ is calculated on each day across different strike prices and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

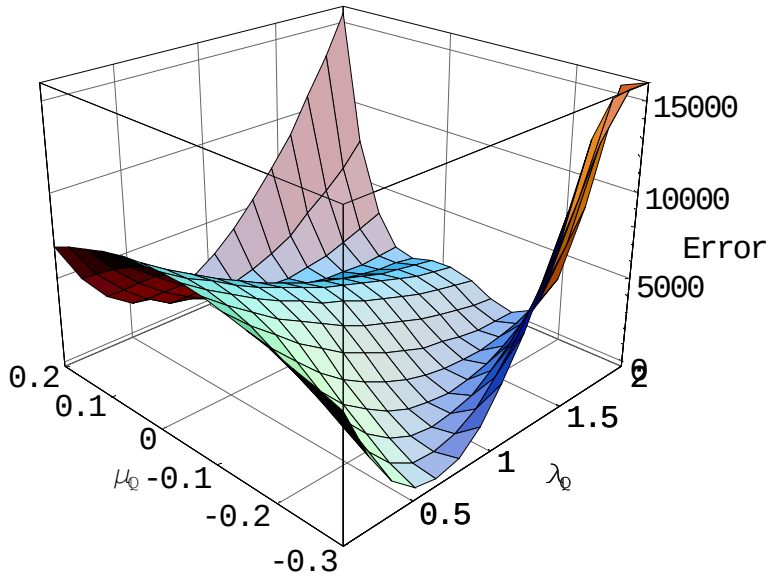
Figure 3.31: Dynamics of out-of-sample dollar pricing errors: CGMY model



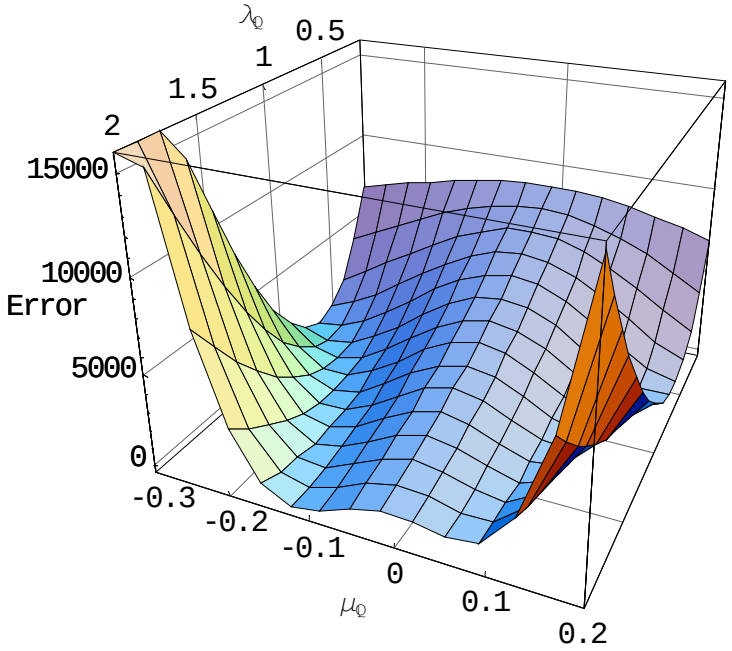
Note: Out-of-sample dollar pricing error defined as $C_t^{\text{model}}(F_t, K, T-t; \eta^{\mathbb{Q}}{}_{t-1})$ $-C_t^{\text{market}}$ is calculated on each day across different strike prices and plotted in a 3D figure across the varying moneyness and the time to maturity using a linear interpolation. Note that the time to maturity is measured in years.

Figure 4.1: 3D plot of the MJD error function

A) From one angle.



B) From another angle.



Note: The sum of squared dollar pricing error function $\sum_{i=1}^{N}\left|C_i^{\text{model}}(\theta^{\mathbb{Q}})-C_i^{\text{market}}\right|^2$ is plotted as a function of parameters $\lambda^{\mathbb{Q}}$ and $\mu^{\mathbb{Q}}$ with other parameters being fixed.

Figure 4.2: Strictly convex function f(x) = xlnx

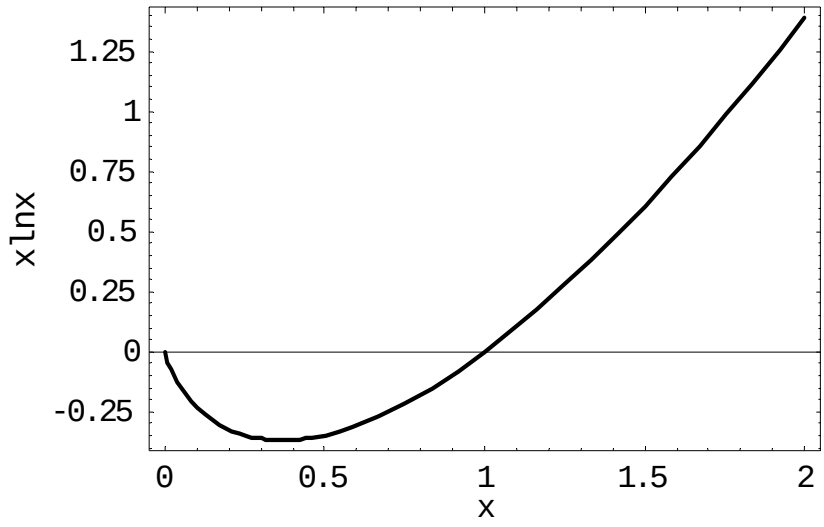
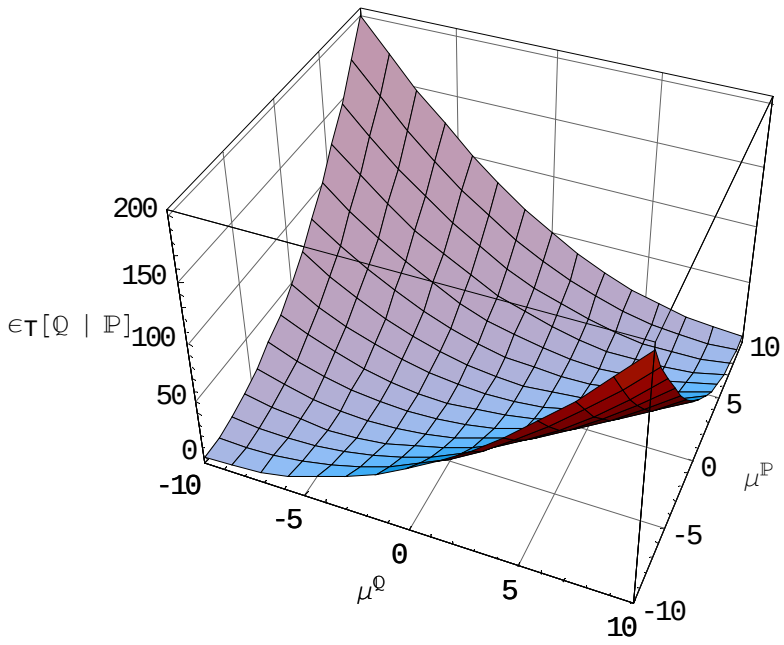
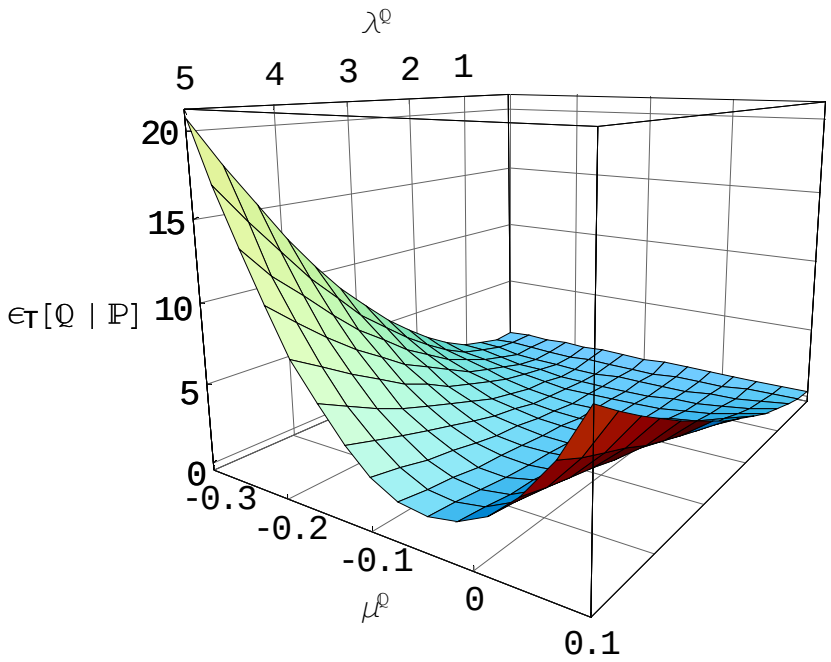


Figure 4.3: Relative entropy function for Gaussian case



Note: We set $T = 1$ and $\sigma = 1$ for simplicity.

Figure 4.4: Relative entropy function for MJD Process



Note: We set $T = 0.25$, $\sigma = 0.1$, $\delta^{\mathbb{Q}} = 0.15$, $\lambda^{\mathbb{P}} = 1$, $\mu^{\mathbb{P}} = -0.1$, and $\delta^{\mathbb{P}} = 0.1$.

Figure 4.5: Statistical daily log-return probability density

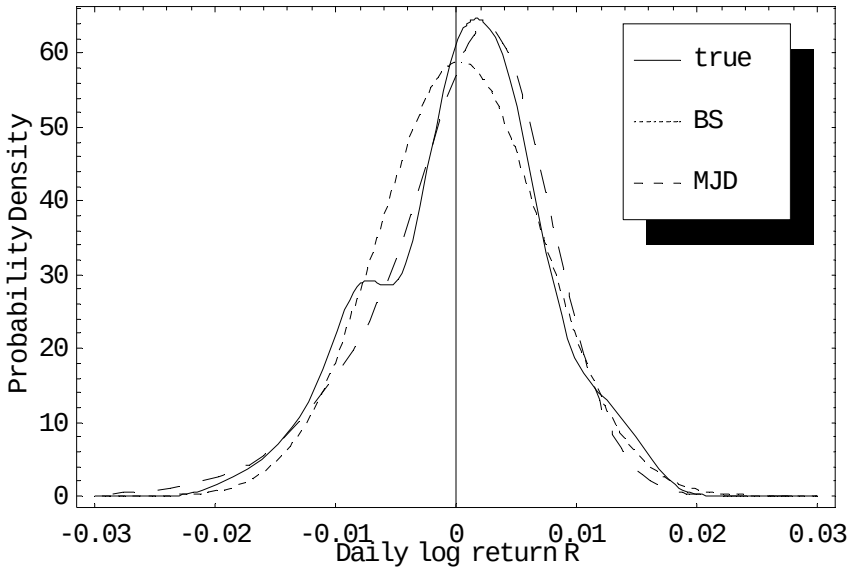


Figure 4.6: Log of statistical daily log-return probability density

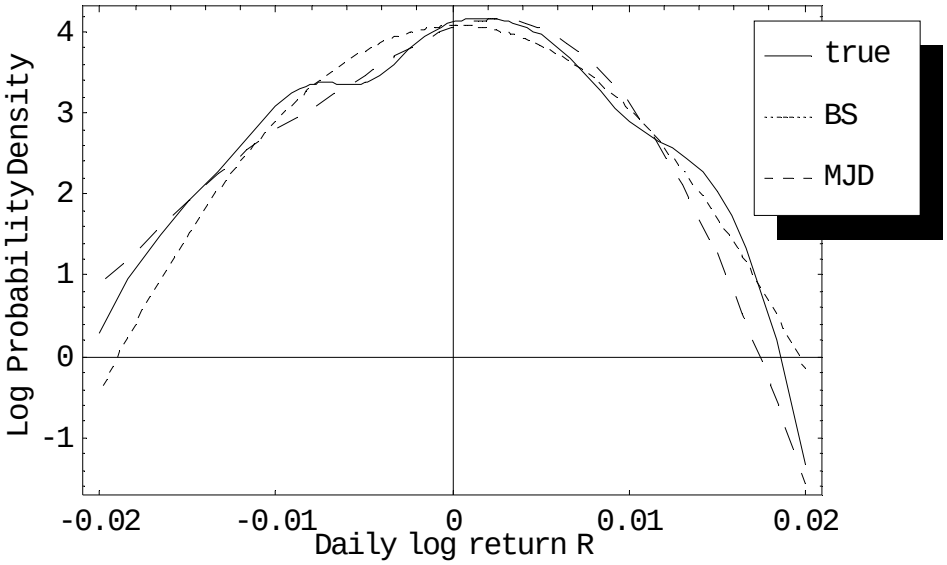


Figure 4.7: Plot of statistically estimated MJD Lévy measure

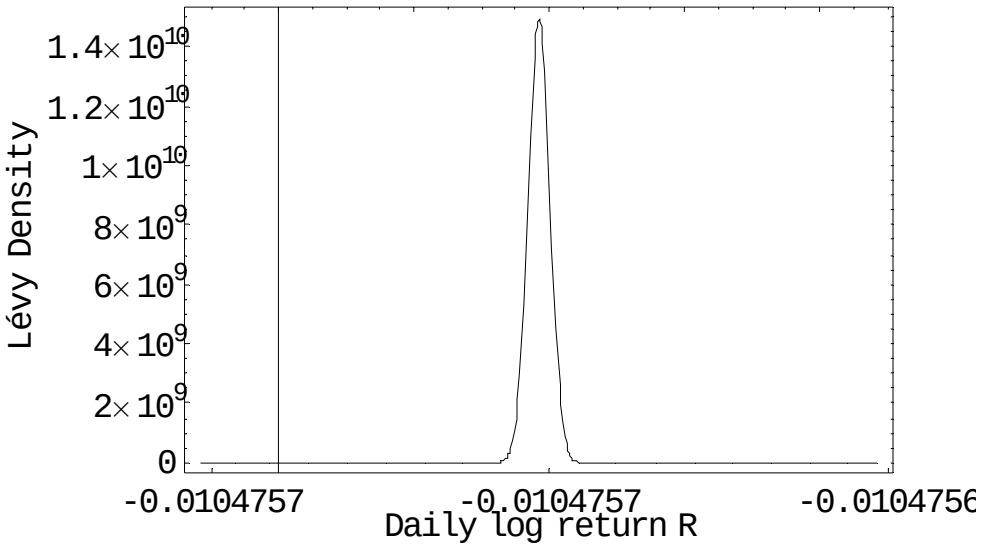


Figure 4.8: Successively-updated risk-neutral prior parameters

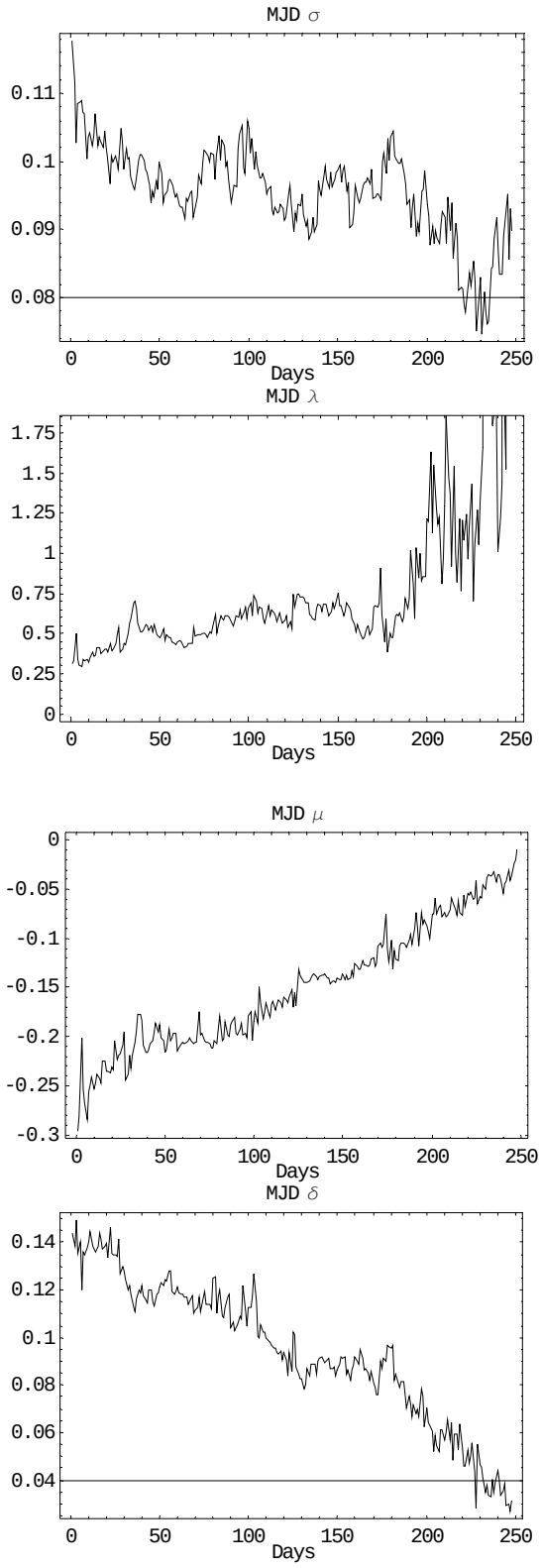
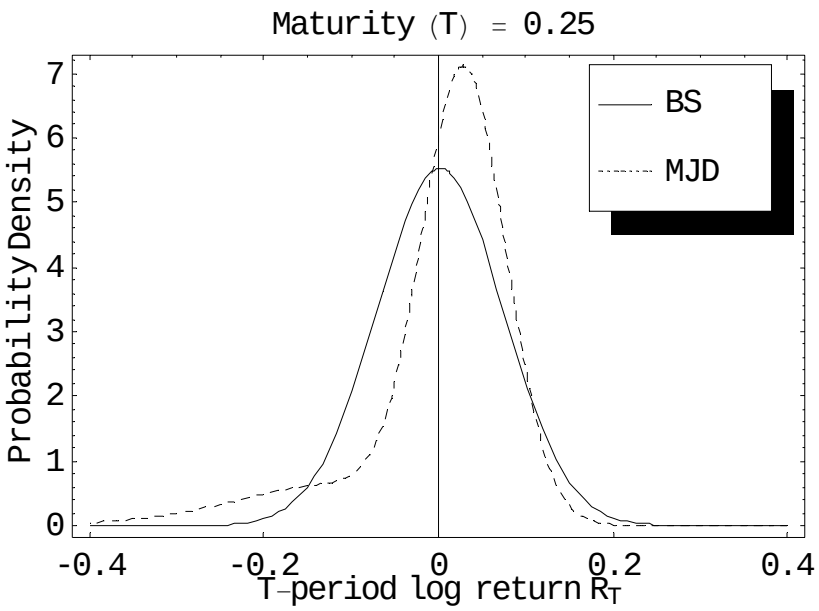


Figure 4.9: Risk-neutral log-return probability density



Note: The maturity is set at $T = 0.25$.

Figure 4.10: Log of risk-neutral daily log-return probability density

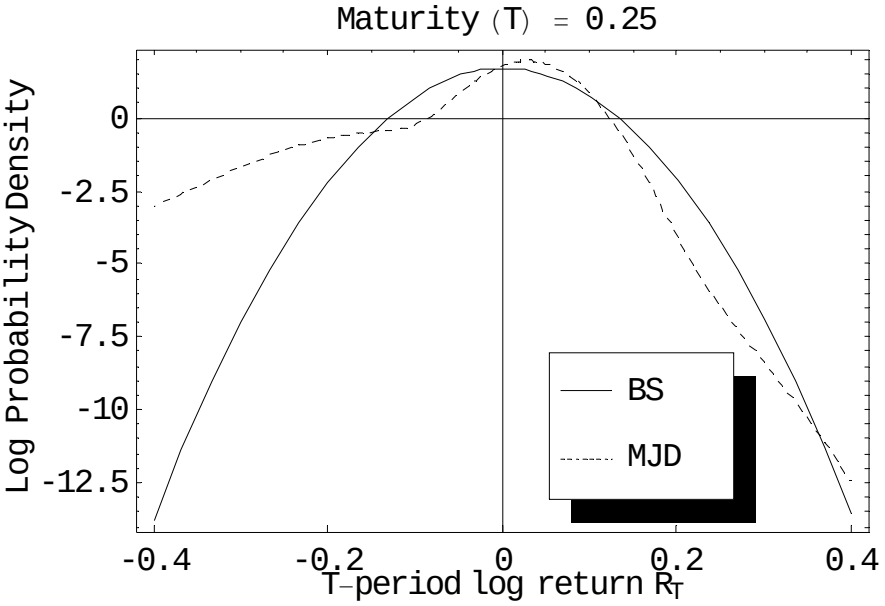


Figure 4.11: Plot of risk-neutral MJD Lévy measure

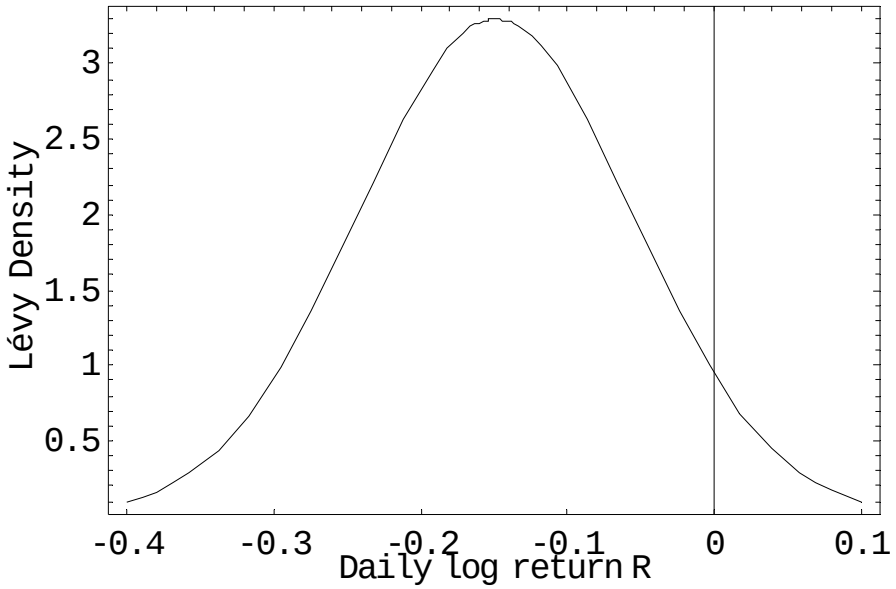
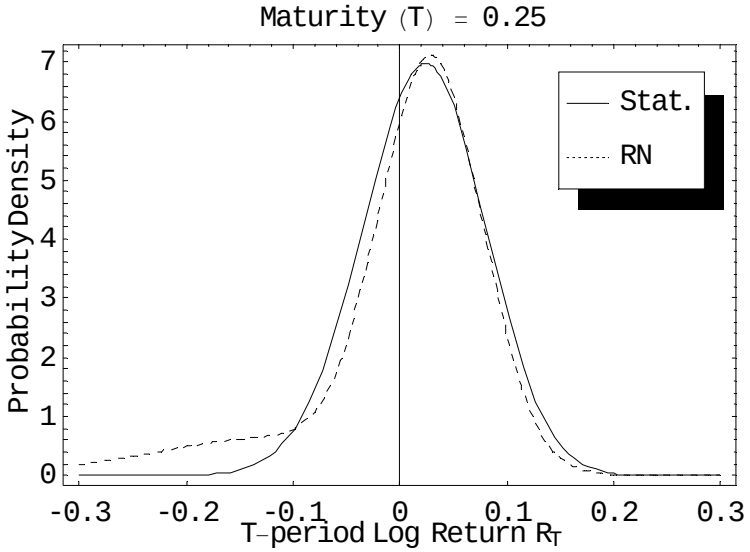
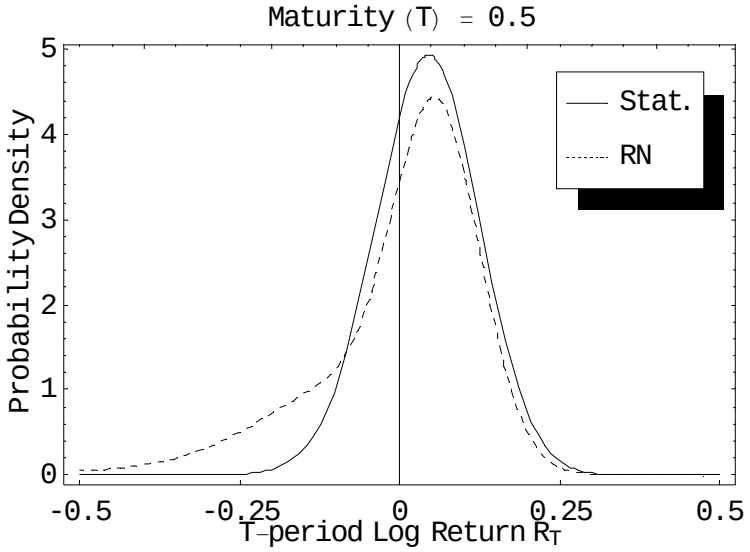


Figure 4.12: Statistical vs. risk-neutral prior log-return probability densities

A) Maturity $T = 0.25$.



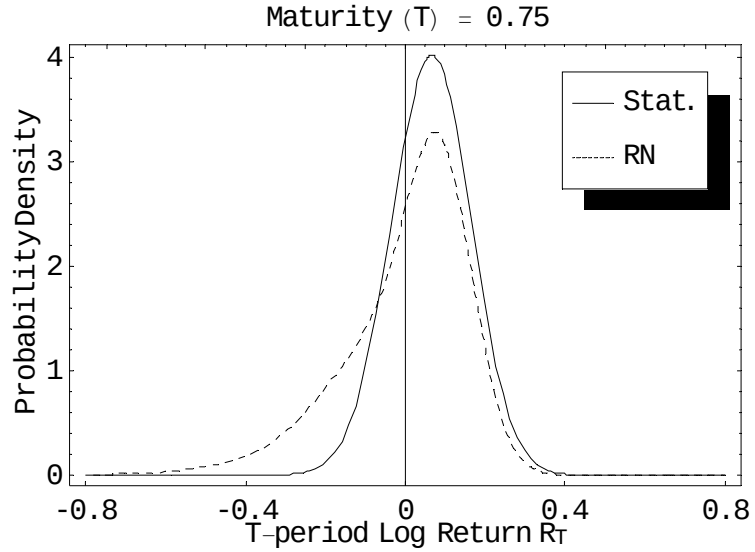
B) Maturity $T = 0.5$.



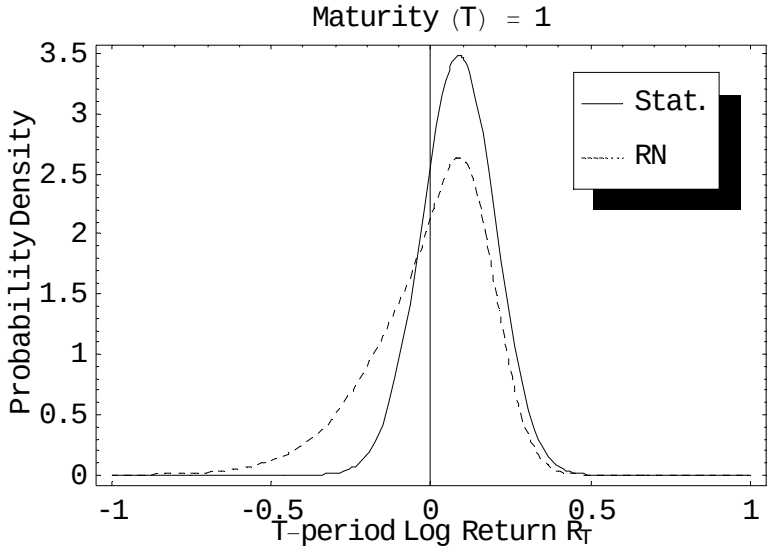
Note: The statistical prior log return probability density is compared to the risk-neutral prior log return probability density for four different maturities $T = 0.25,\ 0.5,\ 0.75,$ and 1.

Figure 4.12: Statistical vs. risk-neutral prior log-return probability densities (Continued)

C) Maturity $T = 0.75$.



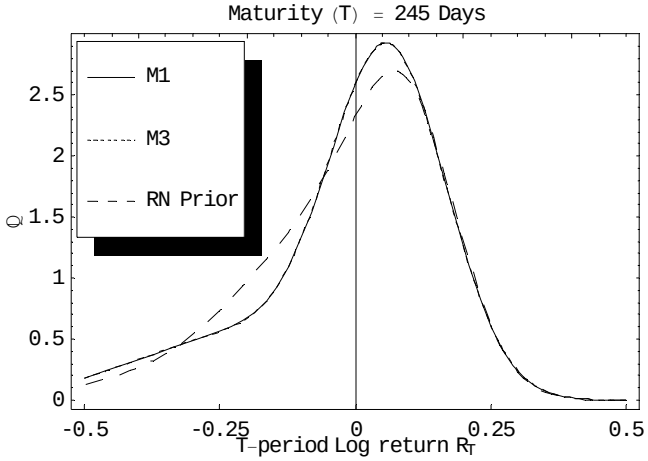
D) Maturity $T = 1$.



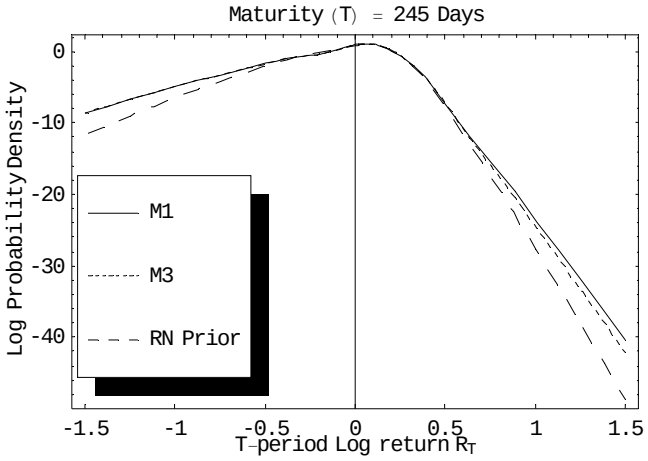
Note: The statistical prior log return probability density is compared to the risk-neutral prior log return probability density for four different maturities $T = 0.25,\ 0.5,\ 0.75,$ and 1.

Figure 4.13: Plot of calibrated log-return probability density and Lévy measure on March 30, 2004 (245 days to maturity)

A) Calibrated log return probability density.



B) Log of Panel A).



C) Calibrated Lévy measure.

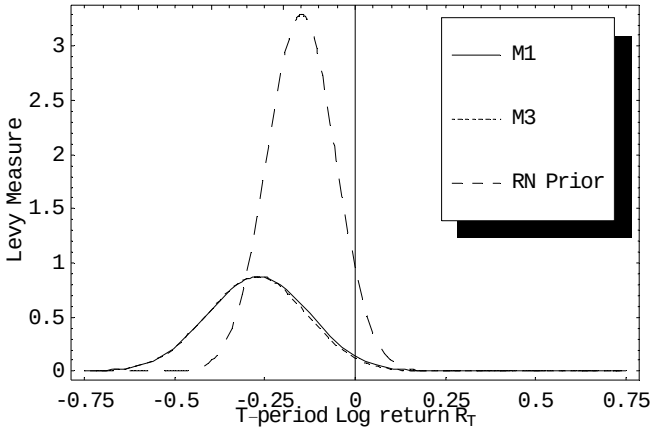
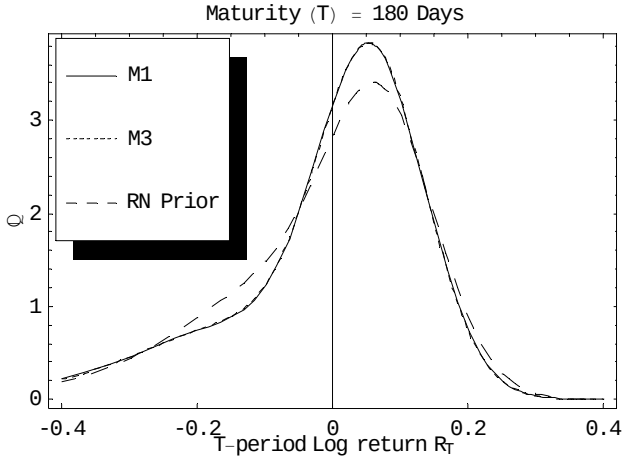
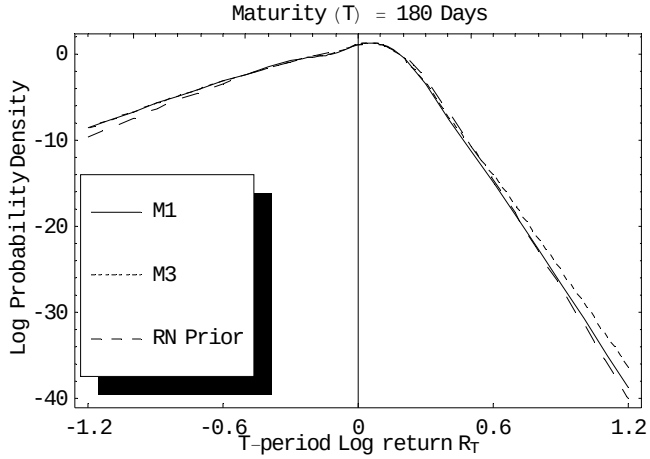


Figure 4.14: Plot of calibrated log-return probability density and Lévy measure on July 1, 2004 (180 days to maturity)

A) Calibrated log return probability density.



B) Log of Panel A).



C) Calibrated Lévy measure.

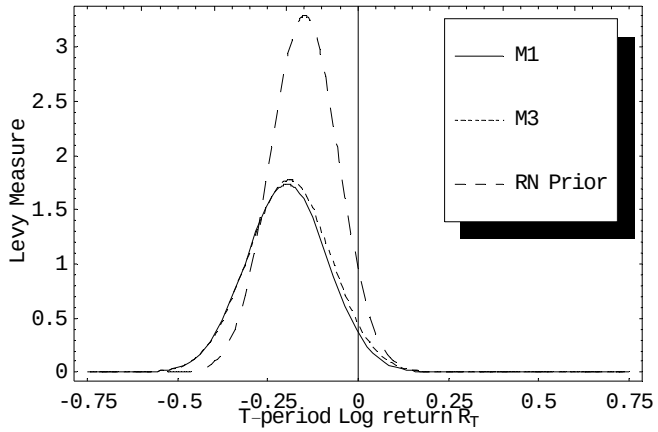
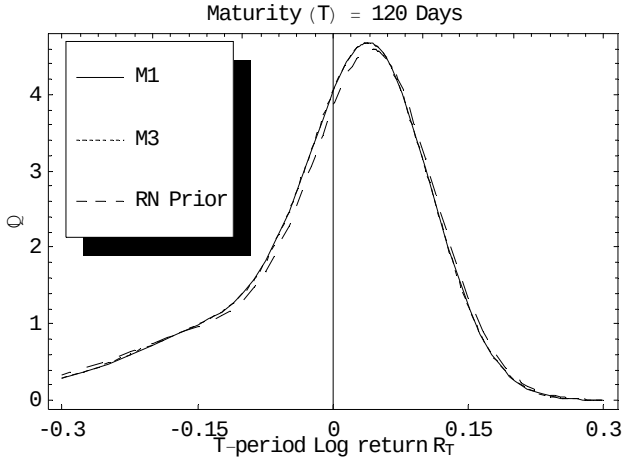
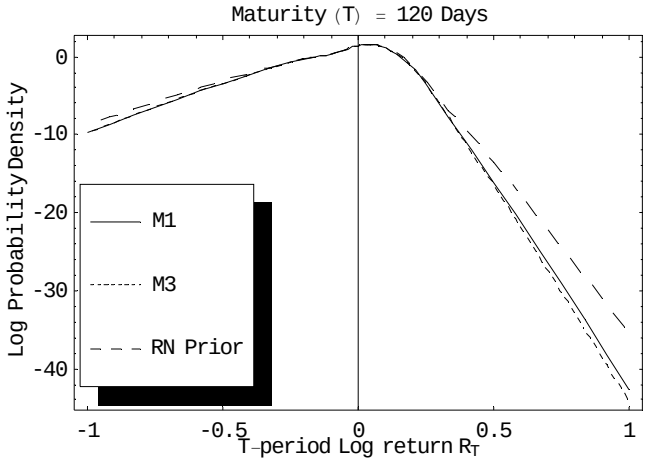


Figure 4.15: Plot of calibrated log-return probability density and Lévy measure on September 27, 2004 (120 days to maturity)

A) Calibrated log return probability density.



B) Log of Panel A).



C) Calibrated Lévy measure.

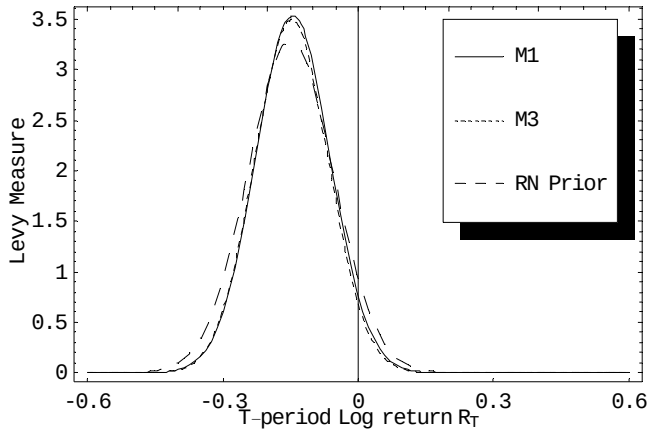
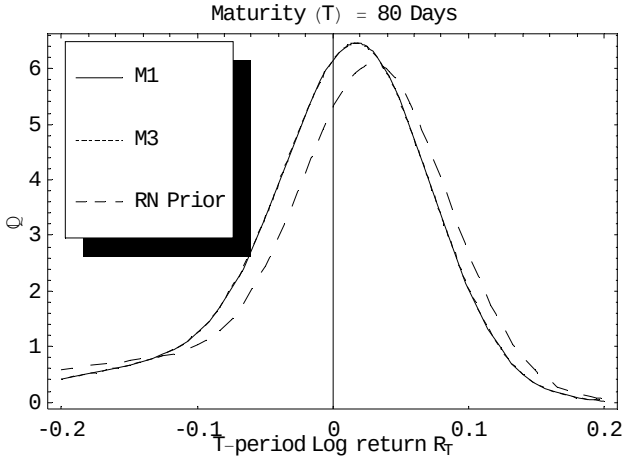
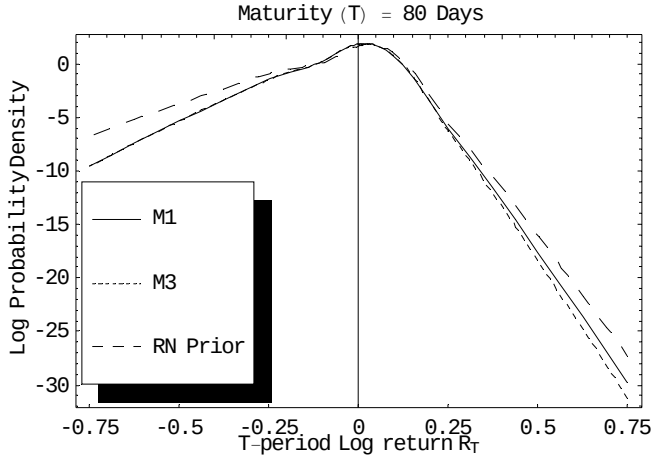


Figure 4.16: Plot of calibrated log-return probability density and Lévy measure on November 22, 2004 (80 days to maturity)

A) Calibrated log return probability density.



B) Log of Panel A).



C) Calibrated Lévy measure.

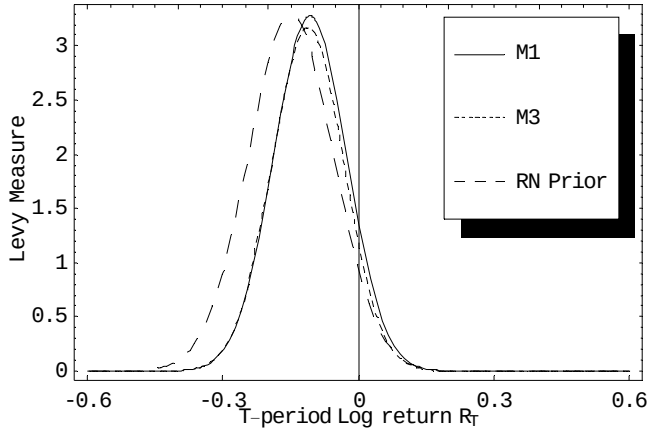
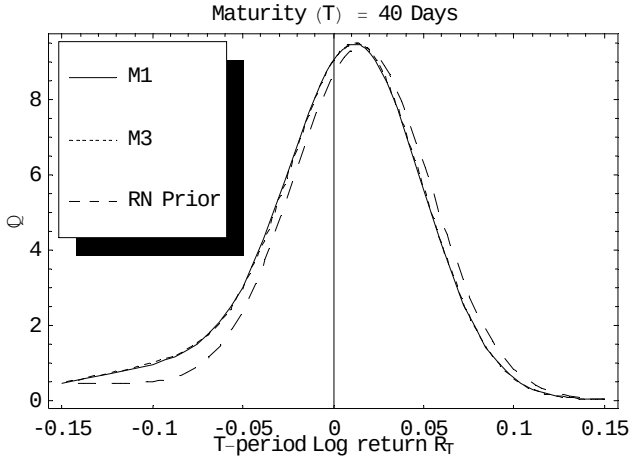
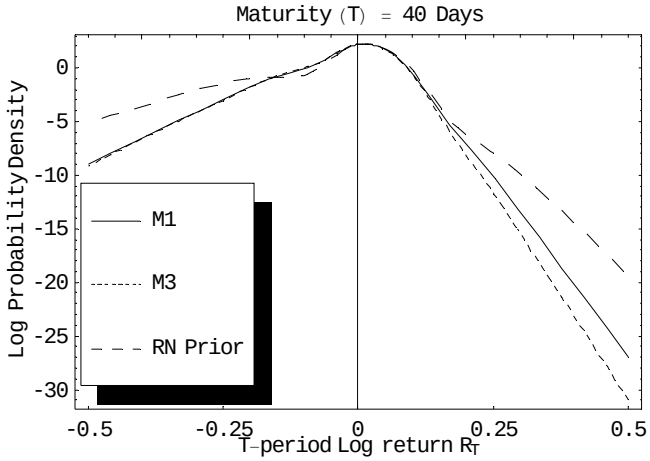


Figure 4.17: Plot of calibrated log-return probability density and Lévy measure on January 20, 2005 (40 days to maturity)

A) Calibrated log return probability density.



B) Log of Panel A).



C) Calibrated Lévy measure.

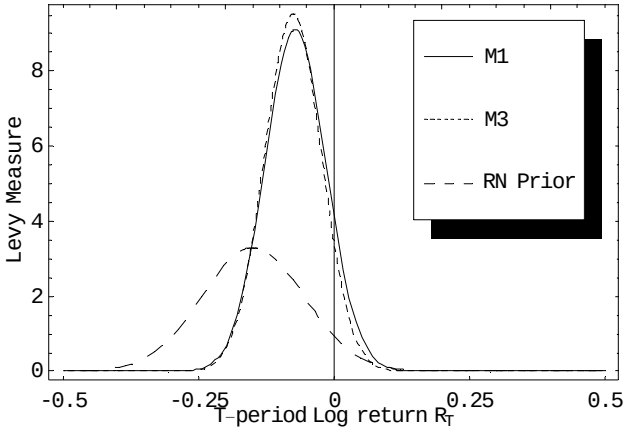
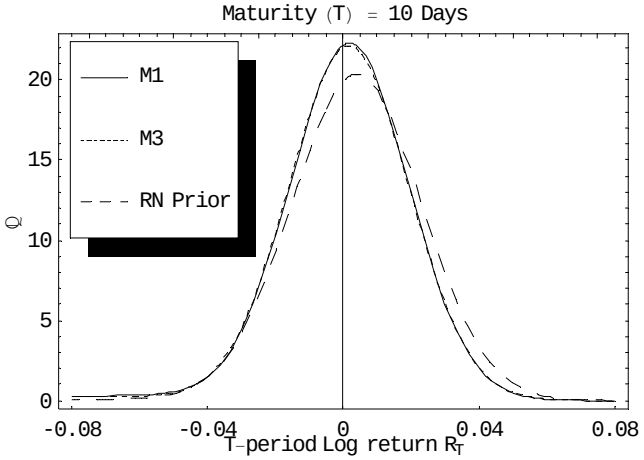
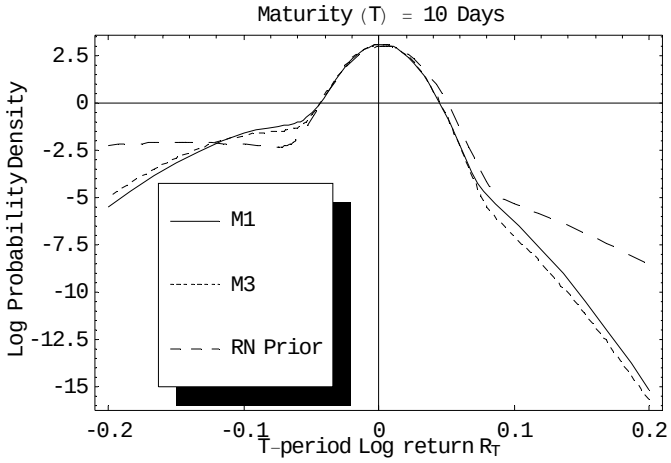


Figure 4.18: Plot of calibrated log-return probability density and Lévy measure on March 4, 2005 (10 days to maturity)

A) Calibrated log return probability density.



B) Log of Panel A).



C) Calibrated Lévy measure.

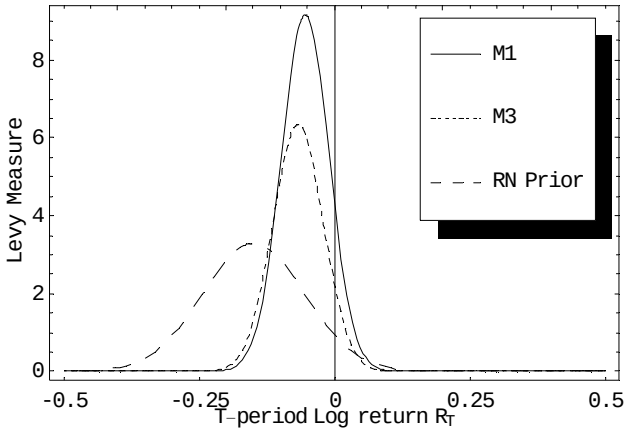
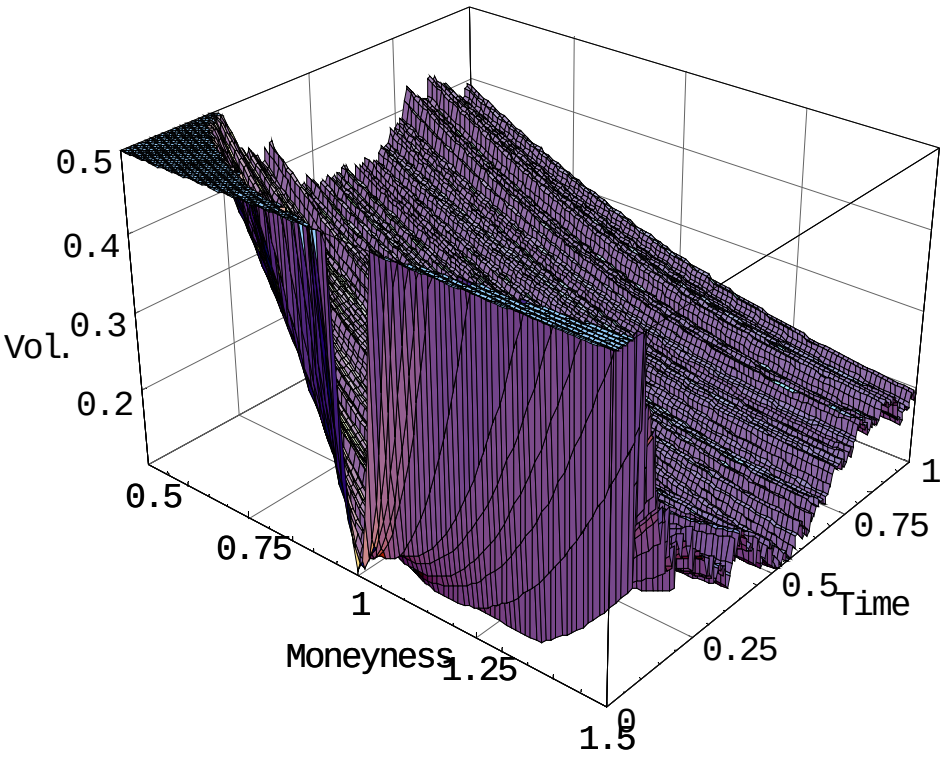
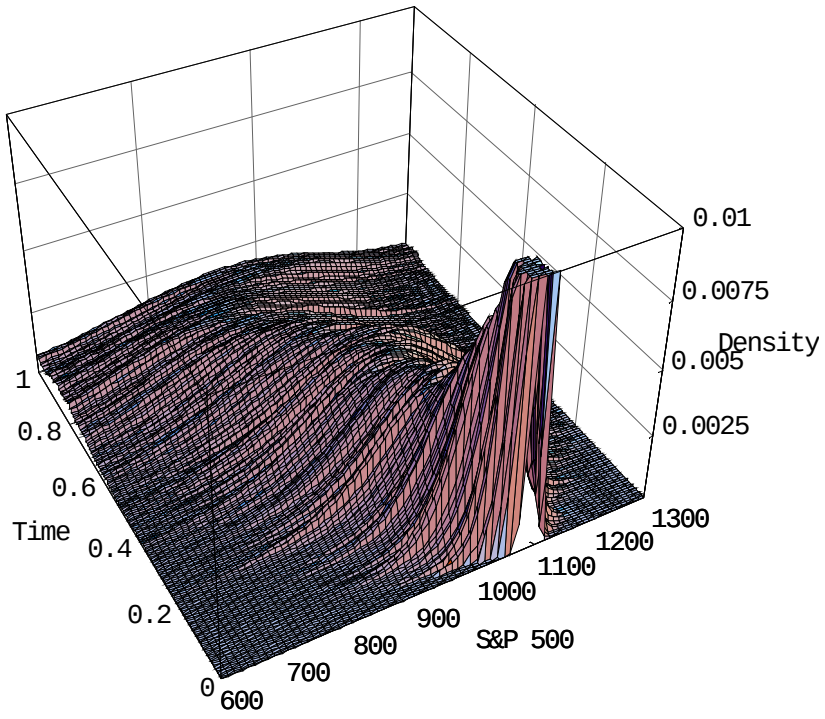


Figure 5.1: Implied volatility surface



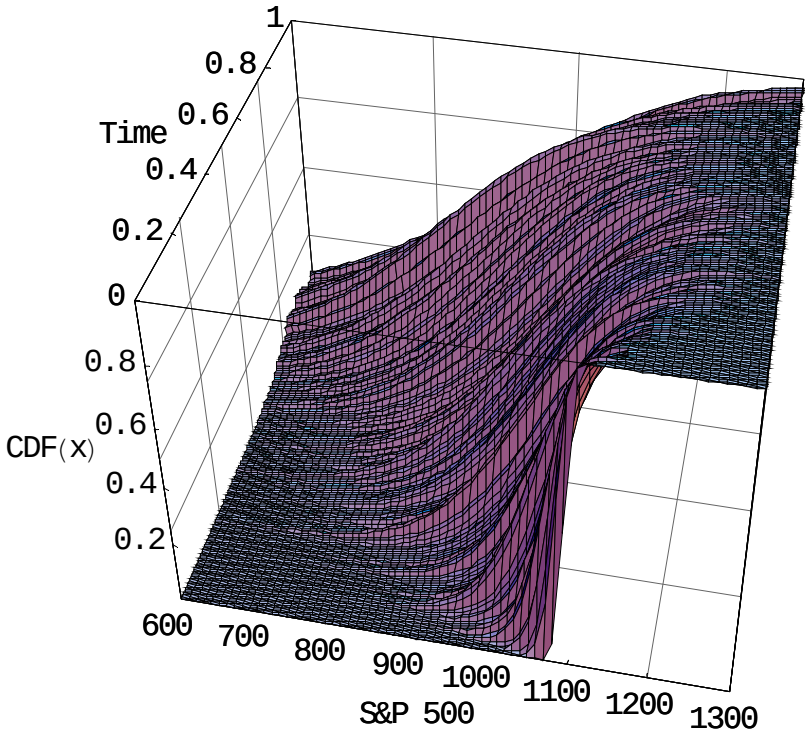
Note: Plot of the dynamics of implied volatility curves of futures options on S&P 500 with December 2003 maturity as a function of time to maturity in years and moneyness = Strike/Futures for the period between December 23 2002 and December 18 2003. Quadratic function is used as a method of interpolation.

Figure 5.2: Dynamics of risk-neutral densities obtained from volatility interpolation method



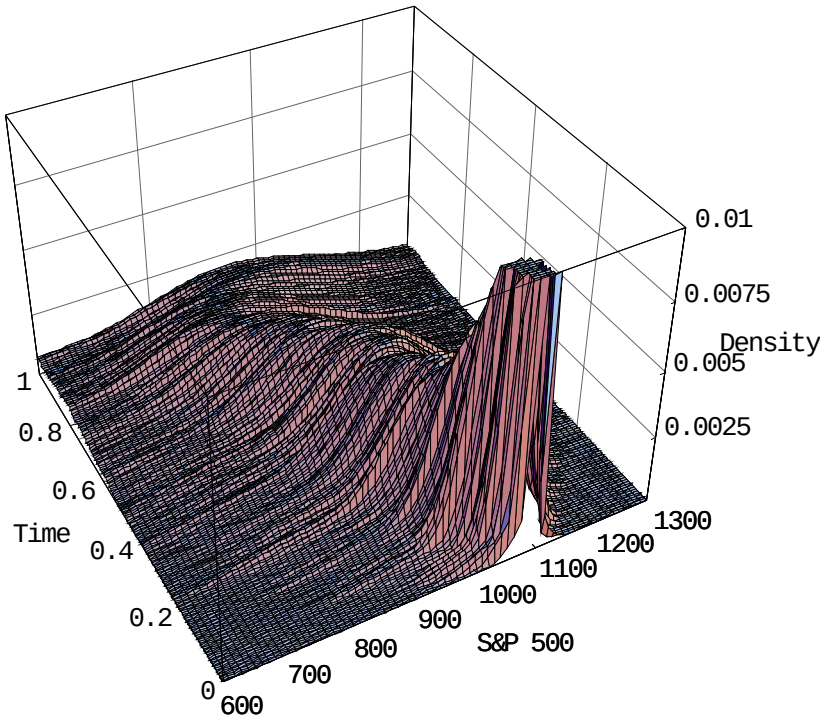
Note: Plot of the dynamics of risk-neutral densities of futures options on S&P 500 with December 2003 maturity obtained from volatility interpolation method for the period between December 23 2002 and December 18 2003. Time to maturity is measured in years. Range of the density plotted is between 0 and 0.01.

Figure 5.3: Dynamics of cumulative risk-neutral densities obtained from volatility interpolation method



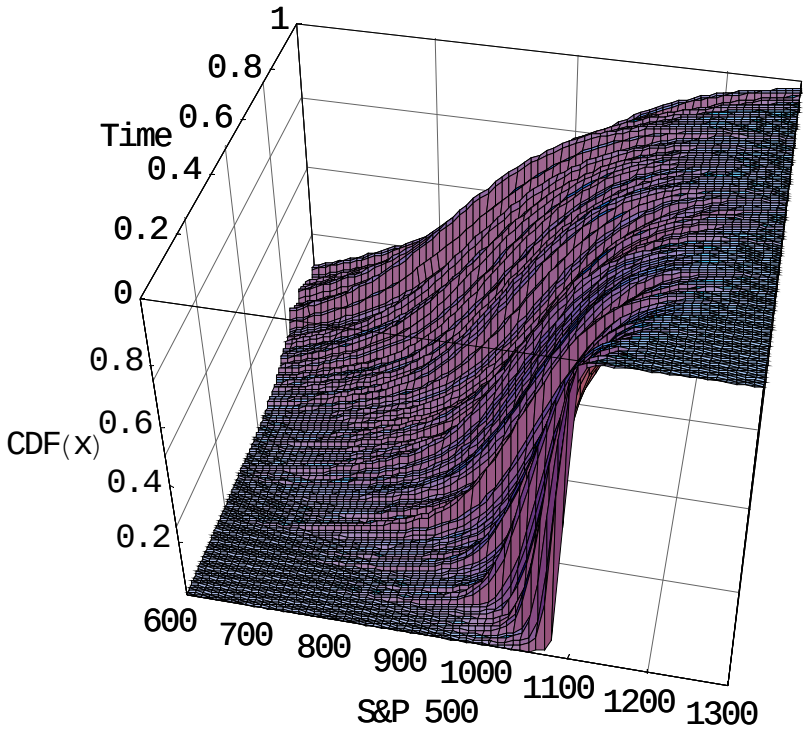
Note: Plot of the dynamics of cumulative risk-neutral densities of futures options on S&P 500 with December 2003 maturity obtained from volatility interpolation method for the period between December 23 2002 and December 18 2003. Time to maturity is measured in years.

Figure 5.4: Dynamics of risk-neutral densities obtained from mixture of two lognormals method



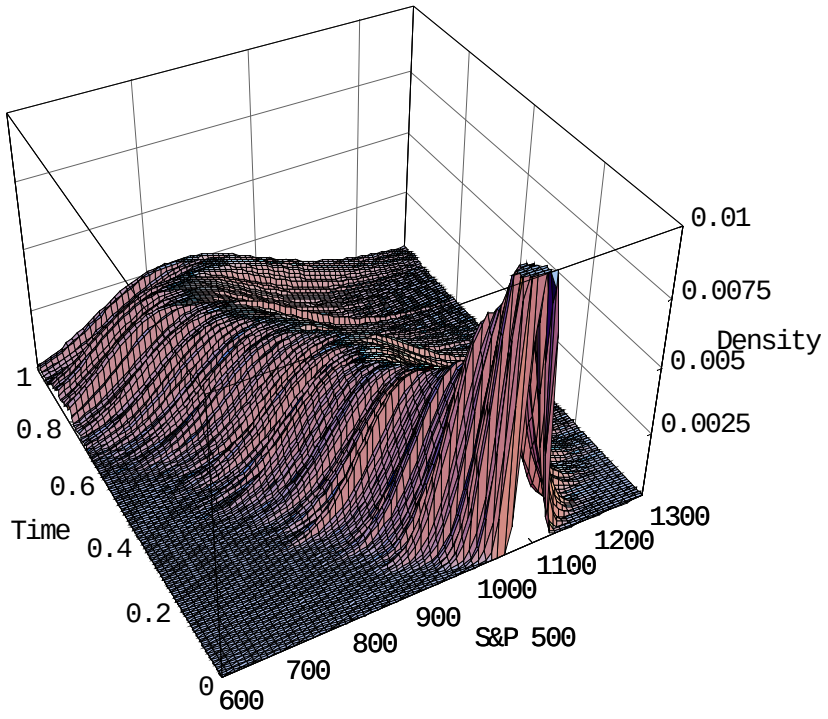
Note: Plot of the dynamics of risk-neutral densities of futures options on S&P 500 with December 2003 maturity obtained from mixture of two lognormals method for the period between December 23 2002 and December 18 2003. Time to maturity is measured in years. Range of the density plotted is between 0 and 0.01.

Figure 5.5: Dynamics of cumulative risk-neutral densities obtained from mixture of two lognormals method



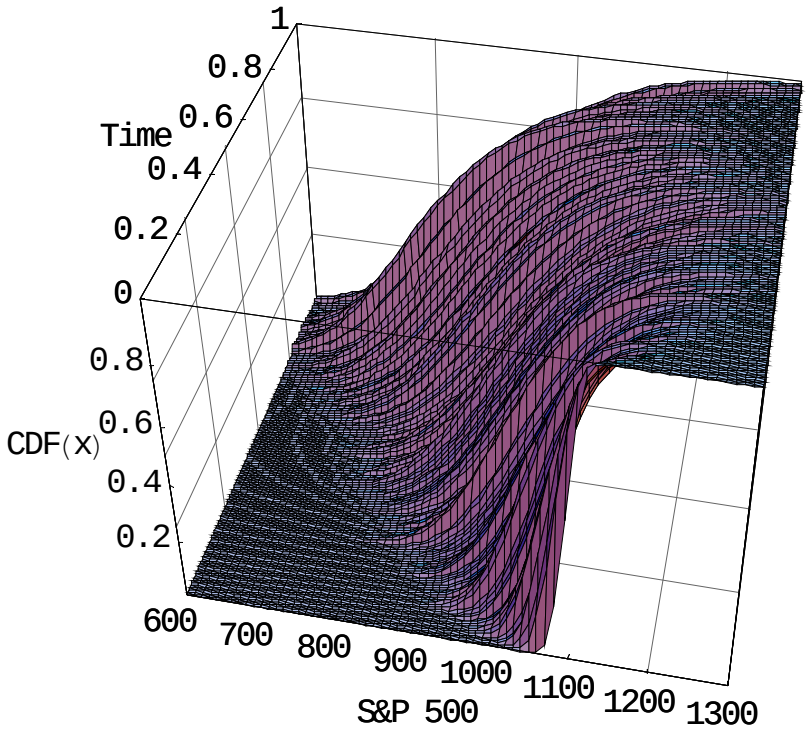
Note: Plot of the dynamics of cumulative risk-neutral densities of futures options on S&P 500 with December 2003 maturity obtained from mixture of two lognormals method for the period between December 23 2002 and December 18 2003. Time to maturity is measured in years.

Figure 5.6: Dynamics of risk-neutral densities obtained from the benchmark BS lognormal



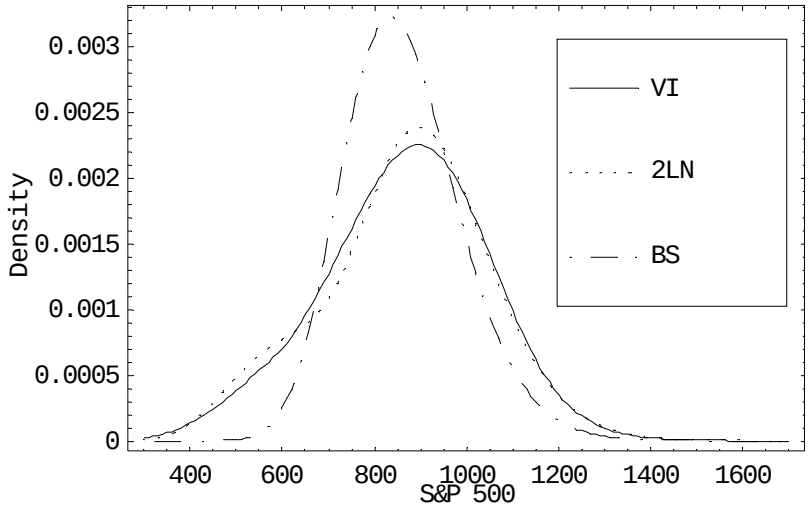
Note: Plot of the dynamics of risk-neutral densities of futures options on S&P 500 with December 2003 maturity obtained from a benchmark Black-Scholes lognormal RND for the period between December 23 2002 and December 18 2003. Time to maturity is measured in years. Range of the density plotted is between 0 and 0.01.

Figure 5.7: Dynamics of cumulative risk-neutral densities obtained from the benchmark Black-Scholes lognormal



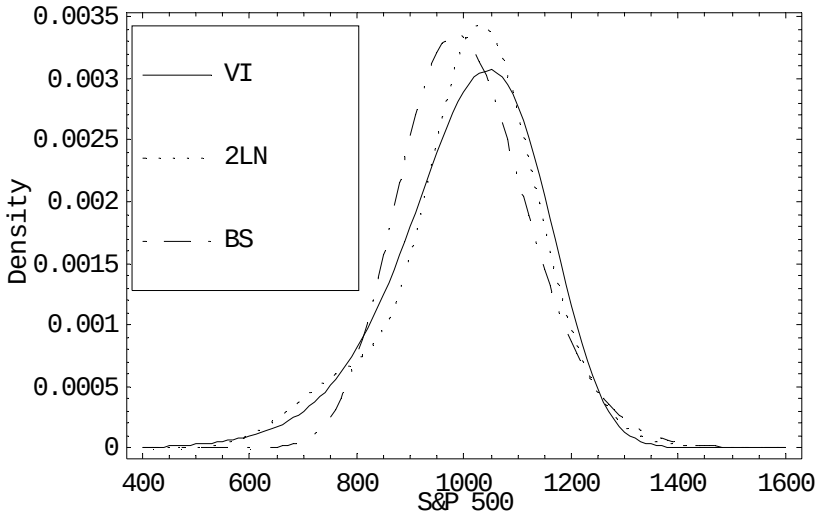
Note: Plot of the dynamics of cumulative risk-neutral densities of futures options on S&P 500 with December 2003 maturity obtained from a benchmark Black-Scholes lognormal RND for the period between December 23 2002 and December 18 2003. Time to maturity is measured in years.

Figure 5.8: Risk-neutral density estimates with 9-month to maturity



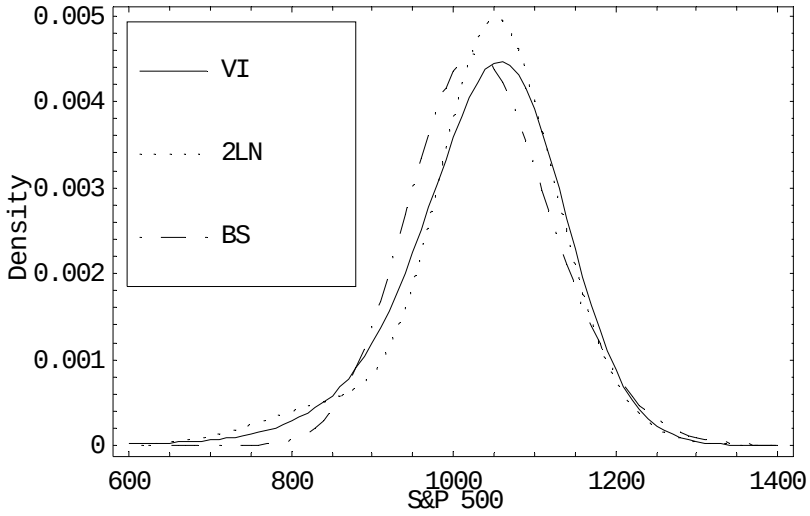
Note: Plot of risk-neutral density estimates of futures options on S&P 500 with December 2003 maturity on March 18 2003 obtained from three different methods.

Figure 5.9: Risk-neutral density estimates with 6-month to maturity



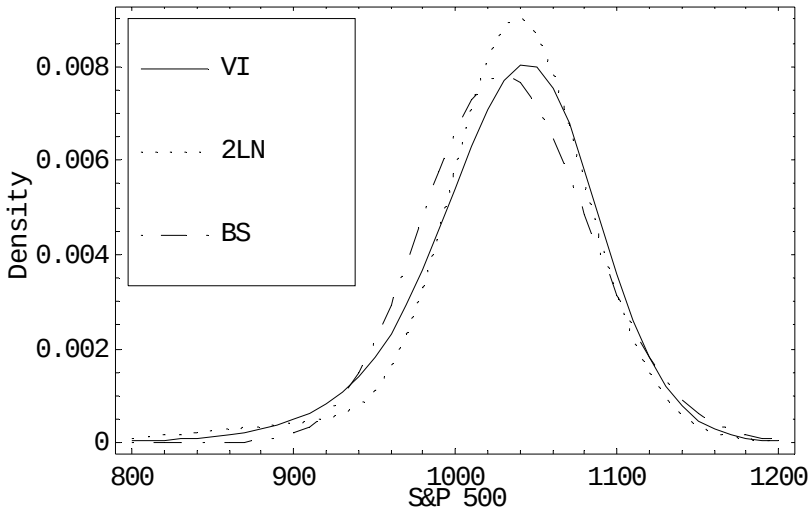
Note: Plot of risk-neutral density estimates of futures options on S&P 500 with December 2003 maturity on June 18 2003 obtained from three different methods.

Figure 5.10: Risk-neutral density estimates with 3-month to maturity



Note: Plot of risk-neutral density estimates of futures options on S&P 500 with December 2003 maturity on September 18 2003 obtained from three different methods.

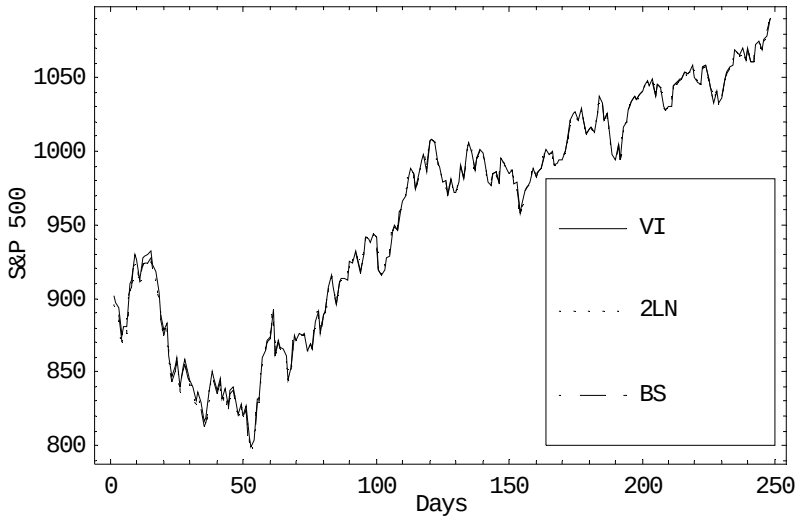
Figure 5.11: Risk-neutral density estimates with 1-month to maturity



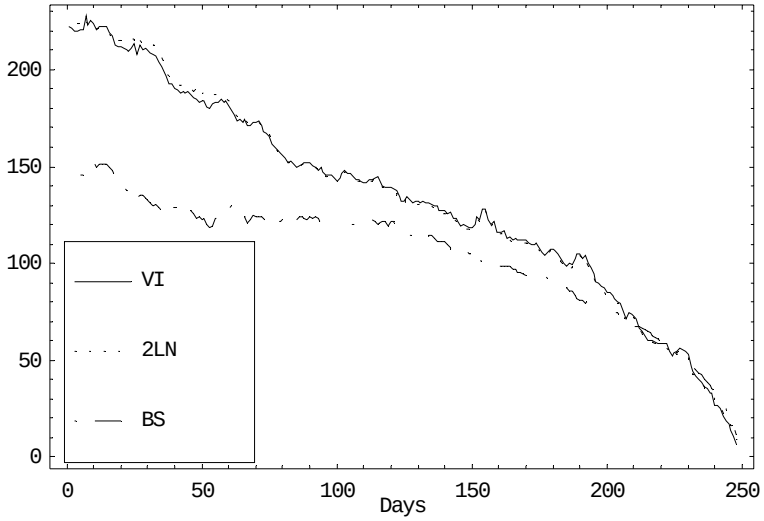
Note: Plot of risk-neutral density estimates of futures options on S&P 500 with December 2003 maturity on November 18 2003 obtained from three different methods.

Figure 5.12: Dynamics of moments of risk-neutral density estimates

A) Mean



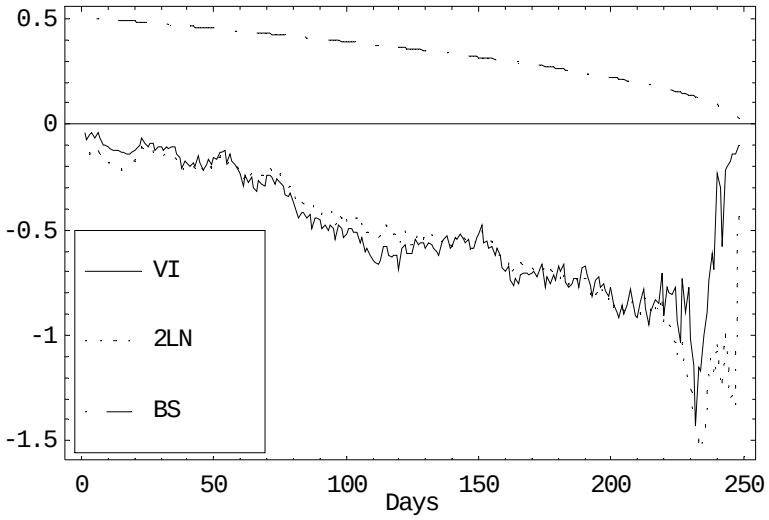
B) Standard Deviation



Note: Plot of the term structure of mean, standard deviation, skewness, and kurtosis of risk-neutral density estimates of futures options on S&P 500 with December 2003 maturity for the period between December 23 2002 and December 18 2003.

Figure 5.12: Dynamics of moments of risk-neutral density estimates (Continued)

C) Skewness



D) Kurtosis

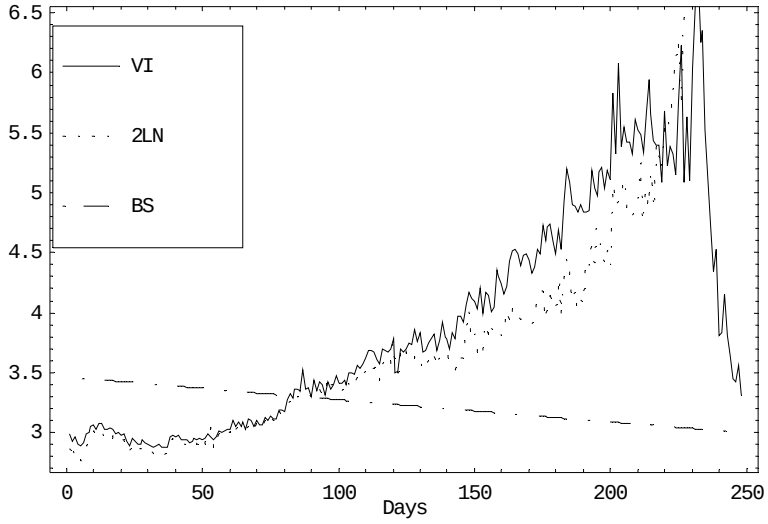
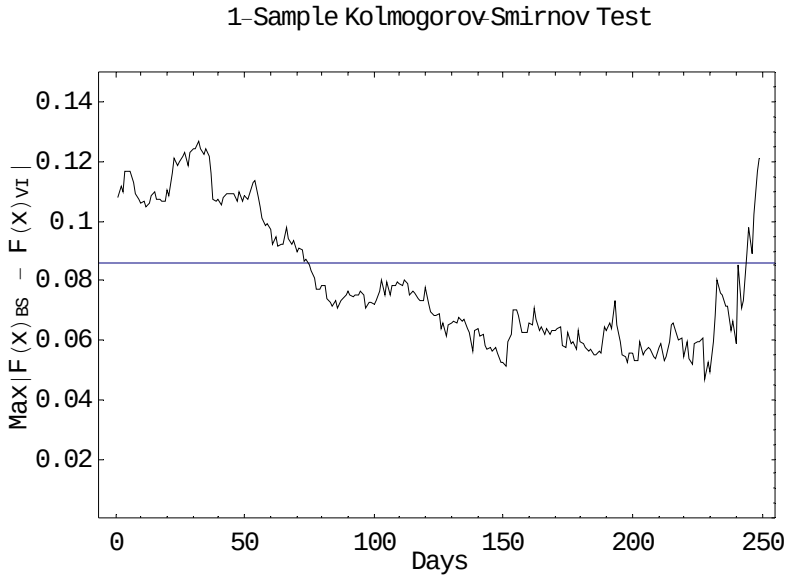
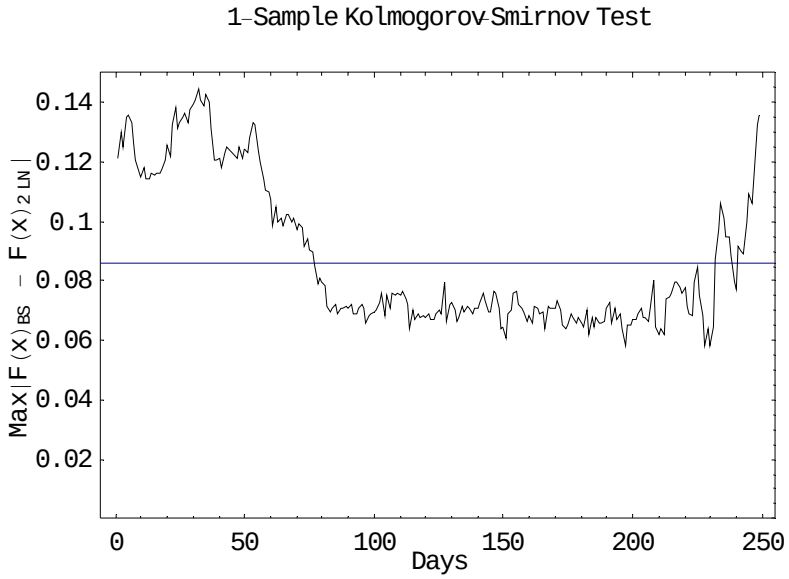


Figure 5.13: Kolmogorov-Smirnov test of risk-neutral density estimates between volatility interpolation and benchmark BS



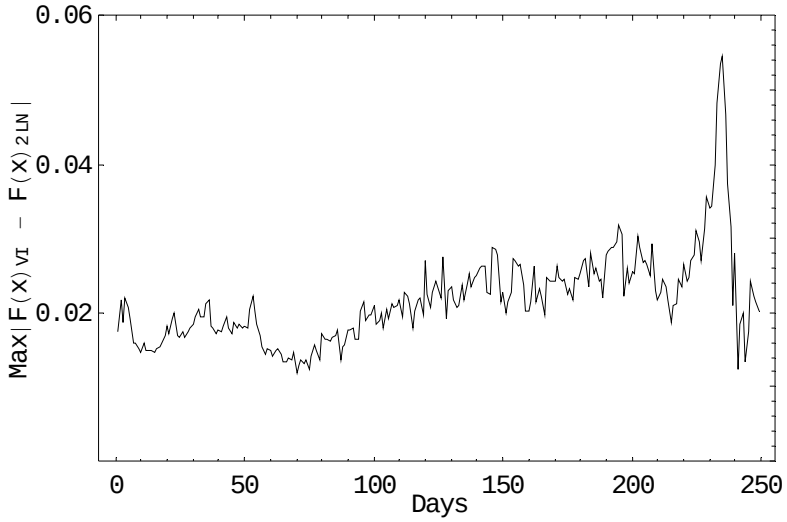
Note: Plot of Test Statistic D Value $Max\left|F_{BS}(x)-F_{VI}(x)\right|$ which is the largest absolute deviation of the cumulative distribution functions obtained from volatility interpolation and Black-Scholes lognormal methods for the period between December 23 2002 and December 18 2003. Critical value is 0.0861865.

Figure 5.14: Kolmogorov-Smirnov test of risk-neutral density estimates between mixture of two lognormals and benchmark BS.



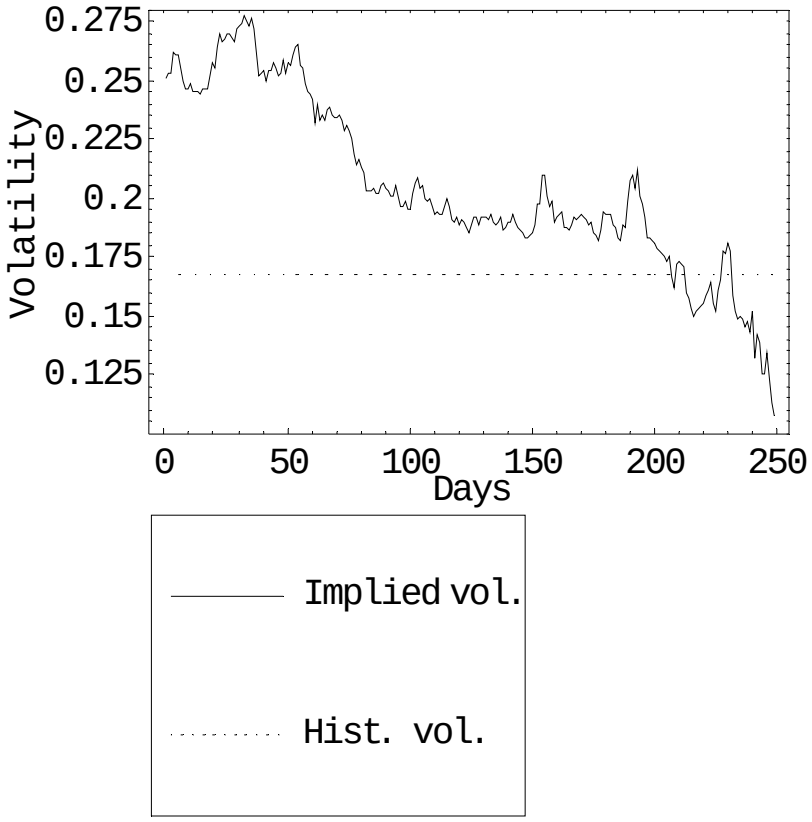
Note: Plot of Test Statistic D Value $Max\left|F_{BS}(x)-F_{2LN}(x)\right|$ which is the largest absolute deviation of the cumulative distribution functions obtained from mixture of two lognormlas and Black-Scholes lognormal methods for the period between December 23 2002 and December 18 2003. Critical value is 0.0861865.

Figure 5.15: Kolmogorov-Smirnov test statistic of risk-neutral density estimates between volatility interpolation and mixture of two lognormals methods



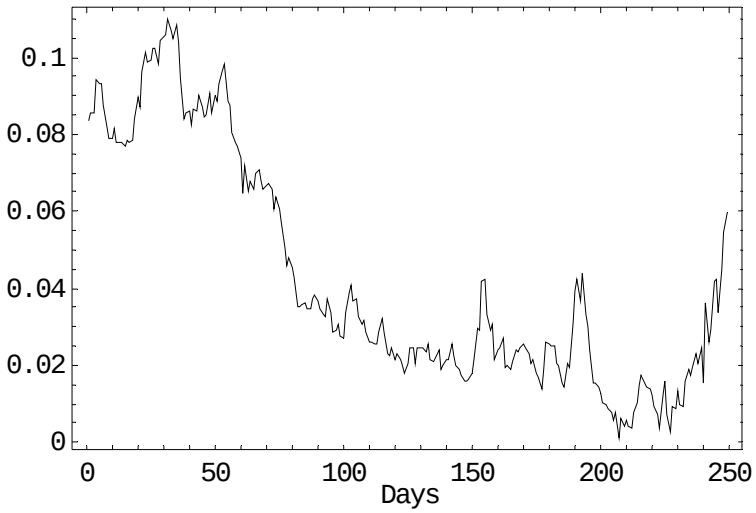
Note: Plot of Test Statistic D Value $Max\left|F_{VI}(x)-F_{2LN}(x)\right|$ which is the largest absolute deviation of the cumulative distribution functions obtained from volatility interpolation and mixture of two lognormals methods for the period between December 23 2002 and December 18 2003. Critical value is 0.0861865.

Figure 5.16: Historical volatility and at-the-money implied volatility



Note: Plot of the historical volatility used to obtain the benchmark Black-Scholes lognormal dynamics and at the money implied volatilities incorporated in the option implied dynamics for the sample period between December 23 2002 and December 18 2003.

Figure 5.17: Absolute deviation between historical volatility and at-the-money implied volatility



Note: Plot of the absolute value of the difference between historical volatilities used to obtain the benchmark Black-Scholes lognormal dynamics and at the money implied volatilities incorporated in the option implied dynamics for the sample period between December 23 2002 and December 18 2003.

Appendix

A.1 Derivation of Fokker-Planck (Forward Kolmogorov) Equation

This section is based on Gnedenko (1962). A stochastic process $\xi(t)$ is said to be continuous if there is only a small probability that $\xi(t)$ will take on an appreciable increment in a short interval of time. This means that for any positive constant δ:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x_T - x_t| \geq \delta} d_{x_T} F(x_T, T; x_t, t) = 0 \qquad \text{(a-1)},$$

where $\Delta t = T - t$. Make the following assumptions:

1) The partial derivatives $\partial F(x_T, T; x_t, t)/\partial x_t$ and $\partial^2 F(x_T, T; x_t, t)/\partial x_t^2$ exist and are continuous for arbitrary values of t, $T > t$, x_t, and x_T.

2) For any $\delta > 0$, the following limits exist and the convergence is uniform in x_t:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x_T - x_t| < \delta} (x_T - x_t) d_{x_T} F(x_T, T; x_t, t) = a(t, x_t) \qquad \text{(a-2)},$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x_T - x_t| < \delta} (x_T - x_t)^2 d_{x_T} F(x_T, T; x_t, t) = b^2(t, x_t) \qquad \text{(a-3)}.$$

3) A probability density function exists:

$$f(x_T, T; x_t, t) = \frac{\partial F(x_T, T; x_t, t)}{\partial x_T} \qquad \text{(a-4)}.$$

4) The following derivatives exist and are continuous:

$$\partial f(x_T, T; x_t, t)/\partial T,$$

$$\frac{\partial}{\partial x_T}[a(T, x_T) f(x_T, T; x_t, t)],$$

$$\frac{\partial^2}{\partial {x_T}^2}[b(T, x_T) f(x_T, T; x_t, t)] \quad \text{(a-5)}.$$

Fokker-Planck Equation (Forward Kolmogorov Equation): *For any continuous stochastic process without after effect satisfying conditions 1) through 4), the probability density function* $f(x_T, T; x_t, t)$ *is a solution of the equation:*

$$\frac{\partial f(x_T, T; x_t, t)}{\partial T} = -\frac{\partial (a(x_T, T) f(x_T, T; x_t, t))}{\partial x_T} + \frac{1}{2}\frac{\partial^2 (b(x_T, T)^2 f(x_T, T; x_t, t))}{\partial {X_T}^2} \quad \text{(a-6)}.$$

Derivation: Let c and d $(c < d)$ denote certain numbers and $R(x_T)$ a nonnegative continuous function having continuous first- and second-order derivatives. Assume:

$$R(x_T) = 0 \quad \text{for} \quad x_T < c \quad \text{and} \quad x_T > d.$$

Due to the condition that the function $R(x_T)$ and its derivatives are continuous,

$$R(c) = R(d) = R'(c) = R'(d) = R''(c) = R''(d) \quad \text{(a-7)}.$$

Note:

$$\int_c^d \frac{\partial f(x_T, T; x_t, t)}{\partial T} R(x_T) dx_T$$

$$= \frac{\partial}{\partial T}\int_c^d f(x_T, T; x_t, t) R(x_T) dx_T$$

$$= \lim_{\Delta T \to 0} \int \frac{f(x_T, T + \Delta T; x_t, t) - f(x_T, T; x_t, t)}{\Delta T} R(x_T) dx_T \quad \text{(a-8)}.$$

Apply the generalized Markov equation:

$$f(x_T, T+\Delta T; x_t, t) = \int f(z, T; x_t, t) f(x_T, T+\Delta T; z, T)\,.$$

Thus, the equation (a-8) can be written down as:

$$\int_c^d \frac{\partial f(x_T, T; x_t, t)}{\partial T} R(x_T) dx_T$$

$$= \lim_{\Delta T \to 0} \frac{1}{\Delta T} [\iint f(z, T; x_t, t) f(x_T, T+\Delta T; z, T) R(x_T) dz dx_T - \int f(x_T, T; x_t, t) R(x_T) dx_T]$$

$$= \lim_{\Delta T \to 0} \frac{1}{\Delta T} [\int f(z, T; x_t, t) \int f(x_T, T+\Delta T; z, T) R(x_T) dx_T dz - \int f(x_T, T; x_t, t) R(x_T) dx_T]$$

$$= \lim_{\Delta T \to 0} \frac{1}{\Delta T} \int f(x_T, T; x_t, t) [\int f(z, T+\Delta T; x_T, T) R(z) dz - R(x_T)] dx_T \qquad \text{(a-9).}$$

First, the order of integration is interchanged. Second, the notation for the variables of integration is changed (Replace x_T by z and z by x_T). By Taylor's theorem:

$$R(z) = R(x_T) + (z - x_T) R'(x_T) + \frac{1}{2}(z - x_T)^2 R''(x_T) + o[(z - x_T)^2]\,.$$

By the bounded nature of the function $R(z)$ and the condition 1):

$$\int_{|x_T - z| \geq \delta} f(z, T+\Delta T; x_T, T) R(z) dz = o(\Delta T)\,,$$

$$\int_{|x_T - z| \leq \delta} f(z, T+\Delta T; x_T, T) R(z) dz = 1 + o(\Delta T)\,.$$

It follows that:

$$\int f(z, T+\Delta T; x_T, T) R(z) dz - R(x_T) = R'(x_T) \int_{|x_T - z| < \delta} (z - y) \int f(z, T+\Delta T; x_T, T) dz$$

$$+\frac{1}{2}R''(x_T)\int_{|x_T-z|<\delta}[(z-x_T)^2+o((z-x_T)^2)]f(z,T+\Delta T;x_T,T)dz+o(\Delta T).$$

Thus:

$$\int_c^d \frac{\partial f(x_T,T;x_t,t)}{\partial T}R(x_T)dx_T =$$

$$\lim_{\Delta T\to 0}\frac{1}{\Delta T}\int f(x_T,T;x_t,t)\{R'(x_T)\int_{|x_T-z|<\delta}(z-y)f(z,T+\Delta T;x_T,T)dz$$

$$+\frac{1}{2}R''(x_T)\int_{|x_T-z|<\delta}[(z-x_T)^2+o((z-x_T)^2)]f(z,T+\Delta T;x_T,T)dz+o(\Delta T)\}dx_T.$$

Let $\Delta t\to 0$. Following the assumption that the limits in 2) and 3) are uniform in x_t, the limit equation of the above can be expressed as:

$$\int_c^d \frac{\partial f(x_T,T;x_t,t)}{\partial T}R(x_T)dx_T =\int f(x_T,T;x_t,t)[a(T,x_T)R'(x_T)+\frac{1}{2}b^2(T,x_T)R''(x_T)]dx_T.$$

Because $R'(x_T)=R''(x_T)=0$ for $x_T\le c$ and $x_T\ge d$:

$$\int_c^d \frac{\partial f(x_T,T;x_t,t)}{\partial T}R(x_T)dx_T =\int_c^d f(x_T,T;x_t,t)[a(T,x_T)R'(x_T)+\frac{1}{2}b^2(T,x_T)R''(x_T)]dx_T$$

(a-10).

Use integration by parts and the equation (a-7),

$$\int_c^d f(x_T,T;x_t,t)a(T,x_T)R'(x_T)dx_T =-\int_c^d R(x_T)\frac{\partial}{\partial x_T}[a(T,x_T)f(x_T,T;x_t,t)]dx_T$$

$$\int_c^d f(x_T,T;x_t,t)b^2(T,x_T)R''(x_T)dx_T =\int_c^d R(x_T)\frac{\partial^2}{\partial {x_T}^2}[b^2(T,x_T)f(x_T,T;x_t,t)]dx_T$$

Substitute these expressions to the equation (a-10):

$$\int_c^d \frac{\partial f(x_T, T; x_t, t)}{\partial T} R(x_T) dx_T$$

$$= \int_c^d \{-\frac{\partial}{\partial x_T}[a(T, x_T) f(x_T, T; x_t, t)] + \frac{1}{2}\frac{\partial^2}{\partial x_T{}^2}[b^2(T, x_T) f(x_T, T; x_t, t)]\} R(x_T) dx_T .$$

This equation can be written in the following form:

$$\int_c^d \{\frac{\partial f(x_T, T; x_t, t)}{\partial T} + \frac{\partial}{\partial x_T}[a(T, x_T) f(x_T, T; x_t, t)]$$

$$-\frac{1}{2}\frac{\partial^2}{\partial x_T{}^2}[b^2(T, x_T) f(x_T, T; x_t, t)]\} R(x_T) dx_T = 0 \qquad \text{(a-11).}$$

Fokker-Planck equation follows from (a-11) since the function $R(x_T)$ is arbitrary.

A.2 Barone-Adesi and Whaley (1987) Quadratic Approximation Method to Adjust for the Early Exercise Premium

Consider a portfolio P of the one long (European or American) option position $V(S,t)$ on a stock S with continuously compounded dividend yield q written at time t and a short position of the stock in quantity Δ:

$$P = V(S,t) - \Delta S \qquad \text{(a-12).}$$

The underlying asset price dynamics follows usual geometric Brownian motion:

$$dS = (\mu - q) S dt + \sigma S dW \qquad \text{(a-13).}$$

Portfolio value changes by:

$$dP = dV - \Delta dS \qquad \text{(a-14).}$$

From Ito's lemma, the change in the value of the option is written as:

$$dV = (\frac{\partial V}{\partial S}(\mu - q) S + \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2) dt + \frac{\partial V}{\partial S}\sigma S dW \qquad \text{(a-15).}$$

Substitution of (a-13) and (a-15) into (a-14) gives the change in the portfolio value as:

$$dP = (\frac{\partial V}{\partial S}(\mu - q)S + \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2)dt + \frac{\partial V}{\partial S}\sigma S dW - \Delta\{(\mu - q)Sdt + \sigma S dW\}.$$

After rearrangement:

$$dP = [(\frac{\partial V}{\partial S} - \Delta)(\mu - q)S + \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2]dt + (\frac{\partial V}{\partial S} - \Delta)\sigma S dW \qquad \text{(a-16)}.$$

Choosing $\Delta = \partial V / \partial S$ makes the portfolio risk-free since randomness dW is eliminated:

$$dP = (\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2)dt \qquad \text{(a-17)}.$$

In an infinitesimal time interval dt, the portfolio holder earns capital gains equal to dP and loses dividends on the stock position (since the portfolio holder has short position in stock):

$$qS\frac{\partial V}{\partial S}dt \qquad \text{(a-18)}.$$

This portfolio is expected to grow at the risk-free interest rate r:

$$(\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2)dt - qS\frac{\partial V}{\partial S}dt = rPdt \qquad \text{(a-19)}.$$

Substitution yields Black-Scholes PDE:

$$\frac{\partial V}{\partial t} + (r - q)S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} = rV \qquad \text{(a-20)}.$$

Let ρ denote an early exercise premium which is the difference between American style and European style option price written on the same underlying with

same maturity. ρ satisfies the Black-Scholes PDE because both American style and European style option price satisfy it:

$$\frac{\partial \rho}{\partial t} + (r-q)S\frac{\partial \rho}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \rho}{\partial S^2} = r\rho \qquad \text{(a-21)}.$$

The "quasi-stationary method" is a standard method for obtaining approximate solutions to differential equations of this type. Apply a change of variable technique of the following, set $\rho = h(\tau)g(S,h)$, and ignore the time-dependence ($\partial\rho/\partial t = 0$):

$$\tau = T - t,\ h(\tau) = 1 - e^{-r\tau},\ k_1 = \frac{2r}{\sigma^2},\ k_2 = \frac{2(r-q)}{\sigma^2}$$

Equation (a-21):

$$\frac{1}{2}S^2\frac{\partial^2 \rho}{\partial S^2}\sigma^2 - r\rho + (r-q)S\frac{\partial \rho}{\partial S} = 0,$$

can now be written down as:

$$S^2\frac{\partial^2 g}{\partial S^2} + k_2 S\frac{\partial g}{\partial S} - \frac{k_1 g}{h} - (1-h)k_1\frac{\partial g}{\partial h} = 0 \qquad \text{(a-22)}.$$

With quasi-stationary approximation $(1-h)k_1\,\partial g/\partial h = 0$:

$$S^2\frac{\partial^2 g}{\partial S^2} + k_2 S\frac{\partial g}{\partial S} - \frac{k_1 g}{h} = 0 \qquad \text{(a-23)}.$$

Equation (a-23) is an equation of homogeneous type, which is easily solved in terms of a power of S. Let $C_{Ameri}(S,t)$ and $P_{Ameri}(S,t)$ denote the American style call and put option prices. Let $C_{Euro}(S,t)$ and $P_{Euro}(S,t)$ denote the European style call and put option prices. After applying boundary conditions to a solution of equation (a-23):

$$C_{Ameri}(S,t)=\begin{cases} C_{Euro}(S,t)+A_2(\frac{S}{B^*})^{\gamma_2} & \text{if } S<B^* \\ S-K & \text{if } S\geq B^* \end{cases} \quad \text{(a-24)},$$

$$P_{Ameri}(S,t)=\begin{cases} P_{Euro}(S,t)+A_1(\frac{S}{B^{**}})^{\gamma_1} & \text{if } S>B^{**} \\ K-S & \text{if } S\leq B^{**} \end{cases} \quad \text{(a-25)}.$$

B^* is the critical stock price above which the call option should be exercised and B^{**} is the critical stock price below which the put option should be exercised. These can be estimated by solving the following equations:

$$B^*-K=C_{Euro}(B^*,t)+[1-e^{-q(T-t)}N(d_1(B^*))]\frac{B^*}{\gamma_2} \quad \text{(a-26)},$$

$$K-B^{**}=P_{Euro}(B^{**},t)-[1-e^{-q(T-t)}N(-d_1(B^{**}))]\frac{B^{**}}{\gamma_1} \quad \text{(a-27)}.$$

Note in above equations:

$$\gamma_1=\frac{1}{2}(1-K_2-\sqrt{(1-K_2)^2+\frac{4K_1}{h}}),$$

$$\gamma_2=\frac{1}{2}(1-K_2+\sqrt{(1-K_2)^2+\frac{4K_1}{h}}).$$

$$A_1=-(\frac{B^{**}}{\gamma_1})[1-e^{-q(T-t)}N(-d_1(B^{**}))]$$

$$A_2=(\frac{B^*}{\gamma_2})[1-e^{-q(T-t)}N(d_1(B^*))]$$

$$d_1(S)=\frac{\ln(S/K)+(r-q+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \quad \text{(a-28)}.$$

Bibliography

Ait-Sahalia, Y., and A. Lo. "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices." *Journal of Finance,* 53 (1998), 499-547.

Avellaneda, M., Friedman, C., Holmes, R., and D. Samperi. "Calibrating Volatility Surfaces via Relative-Entropy Minimization." *Applied Mathematical Finance*, 4 (1997), 37-64.

Applebaum, D. Lévy Processes and Stochastic Calculus, Cambridge University Press (2004).

Bahra, B. "Implied Risk-Neutral Probability Density Functions From Option Prices: Theory and Application." Working Paper No.66, Bank of England (1997).

Bakshi, G., Cao, C., and Z. Chen. "Empirical Performance of Alternative Option Pricing Models." *Journal of Finance*, 52 (1997), 2003-49.

Ball, C. A., and W. N. Torous. "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing." *Journal of Finance*, 40 (1985), 155-173.

Barndorff-Nielsen, O. E. "Exponentially Decreasing Distributions for the Logarithm of the Particle Size." *Proceedings of the Royal Society of London*, A353 (1977), 401-419.

Barndorff-Nielsen, O. E., and O. Halgreen. "Infinite Divisibility of the Hyperbolic and Generalized Inverse Gaussian Distributions." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 38 (1977), 309-312.

Barndorff-Nielsen, O. E. "Processes of Normal Inverse Gaussian Type." *Finance and Stochastics*, 2 (1998), 41-68.

Barndorff-Nielsen, O. E. ed. Lévy Processes: Theory and Applications. Birkhäuser (2001).

Barone-Adesi, G., and R. E. Whaley. "Efficient Analytic Approximation of American option values." *Journal of Finance,* 42 (1987), 301-320.

Bates, D. “The Crash of 87: Was it Expected? The Evidence from Options Markets.” *Journal of Finance*, 46 (1991), 1009-1044.
Bates, D. “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutschemark Options.” *Review of Financial Studies*, 9 (1996), 69-108.

Black, F., and M. Scholes. “The Pricing of Options and Corporate Liabilities.” *Journal of Political Economy*, 81 (1973), 637-654.

Brzezniak, Z., and T. Zastawniak. Basic Stochastic Processes, Springer (1999).

Breeden, D. T., and R. H. Litzenberger. "Prices of State-Contingent Claims Implicit in Option Prices." *Journal of Business*, 51 (1978), 621-651.

Brigham, E. O. The Fast Fourier Transform and Its Applications, Prentice Hall (1988).

Campa, J., Chang, P., and R. Reider. “Implied Exchange Rate Distributions: Evidence from OTC Option Markets.” *Journal of International Money and Finance* 17 (1998), 117-160.

Capinski, M., Kopp, E., and P. E. Kopp. Measure, Integral and Probability, Springer-Verlag (2004).

Carr, P., Chang, E. C., and D. B. Madan. “The Variance Gamma Process and Option Pricing.” *European Finance Review*, 2 (1998), 79-105.

Carr, P., Geman, H., Madan, D. B., and M. Yor. “The Fine Structure of Asset Returns: An Empirical Investigation.” *Journal of Business*, vol 75, issue 2 (2002), 305-332.

Carr, P., and D.B. Madan. “Option Valuation Using the Fast Fourier Transform.” *Journal of Computational Finance*, 2 (1998), 61-73.

Carr, P., Geman, H., Madan, D. B., and M. Yor. “Option Pricing Using Integral Transforms”.

Cont, R., Bouchaud, J. P., and M. Potters. “Scaling in Stock Market Data: Stable Laws and Beyond,” Appeared in Dubrulle, B., Graner, F., and Sornette, D., Scale Invariance and Beyond, Springer Berlin (1997).

Cont, R. “Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues.” *Quantitative Finance*, 1 (2001), 1-14.

Cont, R., and J. da Fonseca . “Dynamics of Implied Volatility Surfaces.” *Quantitative Finance,* 2-1 (2002), 45-60.

Cont, R., and P. Tankov. "Non-Parametric Calibration of Jump-Diffusion Option Pricing Models." *Journal of Computational Finance*, 7 (2004b), 1-49.

Cont, R. and P. Tankov. "Calibration of Jump-Diffusion Option Pricing Models: A Robust Non-Parametric Approach." Rapport Interne 490, CMAP, Ecole Polytechnique (2004).

Cont, R., and P. Tankov. Financial Modelling with Jump Processes. Chapman & Hall/CRC Financial Mathematics Series, (2004).

Cooper, N. "Testing Techniques for Estimating Implied RNDs from the Prices of European and American-Style Options." Working Paper, Bank of England (1999).

Cover, T. M., and J.A. Thomas. *Elements of Information Theory.* John Wiley & Sons (1998).

Derman, E., and I. Kani. "Riding on a Smile." *Risk 7*, (1) (1994), 18-20.

Dumas, B., Fleming, J., and R.E. Whaley. "Implied Volatility Functions: Empirical Tests." *Journal of Finance*, 53, No.6 (1998), 2059-2106.

Dupire, B. "Pricing and Hedging with Smiles." Working Paper (1993), Swaps and Options Research Team, Paribas Capital Markets.

Dupire, B. "Pricing with a smile," *Risk*, 7 (1994), 18-20.

Eberlein, E., and U. Keller. "Hyperbolic Distributions in Finance." *Bernoulli*, 1 (1995), 281-299.

Eberlein, E., Keller, U., and K. Prause. "New Insights into Smile, Mispricing, and Value at Risk: The Hyperbolic Model." *Journal of Business*, Vol. 71 Issue 3 (1998), 371-405.

Engl, H. W., Hanke, M., and A. Neubauer. Regularization of Inverse Problems. Kluwer Academic Publishers Group: Dordrecht (1996).

Feller, W. An Introduction to Probability Theory and Its Applications, John Wiley & Sons (1968).

Gardiner, C. W. Handbook of Stochastic Methods, Springer (1985).

Gemmill, G., and A. Saflekos. "How Useful are Implied Distributions? Evidence from Stock Index Options." *Journal of Derivatives*, Spring (2000)1-17.

Gnedenko, B. V. The Theory of Probability, Chelsea Publishing Company (1962).

Harrison, J. M., and S. R. Pliska. "Martingales and Stochastic Integrals In the Theory of Continuous Trading." *Stochastic Processes and Applications*, 11 (1981), 215-260.

Harrison, J. M., and S. R. Pliska. "A Stochastic Calculous Model of Continuous Trading: Complete Markets." *Stochastic Processes and Applications*, 15 (1983), 313-316.

Heston, S. L. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." *Review of Financial Studies*, 6 (1993), 327-343.

Honoré, P. "Pitfalls in Estimating Jump-Diffusion Models," Working Paper Series No. 18 (1998), CAF Aarhus School of Business, University of Aarhus.

Hull, J. Options, Futures, and Other Derivatives, Prentice Hall (2000).

Hull, J. C., and A. White. "The Pricing of Options on Assets with Stochastic Volatilities." *Journal of Finance*, 42 (1987), 281-300.

Karatzas, I., and S. E. Shreve. Brownian Motion and Stochastic Calculus, Springer-Verlag (1991).

Karlin, S., and H. M. Taylor. A First Course in Stochastic Process, Academic Press (1975).

Koponen, I. "Analytic Approach to the Problem of Convergence of Truncated Lévy Flights Towards the Gaussian Stochastic Process." *Physical Review E*, 52 (1995), 1197-1199.

Madan, D. B., Carr, P., and E. Chang. "The Variance Gamma Process and Option Pricing." *European Finance Review*, 2 (1998), 79-105.

Madan, D. B., and F. Milne. "Option Pricing with VG Martingale Components." *Mathematical Finance*, 1-(4) (1991), 39-55.

Madan, D. B., and E. Seneta. "The Variance Gamma (V.G.) Model for Stock Market Returns." *Journal of Business*, Vol 63, No 4 (1990).

Massey, F. J. Jr. "The Kolmogorov-Smirnov Test of Goodness of Fit." *Journal of the American Statistical Association*, Vol. 46 (1951).

Matsuda, K. "Dynamics of Risk-Neutral Densities Implied By Option Prices." Working Paper (2004a), Graduate School and University Center of the City University of New York.

Matsuda, K. “Introduction to Option Pricing with Fourier Transform: Option Pricing with Exponential Lévy Models.” Working Paper (2004b), Graduate School and University Center of the City University of New York.

Matsuda, K. “Introduction to the Mathematics of Lévy Processes.” Working Paper (2005a), Graduate School and University Center of the City University of New York.

Melick, W. R., and C. P. Thomas. “Recovering an Asset’s Implied PDF from Option Prices: An Application to Crude Oil During the Gulf Crisis.” *Journal of Financial and Quantitative Analysis*, March (1997), 91-115.

Merton, R. C. “Optimum Consumption and Portfolio Rules in a Continuous-Time Model.” *Journal of Economic Theory*, 3 (1971), 373-413.

Merton, R. C. “Theory of Rational Option Pricing.” *Bell Journal of Economics and Management Science*, 4 (1973), 141-183.

Merton, R. C. “Option Pricing When Underlying Stock Returns Are Continuous.” *Journal of Financial Economics*, 3 (1976), 125-144.
Navas, J. F. “A Note on Jump-Diffusion Process for Asset Retuens.” (2000).

Neftci, S. N. An Introduction to the Mathematics of Financial Derivatives, Academic Press (2000).

Oksendal, B. Stochastic Differential Equations: An Introduction with Applications, Springer (2003).

Prause, K. “The Geenralized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures.” PhD Thesis (1999), Universität Freiburg.

Raible, S. “Lévy Processes in Finance: Theory, Numerics and Empirical Facts.” PhD Thesis, Universität Freiburg (1998).

Risken, H. The Fokker-Planck Equation, Springer (1989).

Rogers, L. C. G., and D. Williams. Diffusions, Markov Processes and Martingales, Cambridge University Press (2000).

Ross, S. M. Stochastic Processes, John Wiley & Sons (1983).

Rubinstein, M. “Implied Binomial Trees.” *Journal of Finance*, 49 (1994), 771-818.

Sato, K. Lévy process and Infinitely Divisible Distributions, Cambridge University Press (1999).

Shimko, D. “Bounds of Probability.” *Risk*, 6, 4, April (1993), 33-37.

Shaw, W. T. Modeling Financial Derivatives with Mathematica, Cambridge University Press (1998).

Stephens, M. A. “EDF Statistics for Goodness of Fit and Some Comparisons.” *Journal of the American Statistical Association*, Vol. 69 (1974), 730-737.

Weinberg, S. A. “Interpreting the Volatility Smile: An Examination of the Information Content of Option Prices.” Board of Governors of the Federal Reserve System, International Finance Discussion Papers, Number 706 (2001).

Wilmott, P. Derivatives, John Wiley & Sons (1998).