

# **Handbook of High-Frequency Trading and Modeling in Finance**

---

Wiley Handbooks in

## **FINANCIAL ENGINEERING AND ECONOMETRICS**

Advisory Editor

**Ruey S. Tsay**

*The University of Chicago Booth School of Business, USA*

A complete list of the titles in this series appears at the end of this volume.

# **Handbook of High-Frequency Trading and Modeling in Finance**

*Edited by*

**IONUT FLORESCU  
MARIA C. MARIANI  
H. EUGENE STANLEY  
FREDERI G. VIENS**

**WILEY**

Copyright © 2016 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey  
Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

***Library of Congress Cataloging-in-Publication Data:***

Names: Florescu, Ionut, 1973- editor.

Title: Handbook of high-frequency trading and modeling in finance / edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley, Frederic G. Viens.

Description: Hoboken, NJ : John Wiley & Sons, Inc., [2016] | Includes index.

Identifiers: LCCN 2015043237 (print) | LCCN 2016000501 (ebook) | ISBN 9781118443989 (cloth) | ISBN 9781118593400 (pdf) | ISBN 9781118593325 (epub)

Subjects: LCSH: Investment analysis--Mathematical models. | Investments--Mathematical models. | Finance--Mathematical models.

Classification: LCC HG4529 .H35863 2016 (print) | LCC HG4529 (ebook) | DDC 332.64/20285--dc23

LC record available at <http://lccn.loc.gov/2015043237>

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

# Contents

**Notes on Contributors** *xiii*

**Preface** *xv*

## 1 Trends and Trades 1

*Michael Carlisle, Olympia Hadjiliadis, and Ioannis Stamos*

<b>1.1</b>	Introduction	1
<b>1.2</b>	A trend-based trading strategy	3
<b>1.2.1</b>	Signaling and trends	3
<b>1.2.2</b>	Gain over a subperiod	5
<b>1.3</b>	CUSUM timing	7
<b>1.3.1</b>	Cusum process and stopping time	7
<b>1.3.2</b>	A CUSUM timing scheme	10
<b>1.3.3</b>	US treasury notes, CUSUM timing	11
<b>1.4</b>	Example: Random walk on ticks	12
<b>1.4.1</b>	Random walk expected gain over a subperiod	15
<b>1.4.2</b>	Simple random walk, CUSUM timing	18
<b>1.4.3</b>	Lazy simple random walk, cusum timing	21
<b>1.5</b>	CUSUM strategy Monte Carlo	24
<b>1.6</b>	The effect of the threshold parameter	27
<b>1.7</b>	Conclusions and future work	39
Appendix: Tables		40
References		47

## 2 Gaussian Inequalities and Tranche Sensitivities 51

*Claas Becker and Ambar N. Sengupta*

<b>2.1</b>	Introduction	51
<b>2.2</b>	The tranche loss function	52
<b>2.3</b>	A sensitivity identity	54
<b>2.4</b>	Correlation sensitivities	55
Acknowledgment		58
References		58

**3 A Nonlinear Lead Lag Dependence Analysis of Energy Futures: Oil, Coal, and Natural Gas 61***Germán G. Creamer and Bernardo Creamer***3.1 Introduction 61****3.1.1 Causality analysis 62****3.2 Data 64****3.3 Estimation techniques 64****3.4 Results 65****3.5 Discussion 67****3.6 Conclusions 69****Acknowledgments 69****References 70****4 Portfolio Optimization: Applications in Quantum Computing 73***Michael Marzec***4.1 Introduction 73****4.2 Background 75****4.2.1 Portfolios and optimization 76****4.2.2 Algorithmic complexity 77****4.2.3 Performance 78****4.2.4 Ising model 79****4.2.5 Adiabatic quantum computing 79****4.3 The models 80****4.3.1 Financial model 81****4.3.2 Graph-theoretic combinatorial optimization models 82****4.3.3 Ising and Qubo models 83****4.3.4 Mixed models 84****4.4 Methods 84****4.4.1 Model implementation 85****4.4.2 Input data 85****4.4.3 Mean-variance calculations 85****4.4.4 Implementing the risk measure 86****4.4.5 Implementation mapping 86****4.5 Results 88****4.5.1 The simple correlation model 88****4.5.2 The restricted minimum-risk model 91****4.5.3 The WMIS minimum-risk, max return model 94**

---

<b>4.6</b>	<b>Discussion</b>	<b>95</b>
<b>4.6.1</b>	Hardware limitations	97
<b>4.6.2</b>	Model limitations	97
<b>4.6.3</b>	Implementation limitations	98
<b>4.6.4</b>	Future research	98
<b>4.7</b>	<b>Conclusion</b>	<b>100</b>
Acknowledgments 100		
Appendix 4.A: WMIS Matlab Code 100		
References 103		
<b>5</b>	<b>Estimation Procedure for Regime Switching Stochastic Volatility Model and Its Applications</b>	<b>107</b>
<i>Ionut Florescu and Forrest Levin</i>		
<b>5.1</b>	<b>Introduction</b>	<b>107</b>
<b>5.1.1</b>	The original motivation	108
<b>5.1.2</b>	The model and the problem	108
<b>5.1.3</b>	A brief historical note	109
<b>5.2</b>	<b>The methodology</b>	<b>110</b>
<b>5.2.1</b>	Obtaining filtered empirical distributions at $t_1, \dots, t_T$	110
<b>5.2.2</b>	Obtaining the parameters of the Markov chain	112
<b>5.3</b>	Results obtained applying the model to real data	113
<b>5.3.1</b>	Part i: financial applications	113
<b>5.3.2</b>	Part ii: physical data application. temperature data	119
<b>5.3.3</b>	Part iii: analysis of seismometer readings during an earthquake	121
<b>5.3.4</b>	Analysis of the earthquake signal: beginning	123
<b>5.3.5</b>	Analysis: during the earthquake	125
<b>5.3.6</b>	Analysis: end of the earthquake signal, aftershocks	127
<b>5.4</b>	<b>Conclusion</b>	<b>127</b>
<b>5.A</b>	<b>Theoretical results and empirical testing</b>	<b>128</b>
<b>5.A.1</b>	How does the particle filter work?	128
<b>5.A.2</b>	Theoretical results about convergence and parameter estimates	129
<b>5.A.3</b>	Markov chain parameter estimates	131
<b>5.A.4</b>	Empirical testing	132
<b>5.A.5</b>	A list of supplementary documents	133
References 133		

<b>6 Detecting Jumps in High-Frequency Prices Under Stochastic Volatility: A Review and a Data-Driven Approach</b>	<b>137</b>
<i>Ping-Chen Tsai and Mark B. Shackleton</i>	
<b>6.1 Introduction</b>	<b>137</b>
<b>6.2 Review on the intraday jump tests</b>	<b>140</b>
<b>6.2.1 Realized volatility measure and the BNS tests</b>	<b>140</b>
<b>6.2.2 The ABD and LM tests</b>	<b>142</b>
<b>6.3 A data-driven testing procedure</b>	<b>146</b>
<b>6.3.1 Spy data and microstructure noise</b>	<b>146</b>
<b>6.3.2 A generalized testing procedure</b>	<b>149</b>
<b>6.4 Simulation study</b>	<b>153</b>
<b>6.4.1 Model specification</b>	<b>153</b>
<b>6.4.2 Simulation results</b>	<b>158</b>
<b>6.5 Empirical results</b>	<b>161</b>
<b>6.5.1 Results on the backward-looking test</b>	<b>162</b>
<b>6.5.2 Results on the interpolated test</b>	<b>165</b>
<b>6.6 Conclusion</b>	<b>165</b>
Acknowledgments	<b>166</b>
Appendix 6.A: Least-square estimation of HAR-MA (2) model for log(BP) of SPY	<b>167</b>
Appendix 6.B: Estimation of ARMA (2, 1) model for log(BP) of SPY	<b>168</b>
Appendix 6.C: Minimized loss function loss( $\rho_1, \rho_2$ ) for SV2FJ_2 $\rho$ model, SPY	<b>169</b>
Appendix 6.D.1: Calibration of $\xi$ under SV2FJ_2 $\rho$ model at 2-min frequency, $E[N_t] = 0.08$	<b>170</b>
Appendix 6.D.2: Calibration of $\xi$ under SV2FJ_2 $\rho$ model at 2-min frequency, $E[N_t] = 0.40$	<b>171</b>
Appendix 6.D.3: Calibration of $\xi$ under SV2FJ_2 $\rho$ model at 5-min frequency, $E[N_t] = 0.08$	<b>172</b>
Appendix 6.D.4: Calibration of $\xi$ under SV2FJ_2 $\rho$ Model at 5-min frequency, $E[N_t] = 0.40$	<b>173</b>
Appendix 6.D.5: Calibration of $\xi$ under SV2FJ_2 $\rho$ model at 10-min frequency, $E[N_t] = 0.08$	<b>174</b>
Appendix 6.D.6: Calibration of $\xi$ under SV2FJ_2 $\rho$ model at 10-min frequency, $E[N_t] = 0.40$	<b>175</b>
References	<b>175</b>

---

<b>7 Hawkes Processes and Their Applications to High-Frequency Data Modeling</b>	<b>183</b>
<i>Baron Law and Frederi G. Viens</i>	
<b>7.1</b>	Introduction 183
<b>7.2</b>	Point processes 184
<b>7.3</b>	Hawkes processes 186
<b>7.3.1</b>	Branching structure representation 188
<b>7.3.2</b>	Stationarity 188
<b>7.3.3</b>	Convergence 189
<b>7.4</b>	Statistical inference of Hawkes processes 191
<b>7.4.1</b>	Simulation 191
<b>7.4.2</b>	Estimation 194
<b>7.4.3</b>	Hypothesis testing 197
<b>7.5</b>	Applications of Hawkes processes 198
<b>7.5.1</b>	Modeling order arrivals 199
<b>7.5.2</b>	Modeling price jumps 200
<b>7.5.3</b>	Modeling jump-diffusion 205
<b>7.5.4</b>	Measuring endogeneity (Reflexivity) 205
Appendix 7.A: Point Processes	207
<b>7.A.1</b>	Definition 207
<b>7.A.2</b>	Moments 208
<b>7.A.3</b>	Marked point processes 209
<b>7.A.4</b>	Stochastic intensity 209
<b>7.A.5</b>	Random time change 211
Appendix 7.B: A Brief History of Hawkes processes	211
References	212
<b>8 Multifractal Random Walk Driven by a Hermite Process</b>	<b>221</b>
<i>Alexis Fauth and Ciprian A. Tudor</i>	
<b>8.1</b>	Introduction 221
<b>8.2</b>	Preliminaries 224
<b>8.2.1</b>	Fractional brownian motion and hermite processes 224
<b>8.2.2</b>	Wiener integrals with respect to the hermite process 226
<b>8.2.3</b>	Infinitely divisible cascading noise 229
<b>8.3</b>	Multifractal random walk driven by a Hermite process 231
<b>8.3.1</b>	Definition and existence 231
<b>8.3.2</b>	Properties of the hermite multifractal random walk 233

<b>8.4</b>	Financial applications	234
<b>8.4.1</b>	Simulation of the Hmrw	235
<b>8.4.2</b>	Financial statistics	241
<b>8.5</b>	Concluding remarks	243
	References	247
<b>9</b>	<b>Interpolating Techniques and Nonparametric Regression Methods Applied to Geophysical and Financial Data Analysis</b>	251
	<i>K. Basu and Maria C. Mariani</i>	
<b>9.1</b>	Introduction	251
<b>9.2</b>	Nonparametric regression models	253
<b>9.2.1</b>	Local polynomial regression	255
<b>9.2.2</b>	Lowess/loess method	257
<b>9.2.3</b>	Numerical applications	259
<b>9.3</b>	Interpolation methods	271
<b>9.3.1</b>	Nearest-neighbor interpolation	271
<b>9.3.2</b>	Bilinear interpolation	272
<b>9.3.3</b>	Bicubic interpolation	276
<b>9.3.4</b>	Biharmonic interpolation	277
<b>9.3.5</b>	Thin plate splines	282
<b>9.3.6</b>	Numerical applications	285
<b>9.4</b>	Conclusion	287
	Acknowledgments	292
	References	292
<b>10</b>	<b>Study of Volatility Structures in Geophysics and Finance Using Garch Models</b>	295
	<i>Maria C. Mariani, F. Biney, and I. SenGupta</i>	
<b>10.1</b>	Introduction	295
<b>10.2</b>	Short memory models	297
<b>10.2.1</b>	ARMA(p,q) model	297
<b>10.2.2</b>	GARCH(p,q) model	297
<b>10.2.3</b>	IGARCH(1,1) model	298
<b>10.3</b>	Long memory models	298
<b>10.3.1</b>	ARFIMA(p,d,q) model	299
<b>10.3.2</b>	ARFIMA(p,d,q)-GARCH(r,s)	299
<b>10.3.3</b>	Intermediate memory process	300
<b>10.3.4</b>	Figarch model	300

---

<b>10.4</b>	Detection and estimation of long memory	302
<b>10.4.1</b>	Augmented dickey–fuller test(ADF test)	302
<b>10.4.2</b>	KPSS test	303
<b>10.4.3</b>	Whittle method	304
<b>10.5</b>	Data collection, analysis, and result	306
<b>10.5.1</b>	Analysis on dow Jones index (DJIA) returns	306
<b>10.5.2</b>	Model selection and specification: conditional mean	306
<b>10.5.3</b>	Conditional mean model (returns)	309
<b>10.5.4</b>	Model diagnostics: $ARMA(2, 2)$	309
<b>10.5.5</b>	Test for ARCH effect	311
<b>10.5.6</b>	Model selection and specification: Conditional variance	313
<b>10.5.7</b>	Standardized residuals test	314
<b>10.5.8</b>	Model diagnostics	314
<b>10.5.9</b>	Returns and variance equation	315
<b>10.5.10</b>	standardized residuals test	317
<b>10.5.11</b>	Model diagnostic of conditional returns with conditional variance	318
<b>10.5.12</b>	One-step ahead prediction of last 10 observations	330
<b>10.5.13</b>	Analysis on high-frequency, earthquake, and explosives series	330
<b>10.6</b>	Discussion and conclusion	335
	References	337
<b>11</b>	<b>Scale Invariance and Lévy Models Applied to Earthquakes and Financial High-Frequency Data</b>	341
	<i>M. P. Beccar-Varela, Ionut Florescu, and I. SenGupta</i>	
<b>11.1</b>	Introduction	341
<b>11.2</b>	Governing equations for the deterministic model	342
<b>11.2.1</b>	Application to geophysical (earthquake data)	343
<b>11.2.2</b>	Results	344
<b>11.3</b>	Lévy flights and application to geophysics	345
<b>11.3.1</b>	Truncated Lévy flight distribution	353
<b>11.3.2</b>	Results	356

<b>11.4</b>	Application to the high-frequency market data	360
<b>11.4.1</b>	Methodology	360
<b>11.4.2</b>	Results	361
<b>11.5</b>	Brief program code description	362
<b>11.6</b>	Conclusion	364
<b>11.A</b>	Appendix	366
<b>11.A.1</b>	Stable distributions	366
<b>11.A.2</b>	Characterization of stable distributions	367
	References	368
<b>12</b>	<b>Analysis of Generic Diversity in the Fossil Record, Earthquake Series, and High-Frequency Financial Data</b>	371
	<i>M. P. Beccar Varela, F. Biney, Maria C. Mariani, I. SenGupta, M. Shpak, and P. Bezdek</i>	
<b>12.1</b>	Introduction	371
<b>12.2</b>	Statistical preliminaries and results	373
<b>12.2.1</b>	Sum of exponential random variables with different parameters	374
<b>12.3</b>	Statistical and numerical analysis	377
<b>12.4</b>	Analysis with Lévy distribution	380
<b>12.4.1</b>	Characterization of Stable Distributions	383
<b>12.4.2</b>	Truncated Lévy flight (TLF) distribution	384
<b>12.4.3</b>	Data analysis with TLF distribution	389
<b>12.4.4</b>	Sum of Lévy random variables with different parameters	390
<b>12.5</b>	Analysis of the Stock Indices, high-frequency (tick) data, and explosive series	394
<b>12.6</b>	Results and discussion	409
	Acknowledgments	421
<b>12.A</b>	Appendix A—Big ‘O’ notation	421
	References	422
<b>Index</b>		425

# Notes on Contributors

## Editors

---

**Ionut Florescu**, Financial Engineering Division, Stevens Institute of Technology, Hoboken, NJ, USA

**Maria C. Mariani**, Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX, USA

**H. Eugene Stanley**, Boston University, Boston, MA, USA

**Frederi G. Viens**, Purdue University, West Lafayette, IN, USA

## List of Contributors

---

**K. Basu**, Department of Mathematics, Occidental College, Los Angeles, CA, USA

**M. P. Beccar-Varela**, Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX, USA

**Claas Becker**, Studiengang Angewandte Mathematik, Hochschule Rhein-Main, 65197 Wiesbaden, Germany

**P. Bezdek**, Department of Mathematics, The University of Utah, Salt Lake City, UT, USA

**F. Biney**, Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX, USA

**Michael Carlisle**, Department of Mathematics, Baruch College, City University of New York, New York, NY, USA

**Bernardo Creamer**, Universidad de las Américas, Quito, Ecuador

**Germán G. Creamer**, School of Business, Stevens Institute of Technology, Hoboken, NJ, USA

**Alexis Fauth**, SAMM, Université de Paris 1 Panthéon-Sorbonne, 90, rue de Tolbiac, 75634, Paris, France and Invivoo, 13 rue de l'Abreuvoir, 92400 Courbevoie, France

**Olympia Hadjiliadis**, Department of Mathematics and Statistics, Hunter College, City University of New York, and Departments of Computer Science and Mathematics, Graduate Center, City University of New York, New York, NY, USA

**Baron Law**, Purdue University, West Lafayette, IN, USA

**Forrest Levin**, Financial Engineering Division, Stevens Institute of Technology, Hoboken, NJ, USA

**Michael Marzec**, Stevens Institute of Technology, Hoboken, NJ, USA

**Ambar N. Sengupta**, Department of Mathematics, Louisiana State University, Baton Rouge, LA, USA

**I. SenGupta**, Department of Mathematics, North Dakota State University, Fargo, ND, USA

**Mark B. Shackleton**, Department of Accounting and Finance, Lancaster University Management School, Lancaster, England, United Kingdom

**M. Shpak**, NeuroTexas Institute, St. David's Medical Center, Austin, TX, USA

**Ioannis Stamos**, Department of Computer Science, Hunter College, City University of New York, and Department of Computer Science, Graduate Center, City University of New York, New York, NY, USA

**Ping-Chen Tsai**, Department of Finance, Southern Taiwan University of Science and Technology, Yongkang, Tainan City, Taiwan

**Ciprian A. Tudor**, Laboratoire Paul Painlevé, Université de Lille 1, F-59655 Villeneuve d'Ascq, France, and Department of Mathematics, Academy of Economical Studies, Bucharest, Romania

# Preface

This Handbook is a collection of chapters that describe a range of current empirical and analytical work on financial industry data sampled at high frequency (HF).

Our contemporary Age of Information is a world dominated by ever-increasing quantitative elements that decision makers are expected to take into account. Many fields are confronted with large amounts of data. The phenomenon is particularly challenging in the finance industry, in that evidently relevant data can be sampled with increasingly HF, a trend that started in earnest more than a decade ago and does not seem to be letting down. Some of the special challenges posed by these now staggering amounts of data stem from the uncomfortable evidence that traditional models and information technology tools can be poorly suited to grapple with their size and complexity.

Probabilistic modeling and statistical data analysis attempt to uncover order from apparent disorder. By illustrating this methodological framework in the context of HF finance, the current volume may serve as a guide to various new systematic approaches concerning how to implement these quantitative activities with HF financial data. The chapters herein cover a wide range of topics related to the analysis and modeling of data sampled with HF, principally in finance, as well as in other fields where new ideas may prove helpful to HF finance applications. The first chapters cover the dynamics and complexity of futures and derivatives markets as well as a novel take on the portfolio optimization problem using quantum computers. The following chapters are dedicated to estimating complex model parameters using HF data. The final chapters create links between models used in financial markets and models used in other research areas such as geophysics, fossil records, and earthquake studies.

The editors express their deepest gratitude to all the contributors for their talent and labor in bringing together this Handbook, to the many

anonymous referees who helped the contributors perfect their work, and to Wiley for making the publication a reality.

Ionut Florescu  
*Hoboken, NJ*

Maria C. Mariani  
*El Paso, TX*

H. Eugene Stanley  
*Boston, MA*

Frederi G. Viens  
*Washington, DC*

*August 9, 2015*

## Chapter One

# Trends and Trades

Michael Carlisle<sup>1</sup>, Olympia Hadjiliadis<sup>2</sup>, and Ioannis Stamos<sup>3</sup>

<sup>1</sup>*Department of Mathematics, Baruch College, City University of New York, New York, NY, USA*

<sup>2</sup>*Department of Mathematics and Statistics, Hunter College, City University of New York, and Departments of Computer Science and Mathematics, Graduate Center, City University of New York, New York, NY, USA*

<sup>3</sup>*Department of Computer Science, Hunter College, City University of New York, and Department of Computer Science, Graduate Center, City University of New York, New York, NY, USA*

### 1.1 Introduction

High-frequency data in finance is often characterized by fast fluctuations and noise (see, e.g., [7]), a trait that is known to make the volatility of the data very hard to estimate (see, e.g., [13]). Although this characteristic creates many challenges in modeling, it offers itself to the study of distinguishing “signal” from “noise,” a topic of interest in the area of *quickest detection* (see [25], [5]). One of the most popular algorithms used in quickest detection is known as the *cumulative sum* (CUSUM) stopping rule first introduced by Page [24]. In this work, we employ a sequence of CUSUM stopping rules to construct an online trading strategy. This strategy takes advantage of the relatively frequent number of alarms CUSUM stopping times may provide when applied to high-frequency data as a result of the fast fluctuations present therein. The trading strategy implemented settles frequently and thus eliminates the risk of large positions. This makes the

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

strategy implementable in practice. Prior work has been done by Lam and Yam [20] on drawing connections between CUSUM techniques and the filter trading strategy, yet both the filter trading strategy (see [2, 3]), or its equivalent, the buy and hold strategy (see [12]), run high risks of great losses mainly due to the randomness associated with settling. The well-known trailing stops strategy whose properties have been thoroughly studied in the literature (see, e.g., [15] or [1]) is also related to the filter strategy and thus suffers similar risks.

Although our proposed rule presents clear merits in terms of minimizing the risk of large positions by taking advantage of the high volatility frequently present in high-frequency data, the main purpose of this chapter is to present and illustrate the use of detection techniques (in this case the CUSUM) in high-frequency finance. In particular, the strategy proposed is based on running in parallel two CUSUM stopping rules: one detects an upward (+) change and the other a downward (−) change in the mean of the observations. Once an upward/downward CUSUM alarm (called a “signal”) goes off, there is a buy/short sale of one unit of the underlying asset. At that moment, we repeat a CUSUM stopping rule, and for every alarm of the same sign, we continue buying or short selling one unit of the underlying asset until a CUSUM alarm of the opposite sign is set off, at which time we sell off all of what we bought or buy up all of what we short sold. The high frequency of CUSUM alarms in high-frequency tick data permits the implementation of this rule in practice since large exposures on one side, whether on the buy or on the sell side, are settled relatively quickly.

The algorithmic strategy proposed is applied on real tick data of a 30-year asset and a 5-year note sold at auction on various individual days. It is seen that the algorithm is most profitable in the presence of upward or downward trends (which we call “subperiods”), even in the presence of noise, and is less profitable on periods of price stability. The proposed strategy is, in fact, a trend-following algorithm.

To quantify the performance of the proposed algorithmic strategy, we calculate its expected reward in a simple random walk model. Our diagnostic plots indicate that the more biased the random walk is, the more profitable the proposed strategy becomes, which is consistent with the actual findings when the strategy is applied to real data. This is because in the presence of a bias, trends are more likely to form than in the absence of a bias.

We take the analytical approach of discrete data and a linear random walk model, rather than taking the continuous approach via, for example, the geometric Brownian motion model, because we are analyzing the movement of individual ticks of a price, quantized in a linear fashion (e.g., at the level of 1 cent,  $\frac{1}{32}$  cent, or  $\frac{1}{64}$  cent). Our models focus on tracking the motion of an asset price via these ticks, and so a linear approach is a more realistic setting, when short interest rate effects would be minimal.

We begin our analysis in Section 1.2 by describing a general trading strategy based on following upward or downward trends in a data stream, without specifying the timing mechanism behind such a strategy. We then develop the notion of gain over the time period of an individual trend. In Section 1.3, we build a timing scheme stemming from quickest detection considerations and give a preliminary performance evaluation of the overall strategy on real tick data. Next, in Section 1.4, we analyze the specific case of random walk-based data and calculate the expected value of the gain over a trend in this case. We give an explicit formula for this gain in the special case of simple asymmetric random walk on asset tick changes. Then, in Section 1.5, we give results of Monte Carlo simulations for the asymmetric lazy simple random walk and symmetric lazy random walk on tick changes. In Section 1.6, we discuss the effect of the CUSUM threshold parameter on the trading strategy. We conclude in Section 1.7 by a discussion of ways in which the proposed strategy may be improved with suggestions for further work.

## 1.2 A trend-based trading strategy

Let  $\{S_n\}_{n=0,1,2,\dots}$  be a sequence of data points; for our purposes, they will be samples of the price of an asset. We assume that  $S_0 = s$  is a constant, and  $S_k = 0$  for some  $k$  implies that  $S_n = 0$  for all  $n > k$ . Let  $T_0 = 0$ , and define  $T_k$ ,  $k = 1, 2, \dots$  as an increasing sequence of (stopping) times, called *signals*, noting some trend in the sequence. We call  $T_k$  the  $k$ -th *signal*.

### 1.2.1 SIGNALING AND TRENDS

In this subsection, we construct a trading strategy in the case that there are two types of signals: “+ signals” (declaring the detection of an upward

trend in the data) and “– signals” (declaring the detection of a downward trend in the data). Let “Property  $+(k)$ ” be the property that causes a + signal to occur as the  $k$ th signal, and denote this event by  $\{T_k = T_k^+\}$ . Likewise, let “Property  $-(k)$ ” be the property that causes a – signal to occur as the  $k$ -th signal, and denote this by  $\{T_k = T_k^-\}$ . Only one type of trend can be detected at a time, so we formally define  $T_k^+$  and  $T_k^-$  by

$$T_k^+ := \begin{cases} T_k & \text{if Property } + (k) \text{ occurs} \\ \infty & \text{if Property } - (k) \text{ occurs} \end{cases} \quad (1.1)$$

$$T_k^- := \begin{cases} T_k & \text{if Property } - (k) \text{ occurs} \\ \infty & \text{if Property } + (k) \text{ occurs} \end{cases} \quad (1.2)$$

Thus,  $T_k = T_k^+ \wedge T_k^-$  for every  $k = 1, 2, \dots$ .

Next, we state what it means for the data to stay in a trend. We define the sequence of signal indices  $\alpha(l)$  as follows: let  $\alpha(0) = 0$ , so  $T_{\alpha(0)} = 0$ , and for  $l \geq 1$ , with  $k \geq 2$ , define the properties

- “Property  $+(l, k)$ ” :  $T_j = T_j^-$  for every  $\alpha(l-1) < j < k$  and  $T_k = T_k^+$
- “Property  $-(l, k)$ ” :  $T_j = T_j^+$  for every  $\alpha(l-1) < j < k$  and  $T_k = T_k^-$ .

Then, we define the  $l$ th shift point as, for  $l = 1, 2, \dots$ ,

$$\alpha(l) := \inf \{k \geq \alpha(l-1) + 2 : \text{Property } + (l, k) \text{ or Property } - (l, k) \text{ holds}\}. \quad (1.3)$$

Note that  $T_{\alpha(l)}$  is at least two signals after  $T_{\alpha(l-1)}$ . Definition (1.3) is equivalent to

$$\alpha(l) := \inf \{k \geq \alpha(l-1) + 2 : T_k \text{ has different sign than } T_j, \alpha(l-1) < j < k\}. \quad (1.4)$$

A sequence of the same type of signal will be called a *subperiod* of the sample points. A shift point denotes the *end of a subperiod* of the same type of signal.

Let  $\Delta_n$  be the number of shares of the asset  $S$  held at time  $n$ . Set  $\Delta_0 = 0$ . Note that, for every  $n \in (T_{\alpha(l)}, T_{\alpha(l+1)})$ , the sign of  $\Delta_n$  is invariant, that is,

either  $\Delta_n > 0$  holds for every  $n \in (T_{\alpha(l)}, T_{\alpha(l+1)})$  or  $\Delta_n < 0$  holds for every  $n \in (T_{\alpha(l)}, T_{\alpha(l+1)})$ .

Our trading strategy is as follows:

$$\Delta_{n+1} = \begin{cases} \Delta_n & \text{if no signal at time } n, \text{ i.e. } n \neq T_j \forall j \text{ (no change)} \\ \Delta_n + 1 & \text{if } n = T_j = T_j^+ \text{ for some } j, \alpha(l) < j < \alpha(l+1) \\ & \quad \text{for some } l \text{ (buy one during a + subperiod)} \\ \Delta_n - 1 & \text{if } n = T_j = T_j^- \text{ for some } j, \alpha(l) < j < \alpha(l+1) \\ & \quad \text{for some } l \text{ (sell one during a - subperiod)} \\ 0 & \text{if } n = T_{\alpha(l)} \text{ for some } l \geq 1 \\ & \quad \text{(buy-up if } T_{\alpha(l)}^+; \text{ sell-off if } T_{\alpha(l)}^-). \end{cases} \quad (1.5)$$

We assume a market in which all market orders are instantly fulfilled. The intent of this strategy is to profit from following subperiods of + or - signals by the old adage “buy low, sell high.” The success of this strategy relies mainly on the length of such subperiods.

### 1.2.2 GAIN OVER A SUBPERIOD

We wish to analyze the gain  $G_l$ ,  $l = 1, 2, \dots$ , for this trading strategy over the time period  $(T_{\alpha(l-1)}, T_{\alpha(l)}]$ , called *subperiod*  $l$ ; this is the amount of cash earned or lost by liquidating the transactions made from signals  $T_{\alpha(l-1)+1}, \dots, T_{\alpha(l)-1}$  at  $T_{\alpha(l)}$ .

Note that a subperiod is determined by the first signal on that run: if  $T_1 = T_1^+$ , then the run from signal 1 to signal  $\alpha(1) - 1$  is a “bull run” subperiod of individual buy orders followed by a sell-off at time  $T_{\alpha(1)} = T_{\alpha(1)}^-$ ; if  $T_1 = T_1^-$ , then this run is a “bear run” subperiod of individual short sales followed by a buy-up at  $T_{\alpha(1)} = T_{\alpha(1)}^+$ . Define  $G_l$  to be the gain on subperiod  $l$ ; thus,  $G_1$  is the gain on the first subperiod, starting at signal  $T_{\alpha(0)+1} = T_1$  and ending at signal  $T_{\alpha(1)}$ . We require, as a condition, the sign of the first signal of the subperiod. Let  $c \geq 0$  be the percentage cost per transaction, and define

$$A_l := 1_{\{T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^-\}}, \quad Y_l := \alpha(l) - \alpha(l-1) - 1. \quad (1.6)$$

The gain on a subperiod is calculated as follows:

$$G_l := \begin{cases} (1 - c) \sum_{j=\alpha(l-1)+1}^{\alpha(l)-1} S_{T_j} - (1 + c)(\alpha(l) - \alpha(l-1) - 1)S_{T_{\alpha(l)}} \\ \quad \text{if } T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^-, \\ (1 - c)(\alpha(l) - \alpha(l-1) - 1)S_{T_{\alpha(l)}} - (1 + c) \sum_{j=\alpha(l-1)+1}^{\alpha(l)-1} S_{T_j} \\ \quad \text{if } T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^+, \\ (1 - c) \sum_{j=1}^{Y_l} S_{T_{j+\alpha(l-1)}} - (1 + c)(Y_l)S_{T_{\alpha(l)}} \text{ if } T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^-, \\ (1 - c)(Y_l)S_{T_{\alpha(l)}} - (1 + c) \sum_{j=1}^{Y_l} S_{T_{j+\alpha(l-1)}} \text{ if } T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^+. \end{cases} \quad (1.7)$$

For example, if  $c = 0.01$ ,  $T_1 = T_1^+$ , and  $\alpha(1) = 4$ , then  $T_{\alpha(1)} = T_4 = T_4^-$ . Say the prices at the buy-signal times are  $S_{T_1} = 5$ ,  $S_{T_2} = 7$ ,  $S_{T_3} = 9$ , and we sell everything off at  $S_{T_4} = 8$ . Then  $\Delta_{T_0} = 0$ ,  $\Delta_{T_1} = 1$ ,  $\Delta_{T_2} = 2$ ,  $\Delta_{T_3} = 3$ , and we liquidate at time  $T_4$  to  $\Delta_{T_4} = 0$ . The gain on the first subperiod would then be  $G_1 = (0.99)(3)(8) - (1.01)(5 + 7 + 9) = 2.55$ .

Combining the  $1 - c$  terms and adding on the random variable  $2cY_l S_{\alpha(T_l)}$ , we have after some algebra a sum of price increments:

$$\begin{aligned} G_l + 2cY_l S_{\alpha(T_l)} &= (c + (-1)^{A_l}) \left[ Y_l S_{T_{\alpha(l)}} - \sum_{j=1}^{Y_l} S_{T_{j+\alpha(l-1)}} \right] \\ &= (c + (-1)^{A_l}) \sum_{j=1}^{Y_l} (S_{T_{\alpha(l)}} - S_{T_{j+\alpha(l-1)}}). \end{aligned} \quad (1.8)$$

We can rewrite each difference in the sum as a telescoping sum: setting

$$Z_k := S_{T_{k+1}} - S_{T_k}, \quad k = 1, 2, \dots, \quad (1.9)$$

as the incremental price change between signals  $k$  and  $k + 1$ , we have

$$S_{T_{\alpha(l)}} - S_{T_{j+\alpha(l-1)}} = \sum_{k=j+\alpha(l-1)}^{\alpha(l)-1} (S_{T_{k+1}} - S_{T_k}) = \sum_{k=j+\alpha(l-1)}^{\alpha(l)-1} Z_k = \sum_{k=j}^{Y_l} Z_{k+\alpha(l-1)}.$$

Substituting this back into (1.8) yields

$$G_l + 2cY_l S_{\alpha(T_l)} = (c + (-1)^{A_l}) \sum_{j=1}^{Y_l} \left[ \sum_{k=j+\alpha(l-1)}^{\alpha(l)-1} Z_k \right] = (c + (-1)^{A_l}) \sum_{j=1}^{Y_l} j Z_{j+\alpha(l-1)}. \quad (1.10)$$

Therefore, by (1.11), the gain over subperiod  $l$  is

$$G_l = (c + (-1)^{A_l}) \sum_{j=1}^{Y_l} j Z_{j+\alpha(l-1)} - 2c Y_l S_{\alpha(T_l)}. \quad (1.11)$$

Note that, in the absence of transaction costs (i.e.,  $c = 0$ ), the expected gain  $G_l$  is entirely determined by price increments and the sign of the first signal of the subperiod.

## 1.3 CUSUM timing

Next, we describe a version of the CUSUM statistic process and its associated CUSUM stopping rule, which we will use to devise a timing scheme based on the quickest detection of trends, and incorporate this scheme to our trading strategy.

### 1.3.1 CUSUM PROCESS AND STOPPING TIME

In this section, we begin by introducing the measurable space  $(\Omega, \mathcal{F})$ , where  $\Omega = \mathbb{R}^\infty$ ,  $\mathcal{F} = \cup_n \mathcal{F}_n$ , and  $\mathcal{F}_n = \sigma\{Y_i, i \in \{0, 1, \dots, n\}\}$ . The law of the sequence  $Y_i$ ,  $i = 1, \dots$ , is described by the family of probability measures  $\{P_v\}$ ,  $v \in \mathbb{N}^*$ . In other words, the probability measure  $P_v$  for a given  $v > 0$ , playing the role of the *change point*, is the measure generated on  $\Omega$  by the sequence  $Y_i$ ,  $i = 1, \dots$ , when the distribution of the  $Y_i$ 's changes at time  $v$ . The probability measures  $P_0$  and  $P_\infty$  are the measures generated on  $\Omega$  by the random variables  $Y_i$  when they have an identical distribution. In other words, the system defined by the sequence  $Y_i$  undergoes a “regime change” from the distribution  $P_0$  to the distribution  $P_\infty$  at the change point time  $v$ .

The *CUSUM statistic* is defined as the maximum of the log-likelihood ratio of the measure  $P_v$  to the measure  $P_\infty$  on the  $\sigma$ -algebra  $\mathcal{F}_n$ . That is,

$$C_n := \max_{0 \leq v \leq n} \log \frac{dP_v}{dP_\infty} \Big|_{\mathcal{F}_n} \quad (1.12)$$

is the CUSUM statistic on the  $\sigma$ -algebra  $\mathcal{F}_n$ . The *CUSUM statistic process* is then the collection of the CUSUM statistics  $\{C_n\}$  of (1.12) for  $n = 1, \dots$

The *CUSUM stopping rule* is then

$$T(h) := \inf \left\{ n \geq 0 : \max_{0 \leq v \leq n} \log \frac{dP_v}{dP_\infty} \Big|_{\mathcal{F}_n} \geq h \right\}, \quad (1.13)$$

for some threshold  $h > 0$ . In the CUSUM stopping rule (1.13), the CUSUM statistic process of (1.12) is initialized at

$$C_0 = 0. \quad (1.14)$$

The CUSUM statistic process was first introduced by Page [24] in the form that it takes when the sequence of random variables  $Y_i$  is independent and Gaussian; that is,  $Y_i \sim N(\mu, 1)$ ,  $i = 1, 2, \dots$ , with  $\mu = \mu_0$  for  $i < v$  and  $\mu = \mu_1$  for  $i \geq v$ . Since its introduction by Page [24], the CUSUM statistic process of (1.12) and its associated CUSUM stopping time of (1.13) have been used in a plethora of applications where it is of interest to perform detection of abrupt changes in the statistical behavior of observations in real time. Examples of such applications are signal processing (see [10]), monitoring the outbreak of an epidemic (see [29]), financial surveillance (see [14] and [9]), and more recently computer vision (see [19] or [30]). The popularity of the CUSUM stopping time (1.13) is mainly due to its low complexity and optimality properties (see, for instance, [21], [22, 23], [6] and [27] or [26]), in both discrete and continuous time models.

As a specific example, we now derive the form in which Page [24] introduced the CUSUM. To this effect, let  $Y_i \sim N(\mu_0, \sigma^2)$  that change to  $Y_i \sim N(\mu_1, \sigma^2)$  at the change point time  $v$ . We now proceed to derive the form of the CUSUM statistic process (1.12) and its associated CUSUM stopping time (1.13) in the example set forth in this section. To this effect, let us now denote by  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  the Gaussian kernel. For the sequence of random variables  $Y_i$  given earlier, we can now compute (see also [28] or [25]):

$$\begin{aligned} C_n &= \max_{0 \leq v \leq n} \log \frac{dP_v}{dP_\infty} \Big|_{\mathcal{F}_n} = \max_{0 \leq v \leq n} \log \frac{\prod_{i=1}^{v-1} \phi\left(\frac{Y_i - \mu_0}{\sigma}\right) \prod_{i=v}^n \phi\left(\frac{Y_i - \mu_1}{\sigma}\right)}{\prod_{i=1}^n \phi\left(\frac{Y_i - \mu_0}{\sigma}\right)} \\ &= \frac{1}{\sigma^2} \max_{0 \leq v \leq n} (\mu_1 - \mu_0) \sum_{i=v}^n \left[ Y_i - \frac{\mu_1 + \mu_0}{2} \right]. \end{aligned} \quad (1.15)$$

In view of (1.14), we initialize the sequence (1.15) at  $Y_0 = \frac{\mu_1 + \mu_0}{2}$  and proceed to distinguish the following two cases:

- Case 1:  $\mu_1 > \mu_0$ : divide out  $\mu_1 - \mu_0$ , multiply by the constant  $\sigma^2$  in (1.15), and use (1.13) to obtain the CUSUM stopping rule  $T^+$ :

$$T^+(h^+) = \inf \left\{ n \geq 0 : \max_{0 \leq v \leq n} \sum_{i=v}^n \left[ Y_i - \frac{\mu_1 + \mu_0}{2} \right] \geq h^+ \right\} \quad (1.16)$$

for an appropriately scaled threshold  $h^+ > 0$ .

- Case 2:  $\mu_1 < \mu_0$ : divide out  $\mu_1 - \mu_0$ , multiply by the constant  $\sigma^2$  in (1.15), and use (1.13) to obtain the CUSUM stopping rule  $T^-$ :

$$T^-(h^-) = \inf \left\{ n \geq 0 : \max_{0 \leq v \leq n} \sum_{i=v}^n \left[ \frac{\mu_1 + \mu_0}{2} - Y_i \right] \geq h^- \right\} \quad (1.17)$$

for an appropriately scaled threshold  $h^- > 0$ .

As shown in the study [24] or [11], we can reexpress the stopping times (1.16) and (1.17) in terms of the recurrence relations

$$u_0 = 0; \quad u_n := \max \left\{ 0, u_{n-1} + \left( Y_n - \frac{\mu_1 + \mu_0}{2} \right) \right\} \quad (1.18)$$

$$d_0 = 0; \quad d_n := \max \left\{ 0, d_{n-1} - \left( Y_n - \frac{\mu_1 + \mu_0}{2} \right) \right\}, \quad (1.19)$$

which lead to

$$T^+(h^+) = \inf \{n > 0 : u_n \geq h^+\}, \quad (1.20)$$

$$T^-(h^-) = \inf \{n > 0 : d_n \geq h^-\}. \quad (1.21)$$

The sequences  $u_n$  and  $d_n$  of (1.18) and (1.19), respectively, form a CUSUM according to the deviation of the monitored sequential observations  $Y_n$  from the average of their pre- and postchange means. The first time that one of these sequences reaches its threshold (in (1.20) or (1.21)), the respective alarm  $T^+$  or  $T^-$  fires.

Although the stopping times (1.16) and (1.17) and their respective equivalents (1.20) and (1.21) can be derived by formal CUSUM regime change considerations using the example set forth in this section, they may also be used as general nonparametric stopping rules directly applied to sequential observations as seen in the study by Brodsky and Darkhovsky

[8] or Devore [11]. The former can be used as a general stopping rule to detect an upward change in the mean while the latter a downward one. In many applications, it is of interest to monitor an upward or downward change in the mean of sequential observations simultaneously. This gives rise to the two-sided CUSUM (2-CUSUM), which was first introduced by Barnard [4], and whose optimality properties have been established in Hadjiliadis [17], Hadjiliadis and Moustakides [16], and Hadjiliadis et al. [18]. In the context presented in this section, the 2-CUSUM stopping time takes the form

$$T^+(h^+) \wedge T^-(h^-), \quad (1.22)$$

where  $T^+(h^+)$  appears in (1.20) and  $T^-(h^-)$  in (1.21). The symmetric version of the 2-CUSUM stopping time is that of (1.22) when  $h^+ = h^- = h$ .

### 1.3.2 A CUSUM TIMING SCHEME

We now apply the aforementioned CUSUM stopping rule of (1.22) to a stream of data representing the value of the underlying asset without any model assumptions. In other words, the underlying asset is not necessarily assumed to be independent or normally distributed. That is, we apply the forms (1.16) and (1.17) in a nonparametric fashion. Let  $M > 0$  denote the “tick size” of the asset being monitored (presuming that  $S$  changes in increments of  $M$ ; we do not know the probability distribution of these changes), and  $h > 0$  be a given threshold. Given that  $S_0 = s$ , recall that  $T_0 = 0$ . We monitor the progress of upward or downward adjustments in the price  $S_n$  of the underlying, by individual ticks.

In view of the previous subsection at time  $T_k$ ,  $\mu_0$  is set to the value of the underlying at time  $T_k$ , namely  $\mu_0 = S_{T_k}$ , and  $\mu_1^u = S_{T_k} + M$  and  $\mu_1^d = S_{T_k} - M$  are the two “new” mean levels to be monitored against. Thus, as in equations (1.18) and (1.19), which cumulate the deviations of the monitored sequence from the average of their pre- and postchange means, we now monitor the deviations of the underlying sequence  $S_n$ ,  $n = 1, 2, \dots$ , from the quantities

$$\begin{aligned} m_k^u &:= \frac{(S_{T_k} + M) + S_{T_k}}{2} = S_{T_k} + \frac{M}{2}, \\ m_k^d &:= \frac{(S_{T_k} - M) + S_{T_k}}{2} = S_{T_k} - \frac{M}{2}, \end{aligned} \quad (1.23)$$

where  $k \geq 0$ . To this effect, set  $u_0^k = d_0^k = 0$ , and for  $n \geq 1$ , define the CUSUM statistics

$$\begin{aligned} u_n^k &:= \max\{0, u_{n-1}^k + (S_{n+T_k} - m_k^u)\}, \\ d_n^k &:= \max\{0, d_{n-1}^k - (S_{n+T_k} - m_k^d)\}. \end{aligned} \quad (1.24)$$

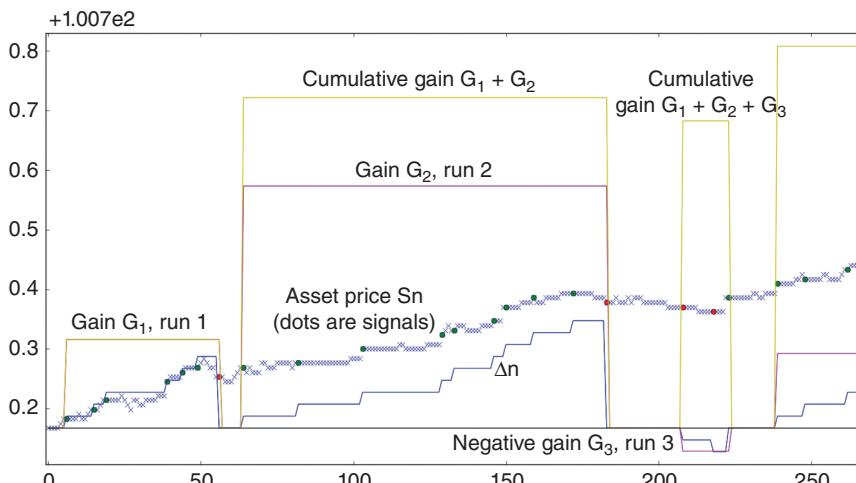
Thus, for  $k \geq 0$ , the CUSUM timing scheme for our trend-following trading strategy is defined by using (1.20) and (1.21) (and coming from (1.1) and (1.2)),

$$\begin{aligned} \text{Property } + (k+1) : u_n^k \geq h; \quad \text{Property } - (k+1) : d_n^k \geq h \\ j_k^* := \min\{n > 0 : \text{Property } + (k+1) \text{ or } - (k+1) \text{ occurs}\} \quad (1.25) \\ T_{k+1} := T_k + j_k^*. \end{aligned}$$

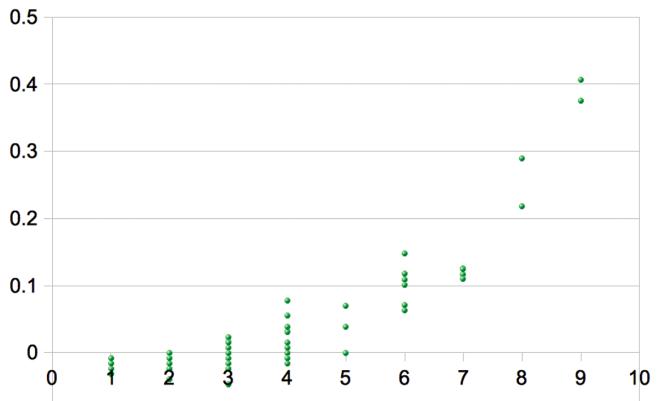
In other words, each  $T_k$  is the symmetric 2-CUSUM stopping time of (1.22) for cycle  $k$ . Finally, at the “end of day,” that is, on the final tick, we close out our position, inducing a final shift point to end trading, for algorithmic purposes.

### 1.3.3 US TREASURY NOTES, CUSUM TIMING

The following figures and chart describe the CUSUM timing scheme (1.25) applied to the trading strategy (1.5) for US Treasury notes sold at auction in 2011. Gains quoted are in increments of \$1000. In Figure 1.1, we show the



**FIGURE 1.1** Plot of the first subperiods, and cumulative gain, for the CUSUM strategy, August 2, 2011, US 5-year treasury note.

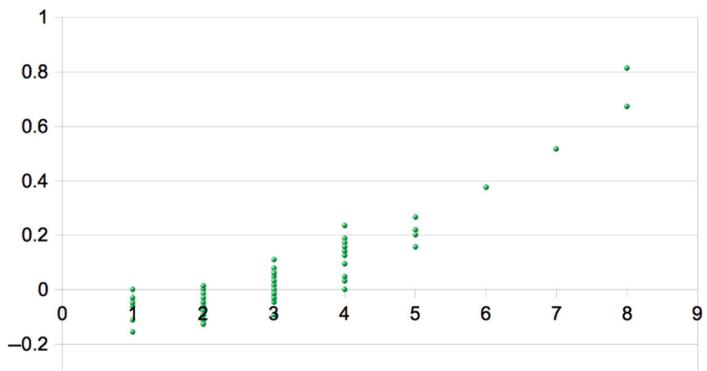


**FIGURE 1.2** Lengths of subperiods versus gains, August 2, 2011, US 5-year treasury note.

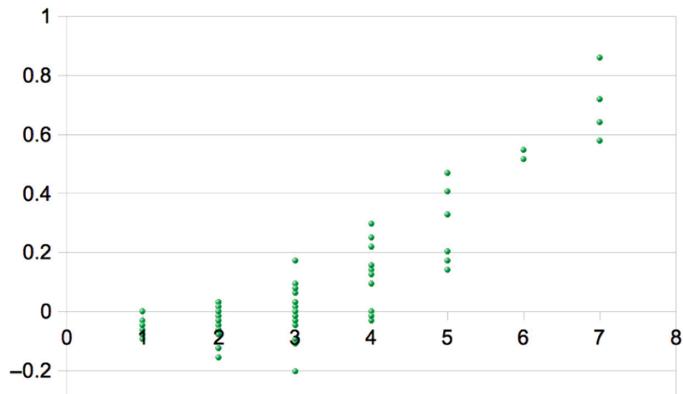
asset price, along with the number of shares held, per-subperiod gain, and running total gain. Figures 1.2, 1.3, 1.4, 1.5 and 1.6 show the individual subperiod gains, plotted by the number of signals during a subperiod, of the gain for 5-year and 30-year treasury notes, and Figure 1.7 aggregates the data from Figures 1.3, 1.4, 1.5 and 1.6 for 30-year notes.

## 1.4 Example: Random walk on ticks

We now describe a simple example to model the asset price motions. Assume that  $\exists N > 0$  such that the sequence  $\{X_j\}_{j \in \mathbb{N}}$  are the steps of a



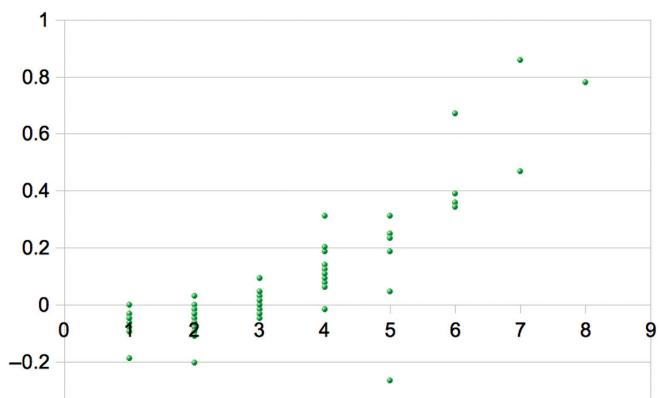
**FIGURE 1.3** Subperiod length versus gain, July 29, 2011, US 30-year treasury note.



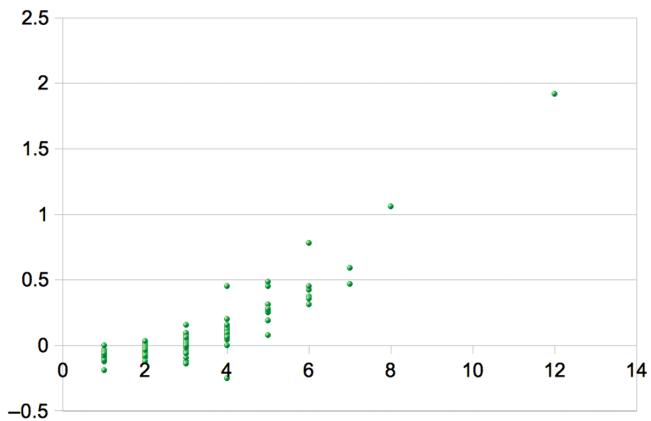
**FIGURE 1.4** Subperiod length versus gain, August 1, 2011, US 30-year treasury note.

random walk taking integer values bounded between  $-N$  and  $N$ , that is,  $|X_j| \leq N$  for all  $j \in \mathbb{N}$ , and that  $X_j \in \{-N, -N + 1, \dots, N - 1, N\}$  for every  $j$ , with  $p_k = P(X_j = k) \geq 0$  and  $\sum_{k=-N}^N p_k = 1$ . Let  $S_0 = s$ , and for  $n \geq 1$ , set  $S_n = s + \sum_{j=1}^n X_j$ . We will consider  $S_n$  to be a random walk on ticks, rather than price itself, and so normalize tick size to  $M = 1$ .

Note that, since  $\Delta_n = 0 \iff n = \alpha(l)$  for some  $l \in \{0, 1, 2, \dots\}$ , the expected gain over a subperiod is the expected gain over an excursion to zero on  $\Delta_n$ , and so we can simply consider the first excursion (independent of other excursions) on the time interval  $(T_{\alpha(0)} = 0, T_{\alpha(1)}]$ . Also, note that



**FIGURE 1.5** Subperiod length versus gain, August 2, 2011, US 30-year treasury note.



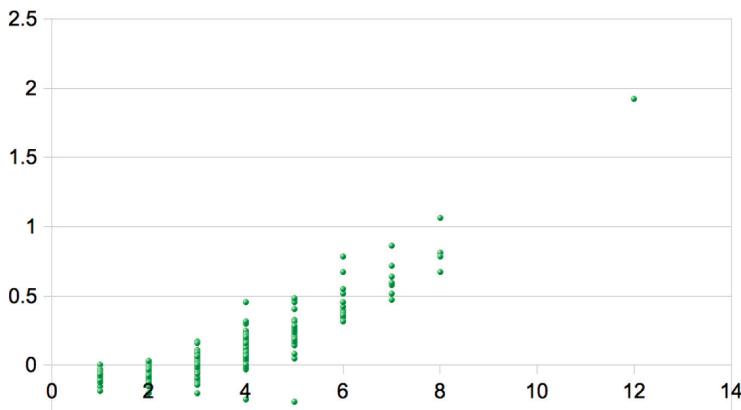
**FIGURE 1.6** Subperiod length versus gain, August 3, 2011, US 30-year treasury note.

in this case, if the transaction cost  $c = 0$ , the  $G_l$  of (1.11) are IID random variables.

Set

$$p^+ := P(T_1 = T_1^+), \quad p^- := 1 - p^+ = P(T_1 = T_1^-), \quad (1.26)$$

and note that signal timing increments are independent. Conditioned on the sign of signal  $\alpha(l-1) + 1$  at time  $T_{\alpha(l-1)+1}$ ,  $Y_l$  is a geometric random



**FIGURE 1.7** Figures 1.3, 1.4, 1.5 and 1.6 combined (30-year).

variable (starting at 1) which gives the number of signals of the same sign in subperiod  $l$ . The distribution of  $Y_l$ , conditioned on  $\mathcal{F}_{T_{\alpha(l-1)+1}}$ , is

$$Y_l \sim \begin{cases} \text{geom}(p^-) & \text{if } T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^+, \\ \text{geom}(p^+) & \text{if } T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^-. \end{cases} \quad (1.27)$$

To explain this, consider the case  $T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^+$  (the first + signal of a bull run subperiod): a subperiod of + has “failure” probability  $p^+$  (a + signal continues the subperiod with another buy) and “success” probability  $p^-$  (a – signal causes a sell-off and ends the subperiod).

$A_l$  is an  $\mathcal{F}_{T_{\alpha(l-1)+1}}$ -measurable random variable, and every increment in the sum in (1.11) is independent of time  $T_{\alpha(l-1)+1}$ . Finally, note that the  $Y_l$  are independent of the walk up to time  $T_{\alpha(l-1)}$ , and if  $c = 0$ , so are the  $G_l$ .

### 1.4.1 RANDOM WALK EXPECTED GAIN OVER A SUBPERIOD

We wish to examine the expected gain  $E(G_l)$  over subperiod  $l$ . For simplicity in our initial analysis, set  $c = 0$ . Since the  $G_l$  are IID, we will calculate  $E(G_1)$ . This is, since  $\alpha(0) = 0$  and  $Y_1 = \alpha(1) - \alpha(0) - 1 = \alpha(1) - 1$ , by (1.11),

$$E(G_1) = E \left[ (-1)^{A_1} \sum_{j=1}^{Y_1} j Z_j \right]. \quad (1.28)$$

We condition over the possible values of  $Y_1$  and  $A_1$ . Note that the sign of  $T_1$  also determines the possibilities of  $Z_j$  for  $j = 1, 2, \dots, Y_1 - 1$ .  $Z_j$  depends on the type of subperiod it resides on, so by the fact that the event  $\{Y_l = n\} \in \mathcal{F}_{T_{\alpha(l)}}$ , and by setting, for  $j = \alpha(l-1) + 1, \dots, \alpha(l)$ ,

$$\begin{aligned} B_{j,l,n}^+ &:= E(Z_j \mid T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^+, Y_l = n), \\ B_{j,l,n}^- &:= E(Z_j \mid T_{\alpha(l-1)+1} = T_{\alpha(l-1)+1}^-, Y_l = n), \end{aligned} \quad (1.29)$$

then, for  $n = 1, 2, \dots$ , we have

$$\begin{aligned} E \left[ \sum_{j=1}^{Y_1} j Z_j \mid T_1 = T_1^+, Y_1 = n \right] &= E \left[ \sum_{j=1}^n j Z_j \mid T_1 = T_1^+, Y_1 = n \right] \\ &= \sum_{j=1}^n j E \left[ Z_j \mid T_1 = T_1^+, Y_1 = n \right] = \sum_{j=1}^n j B_{j,1,n}^+. \end{aligned} \quad (1.30)$$

Since the conditioning on  $B_{j,1,n}^+$  (and, likewise,  $B_{j,1,n}^-$ ) is based only on the walk during the time increments  $(T_0, T_1]$  and  $(T_{\alpha(1)-1}, T_{\alpha(1)}]$ , for  $n > 1$ ,  $B_{j,1,n}^+$  and  $B_{j,1,n}^-$  are numbers for  $j = 1, 2, \dots, n - 1$ . Also, for these  $j$ ,  $B_{j,1,n}^+$  are the same by the strong Markov property at  $T_{j-1}$  since the signs on the  $T_j$  are all +. However, since the signal  $T_{\alpha(1)} = T_{n+1}$  has different sign than  $T_n$ ,  $B_{n,1,n}^+$  has a different distribution. In fact, since this condition implies that  $T_{n+1} = T_{\alpha(1)} = T_{\alpha(1)}^-$ ,  $B_{n,1,n}^+$  can be written by the strong Markov property at  $T_n = T_{\alpha(1)-1}$  as

$$\begin{aligned} B_{n,1,n}^+ &= E \left[ Z_n \mid T_1 = T_1^+, Y_1 = n \right] = E \left[ Z_n \mid T_1 = T_1^+, T_{\alpha(1)=n+1} = T_{\alpha(1)}^- \right] \\ &= E \left[ Z_n \mid T_{\alpha(1)=n+1} = T_{\alpha(1)}^- \right] = B_{1,1,n}^- \end{aligned}$$

To simplify notation, we rewrite  $B_{1,1,n}^+ = B^+$  and  $B_{1,1,n}^- = B^-$ , since they do not depend on  $n$ . In the case  $n = 1$ , we simply have  $B_{1,1,1}^+ = B^-$  and  $B_{1,1,1}^- = B^+$ , and note that  $B^+ \geq 0$  and  $B^- \leq 0$ . Thus, our sum (1.30) becomes

$$\begin{aligned} E \left[ \sum_{j=1}^{Y_1} jZ_j \mid T_1 = T_1^+, Y_1 = n \right] &= \sum_{j=1}^n jB_{j,1,n}^+ = \sum_{j=1}^{n-1} jB_{j,1,n}^+ + nB_{n,1,n}^+ \\ &= \frac{n(n-1)}{2} B^+ + nB^- . \end{aligned} \quad (1.31)$$

The only thing that needs to change for the analogous argument for  $B_{j,1,n}^-$  are the signs; thus, we also have

$$E \left[ \sum_{j=1}^{Y_1} jZ_j \mid T_1 = T_1^-, Y_1 = n \right] = \frac{n(n-1)}{2} B^- + nB^+. \quad (1.32)$$

Next, we give the probability that  $Y_1 = n$ , conditioned on the sign of  $T_1$ . This is easy, since we know that, conditioned on the sign of  $T_1$ ,  $Y_1$  is a geometric random variable. By (1.27), for  $n = 1, 2, \dots$ ,

$$\begin{aligned} P(Y_1 = n \mid T_1 = T_1^+) &= (p^+)^{n-1}(p^-) \\ P(Y_1 = n \mid T_1 = T_1^-) &= (p^-)^{n-1}(p^+). \end{aligned} \quad (1.33)$$

By (1.31), (1.32), and (1.33), and recalling that  $p^- = 1 - p^+$ , the expected gain on a subperiod, given that the subperiod consists of  $n$  signals before a liquidation, is

$$\begin{aligned} E(G_1 \mid Y_1 = n) &= p^+ E \left[ \sum_{j=1}^{Y_1} jZ_j \mid T_1 = T_1^+, Y_1 = n \right] \\ &\quad - p^- E \left[ \sum_{j=1}^{Y_1} jZ_j \mid T_1 = T_1^-, Y_1 = n \right] \\ &= p^+ \left( \frac{n(n-1)}{2} B^+ + nB^- \right) - p^- \left( \frac{n(n-1)}{2} B^- + nB^+ \right) \\ &= \frac{n(n+1)}{2} (B^+ p^+ - B^- p^-) + n(B^- - B^+). \end{aligned} \tag{1.34}$$

The probability that a subperiod lasts  $n$  signals, regardless of its sign, is, by (1.27) and (1.33),

$$\begin{aligned} P(Y_1 = n) &= P(Y_1 = n \mid T_1 = T_1^+) P(T_1 = T_1^+) + P(Y_1 = n \mid T_1 = T_1^-) P(T_1 = T_1^-) \\ &= (p^+)^n (p^-) + (p^-)^n (p^+), \end{aligned} \tag{1.35}$$

which also gives the expected number of same-sign signals in a subperiod

$$E(Y_1) = \sum_{n=1}^{\infty} nP(Y_1 = n) = \frac{p^+}{p^-} + \frac{p^-}{p^+}. \tag{1.36}$$

Note that this necessarily matches the calculation via conditioning on  $T_1$ 's sign; that is, by (1.33),

$$E(Y_1) = E(Y_1 \mid T_1 = T_1^+) p^+ + E(Y_1 \mid T_1 = T_1^-) p^- = \frac{p^+}{p^-} + \frac{p^-}{p^+}.$$

We can sum over all possible values  $n$  in (1.34), and use (1.35) to get the expected gain of a subperiod in terms of  $p^+$ ,  $p^-$ ,  $B^+$ , and  $B^-$ :

$$\begin{aligned} E(G_1) &= \sum_{n=1}^{\infty} E(G_1 | Y_1 = n) P(Y_1 = n) \\ &= \sum_{n=1}^{\infty} \left[ \frac{n(n+1)}{2} (B^+ p^+ - B^- p^-) + n(B^- - B^+) \right] \\ &\quad \times ((p^+)^n (p^-) + (p^-)^n (p^+)) \\ &= B^+ p^+ \left( \left( \frac{p^+}{p^-} \right)^2 - 1 \right) - B^- p^- \left( \left( \frac{p^-}{p^+} \right)^2 - 1 \right). \end{aligned} \quad (1.37)$$

Note that, if  $p^+ = \frac{1}{2}$  (which holds for any symmetric random walk), then  $E(G_1) = 0$ , and as  $p^+ \downarrow 0$  or  $p^+ \uparrow 1$ ,  $E(G_1) \rightarrow \infty$ .

### 1.4.2 SIMPLE RANDOM WALK, CUSUM TIMING

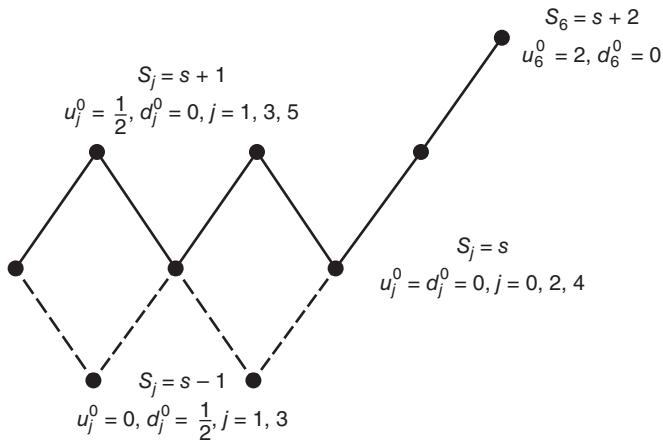
We now calculate the expected return of the first subperiod for a simple random walk asset price, applying CUSUM timing. Set our CUSUM threshold to  $h = 1$ , and our probability measure to the simple asymmetric random walk on ticks, that is,  $N = 1$ , with  $p_1 = p$ ,  $p_{-1} = 1 - p$  for some  $0 < p < 1$ . With  $M = 1$ , we have by (1.23), for every  $k \geq 0$ ,

$$m_k^u = S_{T_k} + \frac{1}{2}, \quad m_k^d = S_{T_k} - \frac{1}{2}.$$

Since  $X_j \in \{-1, 1\}$  for every  $0 \leq j < T_1$ , the possible values of  $u_j^0$  and  $d_j^0$ , by (1.24), are  $\{0, \frac{1}{2}, 2\}$ , where a 2 occurs only with two consecutive ticks of the same type (ending on an even step).  $T_1$  is the first time  $2j$  such that  $u_{2j}^0 \geq 1$  or  $d_{2j}^0 \geq 1$ . Hence,

$$\begin{aligned} T_1 = T_1^+ = 2j &\iff X_{k+1} = -X_k \quad \forall k, 1 \leq k < 2j-1, \quad X_{2j-1} = X_{2j} = 1; \\ T_1 = T_1^- = 2j &\iff X_{k+1} = -X_k \quad \forall k, 1 \leq k < 2j-1, \quad X_{2j-1} = X_{2j} = -1. \end{aligned}$$

Given  $S_0 = s > 0$ ,  $S_{T_1}$  can take only two possible values, from the paths described earlier. For  $j = 1, 2, \dots$ , each possibility takes the form of a geometric random variable conditioned on the final two steps  $X_{T_1-1}, X_{T_1}$ ,



**FIGURE 1.8** The four possible SRW paths for  $T_1^+ = 2(3) = 6$ .

where a “failure” is a sequence of two steps of opposite direction; that is,  $+1$  then  $-1$ , or  $-1$  then  $+1$ .

$$\begin{aligned} \{T_1 = T_1^+\} &\iff \{S_{T_1} = s + 2\}; \\ P(S_{T_1} = s + 2, T_1^+ = 2j) &= [2p(1-p)]^{j-1} p^2 \quad (1.38) \\ \{T_1 = T_1^-\} &\iff \{S_{T_1} = s - 2\}; \\ P(S_{T_1} = s - 2, T_1^- = 2j) &= [2p(1-p)]^{j-1} (1-p)^2. \end{aligned}$$

An illustration of the paths leading to a “+” signal  $T_1^+$  is shown in Figure 1.8. The probabilities of each value of  $S_{T_1}$  occurring are

$$\begin{aligned} P(S_{T_1} = s + 2) &= \sum_{j=1}^{\infty} [2p(1-p)]^{j-1} p^2 = \frac{p^2}{1 - 2p(1-p)} \\ P(S_{T_1} = s - 2) &= \sum_{j=1}^{\infty} [2p(1-p)]^{j-1} (1-p)^2 = \frac{(1-p)^2}{1 - 2p(1-p)}. \end{aligned}$$

Since there is only one possible outcome per signal type, these match the probabilities of each type of signal occurring:

$$\begin{aligned} p^+ &:= P(T_1 = T_1^+) = P(S_{T_1} = s + 2) = \frac{p^2}{1 - 2p(1-p)}; \\ p^- &:= P(T_1 = T_1^-) = P(S_{T_1} = s - 2) = \frac{(1-p)^2}{1 - 2p(1-p)}. \end{aligned} \quad (1.39)$$

The increment  $Z_k = S_{T_{k+1}} - S_{T_k}$  then takes values in  $\{-2, 2\}$  and depends on the sign of the signal of  $T_{k+1}$ . Conditioned on this signal sign, and by the strong Markov property at  $T_k$ , we get the conditional expectations

$$B^+ = E(Z_k | T_{k+1} = T_{k+1}^+) = 2P(S_{T_1} - s = 2 | T_1 = T_1^+) = 2, \quad (1.40)$$

$$B^- = E(Z_k | T_{k+1} = T_{k+1}^-) = -2P(S_{T_1} - s = -2 | T_1 = T_1^-) = -2. \quad (1.41)$$

Thus, by (1.34), (1.40), (1.41), and (1.37), we have the expected gain

$$E(G_1) = \frac{2(p^4 - (1-p)^4)}{1 - 2p(1-p)} \left[ \frac{p^2}{(1-p)^4} - \frac{(1-p)^2}{p^4} \right], \quad (1.42)$$

which can be shown to be symmetric about its minimum  $p = \frac{1}{2}$  (at  $E(G_1) = 0$ ), with  $\lim_{p \downarrow 0} E(G_1) = \lim_{p \uparrow 1} E(G_1) = \infty$ .

We also have the expected time until a signal occurs: by (1.38) and (1.39),

$$\begin{aligned} E(T_1 | T_1 = T_1^+) &= \sum_{j=1}^{\infty} (2j) P(T_1 = 2j | T_1 = T_1^+) \\ &= 2 \sum_{j=1}^{\infty} j \frac{P(T_1 = 2j, T_1 = T_1^+)}{P(T_1 = T_1^+)} = \frac{2p^2}{p^+} \sum_{j=1}^{\infty} j [2p(1-p)]^{j-1} \\ &= \frac{2p^2}{p^+ (1 - [2p(1-p)])^2} = \frac{2}{1 - [2p(1-p)]}; \\ E(T_1 | T_1 = T_1^-) &= \frac{2}{1 - [2p(1-p)]}; \\ E(T_1) &= E(T_1 | T_1 = T_1^+) p^+ + E(T_1 | T_1 = T_1^-) p^- \\ &= \frac{2}{1 - [2p(1-p)]}. \end{aligned} \quad (1.43)$$

Finally, the expected number of same-sign signals in a subperiod is, by (1.36) and (1.39),

$$E(Y_1) = \frac{p^+}{p^-} + \frac{p^-}{p^+} = \frac{p^2}{(1-p)^2} + \frac{(1-p)^2}{p^2} = \frac{p^4 + (1-p)^4}{p^2(1-p)^2}. \quad (1.44)$$

### 1.4.3 LAZY SIMPLE RANDOM WALK, CUSUM TIMING

Introducing a more complicated random walk distribution, such as a lazy simple random walk, with step distribution

$$X_j = \begin{cases} +1 & \text{with probability } p_1 \\ 0 & \text{with probability } p_0 \\ -1 & \text{with probability } p_{-1}, \end{cases} \quad (1.45)$$

where  $p_{-1} + p_0 + p_1 = 1$  increases the complexity of the analysis of the CUSUM timing strategy probabilities, and therefore of calculating the expected gain analytically. We will retain  $h = 1$  and  $M = 1$ .

By introducing a zero tick, we expand the possible cases of “failure” to set off a CUSUM signal. We decompose the lazy random walk path into seven distinct possible components. First, there are three possible patterns that fail to set off a signal, being “up-down” (with probability  $p_1 p_{-1}$ ), “down-up” (with probability  $p_{-1} p_1$ ), and “zero” (a one-step pattern with probability  $p_0$ ). Note that the first two of these are the two possible failure patterns of (1.38). There are, consequently, four “success” patterns:

- the two from (1.38): “up-up” (with probability  $p_1^2$ ) and “down-down” (with probability  $p_{-1}^2$ );
- and two patterns with zero ticks: “up-zero” (with probability  $p_1 p_0$ ) and “down-zero” (with probability  $p_{-1} p_0$ ).

The number of such patterns that occurs up to a signal time is geometric. Define

$$S^* := p_1^2 + p_{-1}^2 + p_1 p_0 + p_{-1} p_0 \quad (1.46)$$

$$F^* := 1 - S^* = 2p_1 p_{-1} + p_0 \quad (1.47)$$

as the respective signal-pattern success and failure probabilities. Then, we define  $V_j$  as the number of failure patterns until signal  $j = 1, 2, \dots$ .  $V_j \sim \text{geom}(S^*)$  (starting at 0), and, conditioned on  $V_j$ , we define  $W_j$  as the number of zero-tick patterns that occur during this time frame. Since the  $W_j$  zero-ticks can take place at any pattern position of the  $V_j$  patterns,  $W_j|_{V_j} \sim \text{bin}(V_j, \frac{p_0}{F^*})$ . Note that if  $p_0 = 0$ , this reduces to the case in the previous section.

We can calculate the expected time before a signal: if there are  $V_j$  failure patterns (of length 1 or 2 ticks) before signal  $j$ ,  $W_j$  of these are the 1-tick zero-tick failures, and, finally, we have a 2-tick success pattern, then the number of ticks before the first signal is

$$T_1 := W_1 + 2(V_1 - W_1) + 2 = 2V_1 - W_1 + 2. \quad (1.48)$$

The expected time until a signal is, then, by (1.48) and (1.47),

$$\begin{aligned} E(T_1) &= 2E(V_1) - E(W_1) + 2 = 2E(V_1) - \sum_{v=0}^{\infty} E(W_1 | V_1 = v)P(V_1 = v) + 2 \\ &= 2E(V_1) - \frac{p_0}{F^*} \sum_{v=0}^{\infty} vP(V_1 = v) + 2 \\ &= \frac{2(1 - S^*)}{S^*} \left( 1 - \frac{p_0}{F^*} \right) + 2 = \frac{2 - p_0}{S^*}. \end{aligned} \quad (1.49)$$

At  $p_0 = 0$ , (1.49) reduces to (1.43).

The zero-tick success patterns increase the possible asset values at a signal. In (1.40) and (1.41), the only possible values for the price change increment  $Z_k$  of (1.9) are  $\{-2, 2\}$ . Here, the possible values of  $Z_k$  are  $\{-2, -1, 1, 2\}$ , and so, by the Markov property at the times  $j - 2$ , and defining  $P_j^T := P(T_1 = j)/S^*$  for  $j \geq 2$ , we have the probabilities

$$\begin{aligned} P(S_{T_1} = s + 2) &= \sum_{j=2}^{\infty} P(S_{T_1} = s + 2, T_1^+ = j) \\ &= p_1^2 \sum_{j=2}^{\infty} P(S_{j-2} = s, T_1 > j - 2) = p_1^2 \sum_{j=2}^{\infty} P_j^T \\ P(S_{T_1} = s + 1) &= p_1 p_0 \sum_{j=2}^{\infty} P_j^T, \quad P(S_{T_1} = s - 1) = p_{-1} p_0 \sum_{j=2}^{\infty} P_j^T, \\ P(S_{T_1} = s - 2) &= p_{-1}^2 \sum_{j=2}^{\infty} P_j^T, \end{aligned} \quad (1.50)$$

which all sum to 1, by the fact that  $\sum_{j=2}^{\infty} P_j^T = \frac{1}{S^*}$ . The equations in (1.50) also yield the conditional probabilities

$$\begin{aligned} P(S_{T_1} = s+2 \mid T_1 = T_1^+) &= \frac{p_1^2}{p_1^2 + p_1 p_0}, \\ P(S_{T_1} = s+1 \mid T_1 = T_1^+) &= \frac{p_1 p_0}{p_1^2 + p_1 p_0}, \\ P(S_{T_1} = s-1 \mid T_1 = T_1^-) &= \frac{p_{-1} p_0}{p_{-1}^2 + p_{-1} p_0}, \\ P(S_{T_1} = s-2 \mid T_1 = T_1^-) &= \frac{p_{-1}^2}{p_{-1}^2 + p_{-1} p_0}. \end{aligned} \quad (1.51)$$

An illustration of possible paths leading to a “+” signal can be found in Figure 1.9.

Retaining the definitions of  $p^+$  and  $p^-$  from (1.26), we get

$$p^+ = \frac{p_1 p_0 + p_1^2}{S^*}; \quad p^- = \frac{p_{-1} p_0 + p_{-1}^2}{S^*}, \quad (1.52)$$

which allows us to calculate the expected number of signals on a subperiod. By (1.52) and (1.36),

$$E(Y_1) = \frac{p^+}{p^-} + \frac{p^-}{p^+} = \frac{p_1 p_0 + p_1^2}{p_{-1} p_0 + p_{-1}^2} + \frac{p_{-1} p_0 + p_{-1}^2}{p_1 p_0 + p_1^2}, \quad (1.53)$$

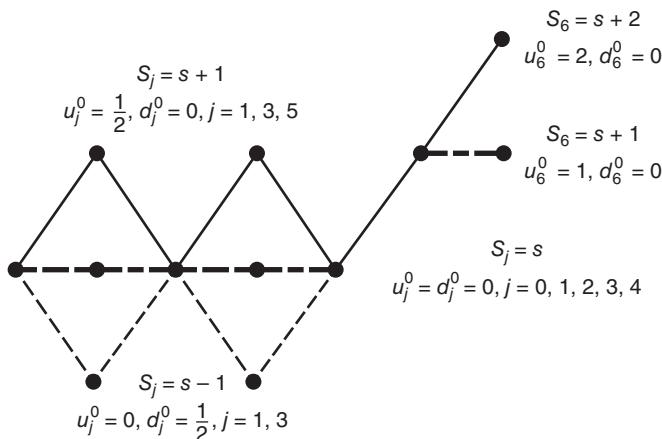


FIGURE 1.9 The 12 possible LSRW paths for  $T_1^+ = 6$ .

which reduces, if  $p_0 = 0$ , to (1.44). Also, if the walk is symmetric, that is,  $p_1 = p_{-1}$ , then  $E(Y_1) = 2$ .

Next, we find  $B^+$  and  $B^-$ , the expected size of the incremental changes  $Z_k$ , conditioned on the type of subperiod. Generalizing (1.40) and (1.41) (where  $p_0 = 0$ ), we have by (1.51)

$$\begin{aligned} B^+ &= 2P(S_{T_1} - s = 2 \mid T_1 = T_1^+) + P(S_{T_1} - s = 1 \mid T_1 = T_1^+) \\ &= 1 + \frac{p_1^2}{p_1 p_0 + p_1^2}, \end{aligned} \quad (1.54)$$

$$\begin{aligned} B^- &= -2P(S_{T_1} - s = -2 \mid T_1 = T_1^-) - P(S_{T_1} - s = -1 \mid T_1 = T_1^-) \\ &= -1 - \frac{p_{-1}^2}{p_{-1} p_0 + p_{-1}^2}. \end{aligned} \quad (1.55)$$

Finally, the expected gain  $E(G_1)$  at the end of a subperiod can be found by combining (1.37) with (1.52), (1.54), and (1.55), generalizing the  $p_0 = 0$  case (1.42).

## 1.5 CUSUM strategy Monte Carlo

Here we provide Monte Carlo simulations of the collection of random walks on ticks given in the previous section to numerically analyze the behavior of our strategy against such walks as asset prices.

The two classes of random walks for our simulations are special subclasses of (1.45): they are the lazy symmetric simple random walk

$$X_j = \begin{cases} +1 & \text{with probability } p_1 = \frac{1-p_0}{2} \\ 0 & \text{with probability } p_0 \in \{0, 0.05, 0.1, \dots, 0.35\} \\ -1 & \text{with probability } p_{-1} = \frac{1-p_0}{2}, \end{cases} \quad (1.56)$$

and the lazy asymmetric simple random walk with upward drift

$$X_j = \begin{cases} +1 & \text{with probability } p_1 = 0.5 - \frac{p_0}{2} + 0.05j, j \in \{0, 1, \dots, 6\} \\ 0 & \text{with probability } p_0 \in \{0, 0.1, 0.2, 0.3, 0.4\} \\ -1 & \text{with probability } p_{-1} = 1 - p_1 - p_0, \end{cases} \quad (1.57)$$

where  $j$  allows  $p_{-1} > 0$ . Each class of walks was run for 200 simulated trading days, with  $N = 5000$  ticks for 1 day's trading, and starting price  $s = 10,000$  ticks each day (to guarantee that 1 day's trading does not bottom out the asset).

Define the *idle time* of a trading strategy during a day as the (random) set of tick times between subperiods, *that is*, when our algorithm declares that our portfolio be empty. If the day consists of  $N$  ticks, then the idle time for the day is defined as

$$\text{idle time} := \{n \in \{1, 2, \dots, N\} : \Delta_n = 0\}.$$

The % idle time in a day is simply  $\frac{|\text{idle time}|}{N}$ . If there are  $R$  subperiods in a day, this is

$$|\text{idle time}| = \sum_{l=0}^{R-1} (T_{\alpha(l)+1} - T_{\alpha(l)}) + (N - T_{\alpha(R)}),$$

where  $T_{\alpha(R)} = N$  if the final subperiod's end is induced by the end-of-day settling the algorithm requires. We can estimate the average number of subperiods per day by  $\frac{N}{E(T)[E(Y)+1]}$ , and so, since there is the length of one signal between each subperiod, we can naively estimate the average amount of idle time in a day as the average number of subperiods per day multiplied by the average time to a signal, that is,  $\frac{N}{E(T)[E(Y)+1]} \cdot E(T_1) = \frac{N}{E(Y)+1}$ . Then, the % idle time in a day is naively estimated by this value divided by the number of ticks per day, or, simply,  $\frac{1}{E(Y)+1}$ .

Tables containing the results of simulations can be found in the Appendix, Section 1.7. Table 1.A.1 contains experimental averages of the following values for the lazy symmetric random walks represented by (1.56):

- average gain per subperiod (1.37), which can be seen to be close to  $E(G_1) = 0$  in all cases due to symmetry;
- average subperiod length, which approximates (1.49) and (1.53)'s  $E(T_1)E(Y_1)$ : for example,  $p_0 = 0.1$  has  $7.670 \approx \left(\frac{2-0.1}{1-2(0.45)^2-0.1}\right)(2) = 3.\overline{83}(2) = 7.\overline{67}$ ;
- average number of signals per subperiod, which approximates (1.53)'s  $E(Y_1) + 1$ : for example,  $p_0 = 0.1$  has  $2.998 \approx 2 + 1 = 3$ ;

- average number of subperiods per day, which approximates  $R$  above (which is itself approximated above by  $\frac{N}{E(T)[E(Y)+1]}$ ); for  $p_0 = 0.1$ , this is  $435.185 \approx \frac{5000}{3.83(3)} \approx 434.21$ ;
- and the average % idle time; for  $p_0 = 0.1$ , this is  $33.2\% \approx \frac{1}{E(Y)+1} = \frac{1}{3}$ .

The remaining tables contain similar experimental data for various lazy simple random walks from Section 1.4.3. Results of simulations using frequencies derived from the real data from the 5-year and 30-year bonds are shown in Tables 1.6, 1.7, and 1.8.

Table 1.A.2 contains detail on the subperiods of these walks:

- the average number of subperiods with a specific number of signals; for example,  $p_0 = 0.1$ , subperiod length  $n = 4$  has 27.31, which, when divided by the average total number of subperiods 435.185 from Table 1.A.1, gives  $\frac{27.31}{435.185} \approx 0.06275 \approx P(Y_1 = 4) = 0.0625$  from (1.35) using (1.52);
- and the average gain on such a subperiod of length  $n = 4$ , which is  $3.64 \approx E(G_1 | Y_1 = 4) = 3.478$  from (1.34) using (1.52), (1.54), and (1.55).

Tables 1.A.3 and 1.A.4 contain the same experimental values as Tables 1.A.1 and 1.A.2, this time from the simple random walk of Section 1.4.2. For example, in Table 1.A.3, examining  $p_1 = 0.65$ , we have

- average gain per subperiod  $16.378 \approx E(G_1) = 16.481$  from (1.42);
- average subperiod length  $13.749 \approx E(Y_1) \cdot E(T_1) \approx 3.7389 \cdot 3.6697 = 13.7208$  from (1.44) and (1.43);
- average number of signals per subperiod  $4.746 \approx E(Y_1) + 1 = 4.7389$  from (1.44);
- average number of subperiods  $288.120 \approx \frac{N}{E(T)[E(Y)+1]} = \frac{5000}{3.6697(4.7389)} = 287.516$ ; and
- average % idle time  $20.8\% \approx \frac{1}{E(Y)+1} = \frac{1}{4.7389} = 21.10\%$ .

Note that, for the simple random walk without a “lazy” probability  $p_0$ , the average amount of idle time per simulation (the percentage of ticks between subperiods) drops as the walk becomes more asymmetric, as the expected amount of time to get a signal (1.43) (and so be in a subperiod)

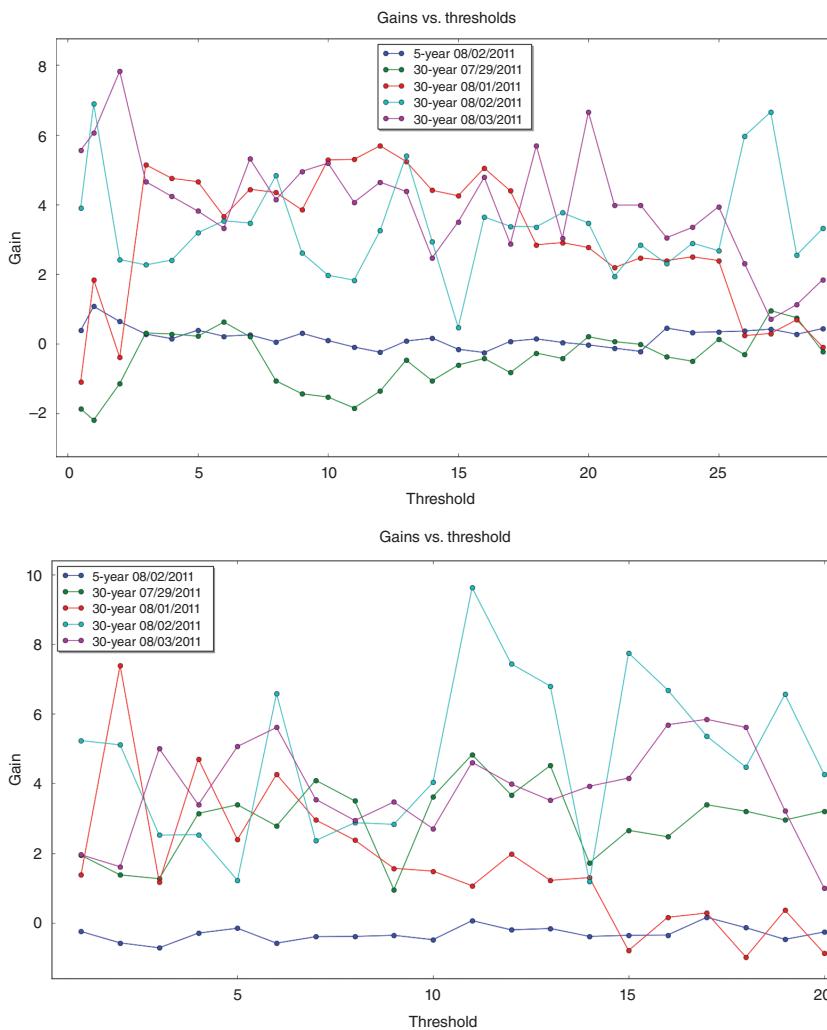
drops. In Table 1.A.4, the first row of each block approximates (1.35) multiplied by the average number of subperiods from Table 1.A.3 for that  $p = p_1$ , and the second row approximates (1.34), which is  $n^2 - 3n$  for  $n$  same-sign signals.

## 1.6 The effect of the threshold parameter

In this section, we discuss the effect of varying the threshold parameter  $h$  on the proposed trading strategy. We first examine this effect on the real data in Section 1.3.3. In particular, Figure 1.10 summarizes the effect of varying thresholds on the gain in all 5 US Treasury bonds of Section 1.3.3. In this figure, it is shown that varying the threshold does not change the sign of the gain. In fact, varying the threshold in the 5-year note leaves the daily gain almost unchanged, while in the 30-year bonds, although a more random variation is observed, no apparent pattern of an increasing or decreasing effect on the gain is observed. This demonstrates a level of robustness of the proposed strategy's gain as a function of the threshold. A closer examination shows that the number of signals per subperiod is almost constant, regardless of the threshold size, as shown in the column "average # of signals per subperiod" in Tables 1.1–1.5. Yet, the number of subperiods per trading day decreases as the threshold increases. This is shown in Figure 1.11, where we note that the number of signals per trading day decreases at the rate of the square root of the threshold. The decrease in the number of subperiods on a given trading day as a result of an increase in the threshold is to be expected since the quantity that varies when the threshold varies is the number of ticks, or equivalently, the amount of time as measured by ticks, required before the completion of a given subperiod. This is true because a smaller threshold gives rise to a more sensitive CUSUM stopping time. In fact, the expected time to a signal (CUSUM alarm) increases as the threshold increases in the order of the square root of the threshold, that is,

$$E[T_1(h)] \approx E[T_1(1)]\sqrt{h}. \quad (1.58)$$

To justify (1.58), note that, on a trend, one of the CUSUM statistics from (1.24) increases quadratically, regardless of the threshold  $h$ . For example,



**FIGURE 1.10** Total gain versus thresholds for 5-year and 30-year notes. Top graph: Small thresholds that vary as follows:  $0.5 \times \text{tick}$ ,  $\text{tick}$ ,  $2 \times \text{tick}$ ,  $3 \times \text{tick}$ , etc. (up to  $29 \times \text{tick}$ ). Bottom graph: Large thresholds that vary as follows:  $50 \times \text{tick}$ ,  $2 \times 50 \times \text{tick}$ ,  $3 \times 50 \times \text{tick}$ , etc. (up to  $20 \times 50 \times \text{tick}$ ).

on an upward trend,  $u_n^k = O(n^2)$ , and so the amount of time  $n$  it takes to break the threshold, that is, the minimum  $n$  to achieve  $u_n^k \geq h$ , is found by observing  $O(n^2) \approx h \Rightarrow n \approx O(\sqrt{h})$ . The coefficient can then be found by checking the baseline threshold  $h = 1$ . We also offer empirical evidence for this from bond data and Monte Carlo simulations.

**TABLE 1.1 5-year 08/02/2011 note. Tick size is  $M = 0.00078125$ , the number of ticks is  $N = 17,074$ . Each threshold  $h$  in column 1 is actually  $hM$ .**

Threshold	# subperiods	# signals	Average gain per subperiod	Average subperiod length	Average # of signals per subperiod	% of idle time	Total gain
0.5	488	1468	0.001	23.340	3.008	33.290	0.385
1	412	1259	0.003	28.488	3.056	31.258	1.074
2	331	1030	0.002	36.332	3.112	29.565	0.637
3	278	864	0.001	43.209	3.108	29.647	0.279
4	241	744	0.001	50.440	3.087	28.804	0.147
5	210	662	0.002	58.895	3.152	27.562	0.389
6	197	609	0.001	60.675	3.091	29.993	0.217
7	187	575	0.001	63.588	3.075	30.356	0.256
8	178	540	0.000	65.157	3.034	32.072	0.054
9	158	500	0.002	74.911	3.165	30.678	0.304
10	155	484	0.001	76.729	3.123	30.344	0.093
50	70	197	-0.004	154.271	2.814	36.752	-0.249
100	44	131	-0.013	245.205	2.977	36.810	-0.577
150	38	111	-0.019	287.921	2.921	35.920	-0.718
200	29	79	-0.010	356.862	2.724	39.387	-0.297
250	27	76	-0.006	397.074	2.815	37.209	-0.157

**TABLE 1.2** 30-year 07/29/2011 note. Tick size is  $M = 0.015625$ , the number of ticks is  $N = 4588$ . Each threshold  $n$  in column 1 is actually  $hM$ .

Threshold	# subperiods	# signals	Average gain per subperiod	Average subperiod length	Average # of signals per subperiod	% of idle time	Total gain
0.5	467	1334	-0.004	6.503	2.857	33.806	-1.865
1	335	988	-0.007	8.931	2.949	34.786	-2.191
2	233	681	-0.005	13.309	2.923	32.411	-1.142
3	192	565	0.002	15.901	2.943	33.457	0.311
4	174	504	0.002	17.632	2.897	33.130	0.282
5	146	430	0.001	20.651	2.945	34.285	0.218
6	127	375	0.005	24.055	2.953	33.413	0.624
7	123	360	0.002	24.894	2.927	33.261	0.204
8	113	331	-0.009	26.186	2.929	35.506	-1.062
9	104	306	-0.014	28.471	2.942	35.462	-1.436
10	103	302	-0.015	28.573	2.932	35.854	-1.530
50	34	115	0.057	97.676	3.382	27.616	1.937
100	23	75	0.060	145.391	3.261	27.114	1.375
150	18	59	0.070	180.778	3.278	29.076	1.266
200	13	50	0.242	278.077	3.846	21.207	3.140
250	14	47	0.242	229.714	3.357	29.904	3.390

**TABLE 1.3** 30-year 08/01/2011 bond. Tick size is  $M = 0.015625$ , the number of ticks is  $N = 3244$ . Each threshold  $h$  in column 1 is actually  $hM$ .

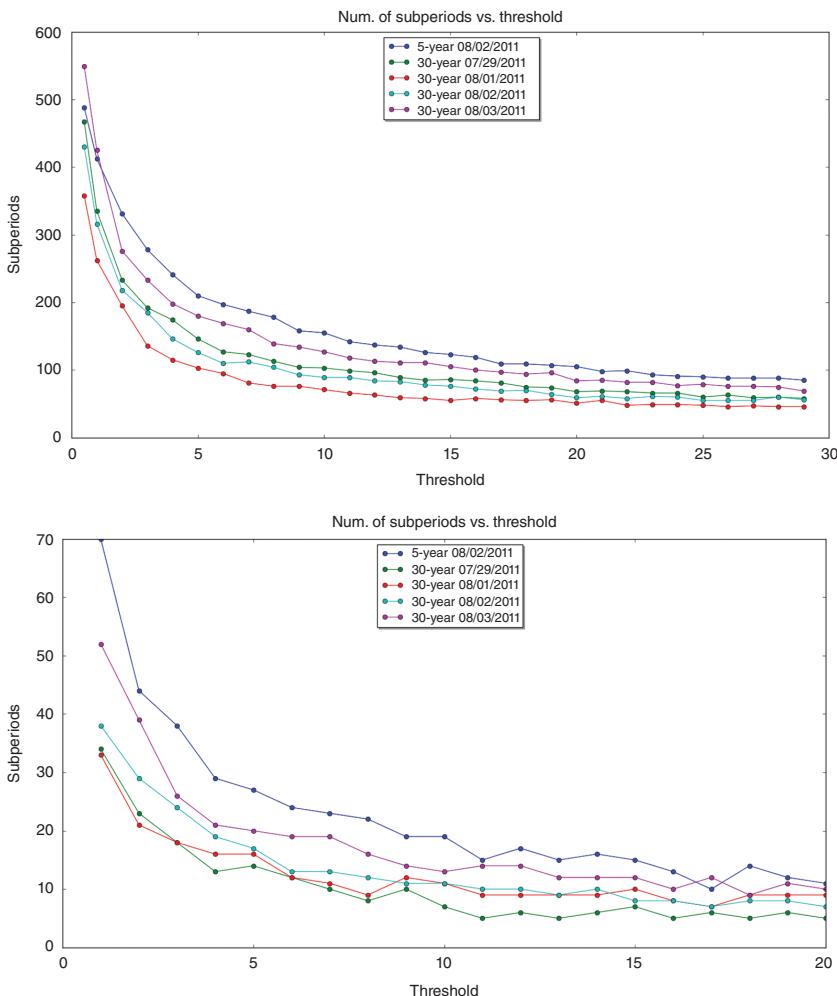
Threshold	# subperiods	# signals	Average gain per subperiod	Average subperiod length	Average # of signals per subperiod	% of idle time	% of total gain
0.5	358	1036	-0.003	6.095	2.894	32.737	-1.100
1	262	792	0.007	8.531	3.023	31.104	1.839
2	195	582	-0.002	11.031	2.985	33.693	-0.392
3	136	457	0.038	17.669	3.360	25.925	5.140
4	115	393	0.041	19.826	3.417	29.716	4.750
5	103	356	0.045	22.136	3.456	29.716	4.656
6	95	324	0.038	23.579	3.411	30.949	3.655
7	81	291	0.055	28.815	3.593	28.052	4.437
8	76	268	0.057	30.500	3.526	28.545	4.344
9	76	254	0.051	30.039	3.342	29.624	3.844
10	71	251	0.074	32.338	3.535	29.223	5.280
50	33	111	0.042	76.818	3.364	21.856	1.375
100	21	70	0.352	112.333	3.333	27.281	7.390
150	18	57	0.065	114.500	3.167	36.467	1.172
200	16	52	0.294	142.312	3.250	29.809	4.703
250	16	46	0.149	139.750	2.875	31.073	2.391

**TABLE 1.4** 30-year 08/02/2011 bond. Tick size is  $M = 0.015625$ , the number of ticks is  $N = 4349$ . Each threshold  $h$  in column 1 is actually  $hM$ .

Threshold	# subperiods	# signals	Average gain per subperiod	Average subperiod length	Average # of signals per subperiod	% of idle time	Total gain
0.5	430	1291	0.009	6.802	3.002	32.743	3.901
1	316	982	0.022	9.497	3.108	30.996	6.888
2	218	704	0.011	13.298	3.229	33.341	2.421
3	185	604	0.012	16.411	3.265	30.191	2.264
4	146	509	0.016	20.856	3.486	29.984	2.404
5	126	444	0.025	24.397	3.524	29.317	3.186
6	110	393	0.032	28.391	3.573	28.190	3.531
7	112	382	0.031	26.804	3.411	30.973	3.468
8	104	354	0.046	28.577	3.404	31.662	4.829
9	93	321	0.028	33.269	3.452	28.857	2.609
10	89	298	0.022	32.921	3.348	32.628	1.967
50	38	129	0.138	80.763	3.395	29.432	5.233
100	29	90	0.176	105.207	3.103	29.846	5.110
150	24	75	0.105	128.375	3.125	29.156	2.516
200	19	61	0.133	179.895	3.211	21.407	2.531
250	17	54	0.072	178.706	3.176	30.145	1.219

**TABLE 1.5** 30-year 08/03/2011 note. Tick size is  $M = 0.015625$ , the number of ticks is  $N = 5153$ . Each threshold  $h$  in column 1 is actually  $hM$ .

Threshold	# subperiods	# signals	Average gain per subperiod	Average subperiod length	Average # of signals per subperiod	% of idle time	Total gain
0.5	549	1674	0.010	6.342	3.049	32.428	5.556
1	425	1311	0.014	8.169	3.085	32.622	6.058
2	276	900	0.028	12.641	3.261	32.292	7.827
3	233	741	0.020	14.670	3.180	33.670	4.654
4	198	634	0.021	17.662	3.202	32.137	4.232
5	180	576	0.021	19.444	3.200	32.078	3.810
6	169	532	0.020	20.604	3.148	32.428	3.325
7	160	507	0.033	22.331	3.169	30.662	5.311
8	139	457	0.030	25.612	3.288	30.914	4.141
9	134	438	0.037	26.560	3.269	30.933	4.953
10	127	414	0.041	28.307	3.260	30.235	5.187
50	52	172	0.038	68.404	3.308	30.972	1.953
100	39	125	0.041	92.282	3.205	30.157	1.609
150	26	92	0.192	150.077	3.538	24.277	5.001
200	21	73	0.161	185.571	3.476	24.374	3.391
250	20	68	0.253	194.600	3.400	24.471	5.063



**FIGURE 1.11 Number of subperiods versus thresholds for 5-year and 30-year notes. Top graph: Small thresholds that vary as follows: 0.5\*tick, tick, 2\*tick, 3\*tick, etc. (up to 29\*tick). Bottom graph: Large thresholds that vary as follows 50\*tick, 2\*50\*tick, 3\*50\*tick, etc. (up to 20\*50\*tick).**

The fifth columns of Tables 1.1–1.5 represent the average subperiod length  $E[T_{\alpha(1)}(h)]$ ; the sixth columns represent the average # of signals per subperiod  $E[Y_1] + 1$ . Therefore, the expected time to a signal can be found as

$$E[T_1(h)] = \frac{E[T_{\alpha(1)}(h)]}{E[Y_1]} \quad (1.59)$$

**TABLE 1.6** Average gains and lengths of subperiods, asymmetric random walk.  $P(-3) = 0.00012$ ,  $P(-2) = 0.00141$ ,  $P(-1) = 0.05348$ ,  $P(0) = 0.88619$ ,  $P(1) = 0.05670$ ,  $P(2) = 0.00182$ , and  $P(3) = 0.00029$ , for various thresholds. These probabilities were computed from the 5-year 08/02/2011 bond. Tick size is 1, the number of ticks is 5000 and starting price 10,000. The number of simulations is 1000.

Threshold	# subperiods	# signals	Average gain per subperiod	Average subperiod length	Average # of signals per subperiod	% of idle time	Total gain
1	162.401	488.191	0.030	20.648	3.006	32.936	4.9
2	125.799	378.175	0.036	26.665	3.006	32.913	4.5
3	103.358	310.196	0.041	32.508	3.001	32.801	4.3
4	88.207	264.402	0.034	38.121	2.998	32.749	3.0
5	77.574	232.413	0.037	43.390	2.996	32.681	2.9
6	69.419	208.576	0.052	48.530	3.005	32.623	3.6
7	63.469	190.459	0.066	53.123	3.001	32.566	4.2
8	58.775	176.465	0.077	57.408	3.002	32.517	4.5
9	54.911	164.795	0.072	61.455	3.001	32.509	4.0
10	51.568	154.781	0.062	65.533	3.001	32.411	3.2
11	48.860	146.595	0.077	69.171	3.000	32.406	3.8
12	46.511	139.607	0.078	72.754	3.002	32.323	3.6
13	44.562	133.657	0.085	75.995	2.999	32.270	3.8
14	42.796	128.334	0.108	79.217	2.999	32.197	4.6
15	41.267	123.685	0.084	82.236	2.997	32.127	3.5
100	14.414	43.619	0.588	245.445	3.026	29.243	8.5
200	9.628	29.632	1.451	381.445	3.078	26.549	14.0
300	7.636	23.615	2.119	489.755	3.093	25.205	16.2
400	6.477	20.074	2.487	588.740	3.099	23.735	16.1
500	5.672	17.819	3.057	687.449	3.142	22.016	17.3

**TABLE 1.7** Average gains and lengths of subperiods, asymmetric random walk.  $P(-4) = 0.00494$ ,  $P(-3) = 0.00997$ ,  $P(-2) = 0.03732$ ,  $P(-1) = 0.12561$ ,  $P(0) = 0.62919$ ,  $P(1) = 0.12747$ ,  $P(2) = 0.04575$ , and  $P(3) = 0.01279$ ,  $P(4) = 0.00697$ , for various thresholds. These probabilities were computed from the 30-year 07/29/2011 and 08/02/2011 bonds after a best fit that was verified by chi-square test. The *p*-value for 07/29/2011 is 0.87 and for 08/02/2011, 0.815. Tick size is 1, the number of ticks is 5000, and the starting price 10,000. The number of simulations is 1000.

Threshold	# subperiods	# signals	Average gain per subperiod	Average subperiod length	Average # of signals per subperiod	% of idle time	Total gain
1	443.934	1336.262	0.083	7.535	3.010	33.099	37.0
2	299.194	903.318	0.161	11.208	3.019	32.935	48.3
3	241.959	731.714	0.240	13.879	3.024	32.838	58.0
4	204.308	618.324	0.289	16.452	3.026	32.776	59.1
5	181.312	549.742	0.352	18.563	3.032	32.684	63.8
6	164.086	498.442	0.421	20.548	3.038	32.569	69.1
7	150.539	457.415	0.453	22.410	3.039	32.529	68.2
8	139.256	423.863	0.507	24.255	3.044	32.446	70.6
9	130.517	397.679	0.549	25.909	3.047	32.369	71.6
10	122.775	374.570	0.617	27.547	3.051	32.358	75.8
11	115.983	354.361	0.683	29.171	3.055	32.333	79.2
12	110.395	337.981	0.775	30.676	3.062	32.269	85.6
13	105.205	322.336	0.848	32.226	3.064	32.194	89.2
14	100.997	309.365	0.902	33.602	3.063	32.127	91.1
15	96.980	297.191	0.927	34.964	3.064	32.183	89.9
100	30.065	96.746	5.522	117.792	3.218	29.172	166.0
200	19.133	63.974	10.603	190.220	3.344	27.210	202.9
300	14.566	50.358	16.365	258.656	3.457	24.648	238.4
400	11.936	42.703	22.771	323.669	3.578	22.734	271.8
500	10.217	37.682	28.222	388.293	3.688	20.656	288.3

**TABLE 1.8** Average gains and lengths of subperiods, asymmetric random walk.  $P(-4) = 0.01023$ ,  $P(-3) = 0.01460$ ,  $P(-2) = 0.04601$ ,  $P(-1) = 0.13666$ ,  $P(0) = 0.58916$ ,  $P(1) = 0.12914$ ,  $P(2) = 0.04548$ ,  $P(3) = 0.01772$ , and  $P(4) = 0.01100$ , for various thresholds. These probabilities were computed from the 30-year 08/01/2011 and 08/03/2011 bonds after a best fit that was verified by chi-square test. The *p*-value for 08/01/2011 is 0.80 and for 08/03/2011, 0.48. Tick size is 1, the number of ticks is 5000 and starting price 10,000. The number of simulations is 1000.

Threshold	# subperiods	# signals	Average gain per subperiod	Average subperiod length	Average # of signals per subperiod	% of idle time	Total gain
1	488.364	1463.672	-0.014	6.834	2.997	33.249	-7.0
2	333.138	998.589	-0.017	10.031	2.998	33.169	-5.6
3	271.557	813.925	-0.015	12.306	2.997	33.162	-4.0
4	229.944	688.463	-0.043	14.532	2.994	33.170	-9.9
5	204.761	613.066	-0.039	16.340	2.994	33.085	-7.9
6	185.382	555.241	-0.034	18.050	2.995	33.077	-6.4
7	170.343	509.950	-0.055	19.636	2.994	33.102	-9.4
8	157.385	471.073	-0.049	21.245	2.993	33.127	-7.7
9	147.463	441.897	-0.020	22.710	2.997	33.022	-3.0
10	138.931	416.129	-0.012	24.127	2.995	32.960	-1.6
11	131.472	393.089	-0.049	25.474	2.990	33.018	-6.5
12	125.161	374.424	-0.053	26.768	2.992	32.993	-6.6
13	119.276	356.992	-0.026	28.096	2.993	32.976	-3.0
14	114.317	341.780	-0.075	29.313	2.990	32.980	-8.5
15	109.788	328.314	-0.068	30.509	2.990	33.009	-7.5
100	35.056	104.634	0.031	96.983	2.985	32.003	1.1
200	22.908	68.150	-0.009	149.200	2.975	31.642	-0.2
300	17.932	53.188	-0.089	192.092	2.966	31.108	-1.6
400	14.932	44.283	-0.198	232.382	2.966	30.602	-3.0
500	13.070	38.673	-0.084	267.170	2.959	30.162	-1.1

To see the square-root effect, let us examine the following rows:

- In rows 2 and 5 of Table 1.2, we can calculate  $E[T_1(1)] = \frac{8.931}{1.949} = 4.582$  and  $E[T_1(4)] = \frac{17.632}{1.897} = 9.295$ , respectively. We now notice that  $E[T_1(4)] \approx E[T_1(1)]\sqrt{4}$ .
- In row 4 of Table 1.2, we can calculate  $E[T_1(3)] = \frac{15.901}{1.943} = 8.184$  and  $E[T_1(3)] \approx E[T_1(1)]\sqrt{3}$ .
- In rows 2 and 8 of Table 1.3, we can calculate  $E[T_1(1)] = \frac{8.531}{2.023} = 4.217$  and  $E[T_1(7)] = \frac{28.815}{2.593} = 11.113$ , respectively. This leads to  $E[T_1(7)] \approx E[T_1(1)]\sqrt{7}$ .

We have also generated simulated data for each of the bonds from which once again we can easily decipher the same square-root effect. To be more specific, we have fitted a lazy random walk model to the 30-year bond series data for 07/29/2011 and 08/02/2011 with the appropriate parameters as designated in the caption of Table 1.7. A simple goodness-of-fit test demonstrates the validity of the model selected. The same process is followed in the remaining 30-year bond data. The results of the simulations are summarized in Tables 1.7 and 1.8, respectively. We again demonstrate the square-root effect once again for the same thresholds used in the observed data:

- In rows 1, 4, and 3 of Table 1.7, we can calculate  $E[T_1(1)] = \frac{7.535}{2.01} = 3.749$ ,  $E[T_1(4)] = \frac{16.452}{2.026} = 8.12$  and  $E[T_1(3)] = \frac{13.879}{2.024} = 6.857$ . Once again we observe the approximations  $E[T_1(4)] \approx E[T_1(1)]\sqrt{4}$  and  $E[T_1(3)] \approx E[T_1(1)]\sqrt{3}$ , respectively.
- In rows 1 and 7 of Table 1.8, we can calculate  $E[T_1(1)] = \frac{6.834}{1.997} = 3.422$  and  $E[T_1(7)] = \frac{19.636}{1.994} = 9.847$ , from which we can extract the approximation  $E[T_1(7)] \approx E[T_1(1)]\sqrt{7}$ .

The square-root effect suggests that increasing the threshold reduces the number of complete subperiods R on any given trading day, and thus the number of transactions completed therein. Thus, although varying the threshold does not have a systematic effect on the gain in the absence of transaction costs, increasing the threshold would decrease the number of transactions but increase the “riskiness” of the trading strategy. A good measure of performance of the strategy over the course of an entire day

of trading is the *total gain*  $\sum_{l=1}^R G_l$ , which, under the zero transaction cost model with IID  $G_l$ , has expected value of  $E(R)E(G_1)$  by Wald's equation. Examining this product as a function of  $h$  under different probabilistic models of the asset price is of interest, especially in terms of maximizing the day's total gain based on the value of  $h$ . However, in the presence of transaction costs, Wald's equation fails and analytical derivations are extremely challenging. Besides, transaction costs often vary from firm to firm and thus the appropriate choice of threshold will depend not only on the selection of a measure of "riskiness" but also on the transaction costs related to the specific product or firm.

## 1.7 Conclusions and future work

---

In Figures 1.2, 1.3, 1.4, 1.5, and 1.6 of Section 1.3.3, it is shown that the proposed CUSUM trading strategy performs well in subperiods of many signals of one sign before a signal of the opposite sign occurs. This is also evident in Tables 1.3, 1.5, 1.7, 1.A.1, and 1.A.3 related to the results of the simulation in the random walk model of Section 1.4. Such subperiods are characterized by consistent upward or downward trends in prices. On the contrary, the proposed strategy is at a loss in the case of few signals of one sign followed by a signal of the opposite sign. Such subperiods are characterized by stability in prices. This observation suggests that the CUSUM trading strategy can be further improved by an online detection of "regimes of stability" (as contrasted to "regimes of trends"). This suggests the construction of new online algorithm possibly inspired by computer vision (see, for instance, Hadjiliadis and Stamos [19] or Stamos et al. (30)).

Another statistic that is indicative of the contrast between times of stability versus times of instability is known as the *speed of reaction* of the CUSUM, which measures the time between the last reset to 0 of the CUSUM statistic process and the time of the CUSUM alarm (see, for instance, [31]). We intend to examine both of these directions of research in order to improve the performance of the proposed algorithm by limiting trading in times of stability.

A parameter that should be investigated in depth is the transaction cost  $c$ . The form of the gain over a subperiod given in (1.11) can be written as

$$G_l = (-1)^{A_l} \sum_{j=1}^{Y_l} j Z_{j+\alpha(l-1)} + c \left( \sum_{j=1}^{Y_l} j Z_{j+\alpha(l-1)} - 2 Y_l S_{T_{\alpha(l)}} \right). \quad (1.60)$$

This second term should be analyzed as a fixed percentage, and on a sliding scale (considering, e.g., high-volume rebates). In addition, transaction costs should be investigated via Monte Carlo simulation, as it requires knowledge of the liquidation price of the asset for that subperiod.

A parameter closely related to the transaction cost is the threshold parameter  $h$  used in the CUSUM timing. A smaller threshold implies more frequent transactions but decreases the “riskiness” of the strategy on any given trading day. The optimal choice of the threshold should thus be based on the trade-off between an appropriately chosen measure of “risk” of the proposed strategy and the transaction costs in the market where it is applied.

Moreover, it should be noted that the random walk examples included here are not intended as actual asset price models (we do not intend to commit a Bachelierian fallacy); these models are merely used to illustrate the strategy and allow for basic calculations. In future work, it would be of interest to examine the best fit random walk model to actual high-frequency asset data (taking into account such real-world considerations as the bid-ask spread). Furthermore, open problems on this topic include extending analysis of this strategy to other models of asset price motion—primarily, building a binomial model (of which our random walks are the simplest case) and limiting to a continuous geometric Brownian motion. Note that our two sets of random walks investigate different types of “time”: the  $p_0 = 0$  case investigates “tick time,” where the clock moves only when the price moves, and the lazy walk, that is,  $p_0 > 0$ , considers clock time (since there may be samples where the price does not move). This simple discrepancy induces extra possible paths into the CUSUM timing process. The general binomial model, which may move a price multiple ticks per sample, and still retain the probability of standing still, is certainly, then, of interest.

Finally, we wish to examine the CUSUM strategy with  $m_k^u$  and  $m_k^d$  set to wait for multiple ticks instead of one (e.g.,  $m_k^u = S_{T_k} + \frac{bM}{2}$  for some  $b > 1$ ).

---

## Appendix: Tables

---

In this section are the tables described in Section 1.5.

**TABLE 1.A.1** Average gains and lengths of subperiods, lazy simple symmetric random walk.

$p_1$	Average total gain	Average gain per subperiod	Average subperiod length	Average # signals per subperiod	Average # subperiods	Average % idle time
0.000	-0.864	-0.002	7.995	2.998	417.472	33.2
0.050	-5.379	-0.013	7.824	2.999	426.918	33.2
0.100	-3.204	-0.007	7.670	2.998	435.185	33.2
0.150	1.720	0.004	7.579	3.001	440.642	33.2
0.200	4.417	0.010	7.505	3.002	445.020	33.2
0.250	7.205	0.016	7.492	3.003	446.076	33.2
0.300	-3.728	-0.008	7.465	2.996	446.920	33.3
0.350	-0.144	-0.000	7.527	3.000	443.581	33.2

**TABLE 1.A.2** Number of subperiods of a specified length, and average gain over those subperiods, lazy simple symmetric random walk.

$p_0$	0.000	0.050	0.100	0.150	0.200	0.250	0.300	0.350
# 1-subperiods	209.05	213.37	218.10	220.11	222.32	223.14	223.83	222.24
average gain, 1-subperiod	-1.99	-1.90	-1.81	-1.73	-1.66	-1.60	-1.53	-1.48
# 2-subperiods	104.61	106.94	108.50	110.43	111.67	111.18	111.86	110.48
average gain, 2-subperiod	-1.99	-1.90	-1.81	-1.74	-1.66	-1.59	-1.53	-1.48
# 3-subperiods	51.87	53.50	54.30	54.94	55.27	55.98	56.05	55.32
average gain, 3-subperiod	0.01	0.01	0.01	0.01	-0.01	0.01	0.01	0.02
# 4-subperiods	26.12	26.60	27.31	27.88	27.95	27.92	27.73	27.80
average gain, 4-subperiod	4.01	3.82	3.64	3.50	3.36	3.23	3.09	2.97
# 5-subperiods	12.81	13.33	13.50	13.64	13.91	13.80	13.62	13.99
average gain, 5-subperiod	10.02	9.55	9.11	8.80	8.36	7.99	7.74	7.44
# 6-subperiods	6.49	6.66	6.79	6.86	6.97	6.88	6.94	6.87
average gain, 6-subperiod	18.01	17.26	16.47	15.71	14.99	14.44	13.87	13.26
# 7-subperiods	3.30	3.33	3.36	3.38	3.40	3.67	3.48	3.45
average gain, 7-subperiod	28.03	26.55	25.50	24.37	23.35	22.48	21.71	20.72
# 8-subperiods	1.58	1.60	1.67	1.65	1.83	1.77	1.68	1.74
average gain, 8-subperiod	40.04	38.00	36.13	34.79	33.05	31.75	30.97	29.57
# 9-subperiods	0.81	0.76	0.86	0.88	0.85	0.86	0.84	0.83
average gain, 9-subperiod	54.03	51.29	48.82	47.12	45.01	43.09	41.67	39.80
# 10 + -subperiods	0.83	0.82	0.81	0.88	0.87	0.88	0.89	0.85
average gain, 10 + -subperiods	2091.33	1393.27	1852.19	1524.21	1230.49	1715.07	1220.75	1318.50

**TABLE 1.A.3 Average gains and lengths of subperiods, asymmetric random walk,  $p_0 = 0.0$ .**

$p_1$	Average total gain	Average gain per subperiod	Average subperiod length	Average # signals per subperiod	Average # subperiods	Average % idle time
0.500	-1.005	-0.002	8.014	3.001	417.140	33.1
0.550	414.950	1.039	8.572	3.165	399.230	31.6
0.600	1,818.090	5.170	10.413	3.705	351.630	26.8
0.650	4,718.815	16.378	13.749	4.746	288.120	20.8
0.700	10,597.510	48.681	19.612	6.667	217.695	14.6
0.750	22,358.635	144.829	29.404	10.181	154.380	9.2
0.800	47,017.940	468.563	47.264	17.080	100.345	5.1

**TABLE 1.A.4** Number of subperiods of a specified length, and average gain over those runs, asymmetric random walk.  
 $p_0 = 0.0$ .

$p_1$	0.500	0.550	0.600	0.650	0.700	0.750	0.800
# 1-subperiods	209.12	192.44	149.00	100.33	56.83	27.99	11.00
average gain, 1-subperiod	-1.99	-1.99	-1.99	-1.99	-1.99	-1.99	-1.98
# 2-subperiods	103.48	95.11	75.41	50.27	29.31	13.47	5.46
average gain, 2-subperiod	-1.99	-2.00	-1.99	-1.99	-1.98	-1.96	-1.94
# 3-subperiods	51.97	49.91	43.08	32.50	20.64	11.08	5.09
average gain, 3-subperiod	0.01	0.01	0.01	0.01	0.03	0.03	0.06
# 4-subperiods	26.20	26.66	27.11	24.35	17.00	10.82	4.56
average gain, 4-run	4.02	4.02	4.02	4.03	4.03	4.08	4.13
# 5-runs	13.76	14.95	17.92	17.76	14.63	9.12	4.30
average gain, 5-run	10.02	10.02	10.02	10.02	10.03	10.08	10.13
# 6-runs	6.18	8.39	11.95	14.38	12.02	8.35	4.28
average gain, 6-run	18.02	18.03	18.04	18.05	18.07	18.13	18.22
# 7-runs	3.29	4.87	8.50	10.80	10.24	7.50	3.98
average gain, 7-run	28.01	28.03	28.00	28.05	28.12	28.10	28.21
# 8-runs	1.70	2.85	5.97	8.47	8.98	6.49	3.79
average gain, 8-run	40.02	40.00	40.03	40.01	40.08	40.08	40.17
# 9-runs	0.69	1.59	3.98	6.70	7.41	5.95	3.31
average gain, 9-run	54.00	54.05	54.11	54.07	54.14	54.19	54.19
# 10 + -runs	0.75	2.44	8.73	22.56	40.63	53.60	54.59
average gain, 10 + -runs	1356.00	2688.00	5096.28	18,888.69	68,970.76	198,829.18	63,3281.20

**TABLE 1.A.5 Average gains and lengths of runs, asymmetric random walk,  $p_0 = 0.1$ .**

$p_1$	Average total gain	Average gain per subperiod	Average subperiod length	Average # signals per subperiod	Average # subperiods	Average % idle time
0.450	7.485	0.017	7.683	3.005	434.700	33.2
0.500	428.825	1.032	8.251	3.173	415.625	31.4
0.550	1,825.045	4.979	9.972	3.705	366.535	26.9
0.600	4,729.605	15.804	13.254	4.767	299.270	20.7
0.650	10,741.020	47.569	18.924	6.727	225.800	14.5
0.700	22,858.345	144.381	28.694	10.364	158.320	9.1
0.750	50,243.475	504.478	47.822	17.994	99.595	4.7

**TABLE 1.A.6** Number of subperiods of a specified length, and average gain over those runs, asymmetric random walk.  
 $p_0 = 0.1$ .

$p_1$	0.450	0.500	0.550	0.600	0.650	0.700	0.750
# 1-runs	217.16	199.44	156.22	104.08	58.37	27.67	10.54
avg gain, 1-run	-1.81	-1.81	-1.80	-1.80	-1.78	-1.77	-1.74
# 2-runs	108.08	99.44	78.25	51.39	29.41	14.06	5.47
avg gain, 2-run	-1.81	-1.79	-1.76	-1.69	-1.63	-1.51	-1.34
# 3-runs	54.92	51.84	44.15	34.02	21.74	11.57	4.70
avg gain, 3-run	0.01	0.04	0.14	0.31	0.47	0.64	0.94
# 4-runs	27.02	28.05	28.40	25.00	18.00	10.38	4.34
avg gain, 4-run	3.61	3.73	3.88	4.12	4.29	4.58	4.96
# 5-runs	14.05	15.66	18.58	19.49	15.38	9.06	4.29
avg gain, 5-run	9.02	9.27	9.48	9.82	10.07	10.51	11.06
# 6-runs	6.57	8.90	12.41	14.54	13.19	8.81	4.07
avg gain, 6-run	16.28	16.64	16.96	17.45	17.66	18.25	18.87
# 7-runs	3.58	4.81	8.80	11.25	10.62	7.71	3.69
avg gain, 7-run	25.49	25.62	26.25	26.72	27.38	27.79	28.72
# 8-runs	1.64	2.96	6.07	8.87	8.64	6.85	3.26
avg gain, 8-run	36.23	36.95	37.57	37.91	38.84	39.49	40.15
# 9-runs	0.85	1.78	4.13	6.91	7.64	6.20	3.40
avg gain, 9-run	49.42	50.21	50.84	50.91	51.87	52.39	53.87
# 10 + -runs	0.83	2.73	9.52	23.73	42.81	56.00	55.85
avg gain, 10 + -runs	1119.42	2305.77	6948.10	15,781.54	61,048.80	178,378.58	711,759.83

## References

1. Abramov, V., Khan, M. K., and Khan, R. A., A probabilistic analysis of the trading the line strategy, *Quantitative Finance*, Vol. 8, No. 5, pp. 499–512, 2008.
2. Alexander, S. S., Price movements in speculative markets: trends or random walk, *Industrial Management Review*, Vol. 2, No. 2, pp. 7–26, 1961.
3. Alexander, S. S., Price movements in speculative markets: trends or random walk no. 2, *Industrial Management Review*, Vol. 5, No. 2, pp. 338–372, 1964.
4. Barnard, G., Control charts and stochastic processes, *Journal of the Royal Statistical Society B*, Vol. 11, No. 1, pp. 239–271, 1959.
5. Basseville, M., and Nikiforov, I., *Detection of abrupt changes: Theory and applications*, 1993 (Prentice Hall, Englewood Cliffs, NJ).
6. Beibel, M., A note on Ritov's Bayes approach to the minimax property of the CUSUM procedure, *Annals of Statistics*, Vol. 24, No. 2, pp. 1804–1812, 1996.
7. Bozdog, B., Florescu, I., Khashanah K., and Wang, J., A study of persistence in price movement using high-frequency financial data. In *Handbook of modeling high-frequency data in finance*, edited by F. Viens, M. Mariani, and I. Florescu, pp. 27–46, 2012 (John Wiley & Sons, Hoboken, NJ).
8. Brodsky, B. E., and Darkhovsky, B. K., *Nonparametric Methods in Change-Point Problems* (Kluwer, Dordrecht), 1993.
9. Broemling, L. D., and Tsurumi, H., *Econometrics and Structural Change*, 1987 (Marcel Dekker, New York).
10. Cohen, A., *Biomedical Signal Processing*, 1987 (CRC Press, Boca Raton, FL).
11. Devore, J., *Probability and Statistics for Engineering and the Science, 8th edition*, 2012 (Brooks/Cole, Boston, MA).
12. Fama, B., and Blum, M., Filter rules and stock market trading, *Journal of Business*, Vol. 40, pp. 226–241, 1966.
13. Figueroa-Lopez, J. E., Lancette, S., Lee, K., and Yanhun, M., Estimation of NIG and VG models for high-frequency financial data. In *Handbook of modeling high-frequency data in finance*, edited by

- F. Viens, M. Mariani, and I. Florescu, pp. 3–26, 2012 (John Wiley & Sons, Hoboken, NJ).
14. Frisen, M., (Ed.), *Financial surveillance*, 2008 (John Wiley and Sons, Hoboken, NJ).
  15. Glynn, P. W., and Iglehart, D. L., Trading securities using trailing stops, *Management Science*, Vol. 41, No. 6, pp. 1096–1106, 1995.
  16. Hadjiliadis, O., and Moustakides, G. V., Optimal and asymptotically optimal CUSUM rules for change point detection in the Brownian motion model with multiple alternatives, *Theory of Probability and Its Applications*, Vol. 50, No. 1, pp. 131–144, 2006.
  17. Hadjiliadis, O., Optimality of the 2-CUSUM drift equalizer rules for detecting two-sided alternatives in the Brownian motion model, *Journal of Applied Probability*, Vol. 42, No. 4, pp. 1183–1193, 2005.
  18. Hadjiliadis, O., Hernandez-del-Valle, G., and Stamos, I., A comparison of 2-CUSUM stopping rules for quickest detection of two-sided alternatives in a Brownian motion model, *Journal of Sequential Analysis*, Vol. 28, No. 1, pp. 92–114, 2009.
  19. Hadjiliadis, O., and Stamos, I., Sequential classification in point clouds of urban scenes, In *Proceedings of 5th International Symposium on 3D Data Processing Visualization and Transmission*, Paris, France, May 17–20, 2010.
  20. Lam, K., and Yam, H. C. CUSUM techniques for technical trading in financial markets, *Financial Engineering and the Japanese Markets*, Vol. 4, pp. 257–274, 1997.
  21. Lorden, G., Procedures for reacting to a change in distribution, *Annals of Mathematical Statistics*, Vol. 42, No. 6, pp. 1897–1908, 1971.
  22. Moustakides, G. V., Optimal stopping times for detecting changes in distributions, *Annals of Statistics*, Vol. 14, No. 4, pp. 1379–1387, 1986.
  23. Moustakides, G. V., Optimality of the CUSUM procedure in continuous time, *Annals of Statistics*, Vol. 32, No. 1, pp. 302–315, 2004.
  24. Page, E. S., Continuous inspection schemes, *Biometrika*, Vol. 41, Nos. 12, pp. 100–115, 1954.
  25. Poor, H. V., and Hadjiliadis, O., *Quickest Detection*, 2008 (Cambridge University Press, Cambridge, UK).

26. Shiryaev, A. N., On optimum methods in quickest detection problems, *Theory of Probability and Its Applications*, Vol. 8, No. 1, pp. 22–46, 1963.
27. Shiryaev, A. N., Minimax optimality of the method of cumulative sums (CUSUM) in the continuous case, *Russian Mathematical Surveys*, Vol. 51, No. 4, pp. 750–751, 1996.
28. Siegmund, D., *Sequential Analysis*, 1985 (Springer-Verlag, New York).
29. Sonesson, C., and Bock, D., A review and discussion of prospective statistical surveillance in public health, *Journal of the Royal Statistical Society: Series A*, Vol. 166, No. 1, pp. 5–21, 2003.
30. Stamos, I., Hadjiliadis, O., Zhang, H., and Flynn, T., Online algorithms in the classification of urban objects in 3D point clouds, *Proceedings of the 2nd 3D Imaging Processing Visualization and Transmission Conference*, Zurich, Switzerland, October 13–15, 2012.
31. Zhang, H., and Hadjiliadis, O., Drawdowns and the speed of a market crash, *Methodology and Computing in Applied Probability*, Vol. 14, No. 3, pp. 739–752, 2012.

## Chapter Two

# Gaussian Inequalities and Tranche Sensitivities

Claas Becker<sup>1</sup> and Ambar N. Sengupta<sup>2</sup>

<sup>1</sup>Studiengang Angewandte Mathematik, Hochschule RheinMain,  
65197 Wiesbaden, Germany

<sup>2</sup>Department of Mathematics, Louisiana State University, Baton Rouge, LA, USA

## 2.1 Introduction

---

With the CDO (collateralized debt obligation) market picking up [16], it is important to build a stronger understanding of pricing and risk management models. The role of the Gaussian copula model, whose importance grew after the work of Li [11], has well-known deficiencies and has been criticized in the technical literature (e.g., Hull and White [8]) as well as in the popular press, but it continues to be fundamental as a starter model. In this chapter, we draw attention to the applicability of Gaussian inequalities in analyzing tranche loss sensitivity to correlation parameters for the Gaussian copula model. Some of these methods are also applicable to other models. Detailed proofs and more general results, applicable to non-Gaussian copulas, are available in our longer paper [2].

We work with an  $\mathbb{R}^N$ -valued Gaussian random variable  $X = (X_1, \dots, X_N)$ , where each  $X_j$  is normalized to mean 0 and variance 1, and study the equity tranche loss

$$L_{[0,a]} = \sum_{m=1}^N l_m 1_{[X_m \leq c_m]} - \left\{ \sum_{m=1}^N l_m 1_{[X_m \leq c_m]} - a \right\}_+,$$

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

where  $l_1, \dots, l_N > 0$ ,  $a > 0$ , and  $c_1, \dots, c_N \in \mathbb{R}$  are parameters. Our results establish an identity between the sensitivity of  $\mathbb{E}[L_{[0,a]}]$  to the correlations  $r_{jk} = \mathbb{E}[X_j X_k]$  and the parameters  $c_j$  and  $c_k$ . From this, we also prove (in [2]) the inequality

$$\frac{\partial \mathbb{E}[L_{[0,a]}]}{\partial r_{jk}} \leq 0. \quad (2.1)$$

Applying this inequality to a CDO containing  $N$  names whose default behavior is governed by the Gaussian variables  $X_j$  shows that an increase in name-to-name correlation decreases expected loss in an equity tranche. This is a generalization of the well-known result for Gaussian copulas with uniform correlation.

Our approach is in the spirit of Slepian inequalities (Slepian [15]) and our goal here is to bring attention to such Gaussian inequalities in the context of CDO loss behavior. Narrower forms of the inequality (2.1) were the focus of the papers by Hillebrand et al. [6, 7] and Meng and Sengupta [12]. Other related works include Cousin and Laurent [3], who use a general theory of stochastic orders (see Müller and Stoyan [13]). Jarrow and van Deventer [9] explore dependence of the risk of the equity tranche on correlation under various scenarios and different models. Ağça and Islam [1] explain how an increase in correlation can impact market-implied default probabilities, potentially increasing tranche losses.

## 2.2 The tranche loss function

Consider a CDO consisting of  $N$  names, with  $\tau_j$  denoting the (random) default time of the  $j$ -th name. Let

$$X_j = \Phi_j^{-1}(F_j(\tau_j)), \quad (2.2)$$

where  $F_j$  is the distribution function of  $\tau_j$  (relative to the market pricing measure), assumed to be continuous and strictly increasing, and  $\Phi_j$  is the standard Gaussian distribution function. Then for any  $x \in \mathbb{R}$  we have

$$\mathbb{P}[X_j \leq x] = \mathbb{P}\left[\tau_j \leq F_j^{-1}(\Phi_j(x))\right] = F_j\left(F_j^{-1}(\Phi_j(x))\right) = \Phi_j(x), \quad (2.3)$$

which means that  $X_j$  has standard Gaussian distribution. The Gaussian copula model posits that the joint distribution of the  $X_j$  is Gaussian; thus,

$$X = (X_1, \dots, X_N) \quad (2.4)$$

is an  $\mathbb{R}^N$ -valued Gaussian variable whose marginals are all standard Gaussian. The correlation

$$r_{jk} = \mathbb{E}[X_j X_k] \quad (2.5)$$

reflects the default correlation between the names  $j$  and  $k$ . Now let

$$p_j = \mathbb{E}[\tau_j \leq T] = \mathbb{P}[X_j \leq c_j], \quad (2.6)$$

be the probability that the  $j$ -th name defaults within a time horizon  $T$  (which we hold fixed through the entire analysis), and

$$c_j = \Phi_j^{-1}(F_j(T)) \quad (2.7)$$

is the *default threshold* of the  $j$ -th name.

In the schematic model, we use to explore the essential phenomena, the default of name  $j$ , which happens if the default time  $\tau_j$  is within the time horizon  $T$ , results in a loss of amount  $l_j > 0$  in the CDO portfolio. Thus, the total loss during the time period  $[0, T]$  is

$$L = \sum_{m=1}^N l_m 1_{[X_m \leq c_m]}. \quad (2.8)$$

In this simplified model, we are essentially working with a one-period CDO and are ignoring discounting from the random time of actual default; these are fairly standard assumptions for a first analysis.

A *tranche* is simply a range of loss for the portfolio; it is specified by a closed interval  $[a, b]$  with  $0 \leq a \leq b$ . If the loss  $x$  is less than  $a$ , then this tranche is unaffected, whereas if  $x \geq b$  then the entire tranche value  $b - a$  is eaten up by loss; in between, if  $a \leq x \leq b$ , the loss to the tranche is  $x - a$ . Thus, the *tranche loss function*  $t_{[a,b]}$  is given by

$$t_{[a,b]}(x) = \begin{cases} 0 & \text{if } x < a; \\ x - a & \text{if } x \in [a, b]; \\ b - a & \text{if } x > b. \end{cases} \quad (2.9)$$

We can rewrite this more compactly as

$$t_{[a,b]}(x) = (x - a)_+ - (x - b)_+. \quad (2.10)$$

From this, it is clear that  $t_{[a,b]}(x)$  is continuous in  $(a, b, x)$ , and from (2.9) we see that it is a non-decreasing function of  $x$ . Thus, the loss in an equity tranche  $[0, a]$  is given by

$$t_{[0,a]}(L) = L - (L - a)_+, \quad (2.11)$$

where  $a > 0$ .

## 2.3 A sensitivity identity

Gaussian inequalities were pioneered by Slepian [15] in the context of extrema of Gaussian processes. Such inequalities have been developed further by many researchers, such as Joag-Dev et al. [10]. Some of these inequalities are based on the following remarkable identity, first discovered by Plackett [14, equation (3)], for the density of a correlated centered multi dimensional Gaussian variable:

**Proposition 2.1.** *Let  $Q(R, x)$  be the Gaussian density*

$$Q(R, x) = (\det(2\pi R))^{-1/2} e^{-\frac{1}{2}\langle x, R^{-1}x \rangle}, \quad (2.12)$$

where  $R$  is a  $d \times d$  strictly positive-definite matrix  $R = [r_{jk}]$  and  $x \in \mathbb{R}^d$ . Then

$$\frac{\partial Q(R, x)}{\partial r_{jk}} = \begin{cases} \frac{1}{2} \frac{\partial^2 Q(R, x)}{\partial x_j^2} & \text{if } j = k; \\ \frac{\partial^2 Q(R, x)}{\partial x_j \partial x_k} & \text{if } j \neq k. \end{cases} \quad (2.13)$$

For a proof of this identity, see, for example, our longer paper [2], where we use the following basic and very convenient identity:

$$\mathbb{E}[e^{aX}] = e^{a\mathbb{E}[X] + \frac{a^2}{2}\text{var}(X)}, \quad (2.14)$$

for any Gaussian variable  $X$  and complex number  $a$ .

Using (2.13), we prove in [2] the following identity that is applicable to understanding expected tranche loss sensitivities:

**Theorem 2.1.** *Let  $(X_1, \dots, X_N)$  be an  $\mathbb{R}^N$ -valued Gaussian variable, with each  $X_m$  having mean 0 and variance 1, and covariance matrix  $R = [r_{jk}]$  that is strictly positive definite. Let*

$$L = \sum_{m=1}^N l_m 1_{[X_m \leq c_m]},$$

where  $c_1, \dots, c_N \in \mathbb{R}$ , and  $l_1, \dots, l_N > 0$ . Then

$$\partial_{r_{jk}} \mathbb{E}[(L - a)_+] = \partial_{c_j} \partial_{c_k} \mathbb{E}[(L - a)_+] \quad (2.15)$$

for any  $a \geq 0$ , and any distinct  $j, k \in \{1, \dots, N\}$ .

What we see here is a remarkable relation between the sensitivity of the expected loss

$$\mathbb{E}[L - (L - a)_+]$$

in the equity tranche  $[0, a]$  (see (2.11)) with respect to the correlation parameter  $r_{jk}$  and the default thresholds  $c_j$  and  $c_k$ .

## 2.4 Correlation sensitivities

We turn now to the nature of correlation sensitivities for the tranche loss. Our result from [2] is:

**Theorem 2.2.** *Let  $(X_1, \dots, X_N)$  be an  $\mathbb{R}^N$ -valued Gaussian variable, with each  $X_m$  having mean 0 and variance 1, and covariance matrix  $R = [r_{jk}]$  that is strictly positive definite. Let*

$$L = \sum_{j=1}^N l_j 1_{[X_j \leq c_j]},$$

where  $c_1, \dots, c_N \in \mathbb{R}$ , and  $l_1, \dots, l_N > 0$ . Then, for any  $a \geq 0$ ,

$$\partial_{r_{jk}} \mathbb{E}[(L - a)_+] = \partial_{c_j} \partial_{c_k} \mathbb{E}[(L - a)_+] \quad (2.16)$$

for any distinct values of  $j, k \in \{1, \dots, N\}$ .

We sketch here the essential ideas of the proof; a complete proof is available in [2].

**Sketch Proof.** The expected tranche loss  $\mathbb{E}[(L - a)_+]$  is given by

$$\mathbb{E}[(L - a)_+] = \int_{\mathbb{R}^N} (l_N(x) - a)_+ Q(R, x) dx, \quad (2.17)$$

where  $Q(R, x)$  is the Gaussian density

$$Q(R, x) = (\det(2\pi R))^{-1/2} e^{-\frac{1}{2}\langle x, R^{-1}x \rangle}, \quad (2.18)$$

and

$$l_N(x) = \sum_{m=1}^N 1_{(-\infty, c_m]}(x_m). \quad (2.19)$$

Taking the derivative with respect to the correlation  $r_{jk}$  and using the second Gaussian identity (2.13), we have

$$\begin{aligned}\partial_{r_{jk}} \mathbb{E}[(L - a)_+] &= \int_{\mathbb{R}^N} (l_N(x) - a)_+ \partial_{r_{jk}} Q(R, x) dx \\ &= \int_{\mathbb{R}^N} (l_N(x) - a)_+ \frac{\partial^2 Q(R, x)}{\partial x_j \partial x_k} dx.\end{aligned}\quad (2.20)$$

Now using integration by parts, we move the partial derivatives over to the first term and obtain, formally,

$$\int_{\mathbb{R}^N} \frac{\partial^2 (l_N(x) - a)_+}{\partial x_j \partial x_k} Q(R, x) dx.$$

This is formal because we have not given a precise meaning to the derivatives involved here. A precise meaning can be given either in the sense of distributions or, more simply, by working out the full integration-by-parts process in terms of limits of difference quotients instead of partial derivatives. Writing the partial derivative  $\partial_{x_j}$  as the limit of a difference quotient and using dominated convergence, and repeating this for  $x_k$ , we have

$$\partial_{r_{jk}} \mathbb{E}[(L - a)_+] = \lim_{\epsilon_j \downarrow 0, \epsilon_k \downarrow 0} \frac{1}{\epsilon_j \epsilon_k} \int_{\mathbb{R}^N} (l_N(x) - a)_+ [*] dx, \quad (2.21)$$

where

$$[*] = Q(R, x + \epsilon_j e_j + \epsilon_k e_k) - Q(R, x + \epsilon_j e_j) - Q(R, x + \epsilon_k e_k) + Q(R, x). \quad (2.22)$$

We split the integration on the right in (2.21) into four integrals corresponding to the terms in (2.22) and replace  $x_j$  by  $x_j - \epsilon_j$  and  $x_k$  by  $x_k - \epsilon_k$ , and then we use

$$1_{(-\infty, c]}(x - \epsilon) = 1_{(-\infty, c+\epsilon]}(x).$$

This relation allows us to transform the differences involving the  $x_j$  and  $x_k$  into differences involving  $c_j$  and  $c_k$ . Putting these ideas together, dividing by  $\epsilon_j$  and letting  $\epsilon_j \downarrow 0$ , and then dividing by  $\epsilon_k$  and letting  $\epsilon_k \downarrow 0$ , we obtain

$$\frac{\partial^2 \mathbb{E}[(L - a)_+]}{\partial c_j \partial c_k}.$$

Going back to (2.21), we conclude that

$$\partial_{r_{jk}} \mathbb{E}[(L - a)_+] = \partial_{c_j} \partial_{c_k} \mathbb{E}[(L - a)_+].$$

QED

We turn now to our main inequality.

**Theorem 2.3.** Let  $(X_1, \dots, X_N)$  be an  $\mathbb{R}^N$ -valued Gaussian variable, with each  $X_m$  having mean 0 and variance 1, and covariance matrix  $R = [r_{jk}]$  that is strictly positive definite. Let

$$L = \sum_{m=1}^N l_m 1_{[X_m \leq c_m]},$$

where  $c_1, \dots, c_N \in \mathbb{R}$ , and  $l_1, \dots, l_N > 0$ . Then, for any  $a \geq 0$ ,

$$\frac{\partial \mathbb{E}[(L - a)_+]}{\partial r_{jk}} \geq 0, \quad (2.23)$$

and

$$\frac{\partial \mathbb{E}[t_{[0,a]}(L)]}{\partial r_{jk}} \leq 0, \quad (2.24)$$

for all distinct  $j, k \in \{1, \dots, N\}$ .

Again, we present here only the essence of the argument and refer to [2] for details.

As we saw in the remarks following (2.21)

$$\frac{\partial \mathbb{E}[(L - a)_+]}{\partial r_{jk}}$$

is the limit of a difference quotient integrated against  $Q(R, x)$ . The difference term is of the form

$$\{l_1 + l_2 + w\}_+ - \{l_1 + w\}_+ - \{l_2 + w\}_+ + \{w\}_+, \quad (2.25)$$

where  $l_1, l_2, w \geq 0$ ; the term (2.25) is non negative and this leads to the inequality (2.23).

Next, we recall from (2.11) that the equity tranche loss is given by

$$t_{[0,a]}(L) = L - (L - a)_+.$$

The full portfolio loss  $L$  is insensitive to the correlations  $r_{jk}$  because it is simply the sum of the individual expected losses:

$$\mathbb{E}[L] = \sum_{j=1}^N l_j \mathbb{P}[X_j \leq c_j].$$

Hence

$$\frac{\partial \mathbb{E}[t_{[0,a]}(L)]}{\partial r_{jk}} = -\frac{\partial \mathbb{E}[(L - a)_+]}{\partial r_{jk}} \leq 0.$$

In [2], we show that these results extend to copulas more general than the Gaussian copula (such as elliptically contoured ones discussed in Gordon [5]).

## Acknowledgment

---

Sengupta acknowledges research support from NSA grant H98230-13-1-0210.

## References

1. Ağça, S., and Islam, S., 2010, Spring. Can CDO equity be short on correlation? *The Journal of Alternative Investments*, Vol. 12, No. 4, pp. 85–96.
2. Becker, C., and Sengupta, A. N., 2013. Identities and inequalities for tranche sensitivities. *Communications on Stochastic Analysis*, Vol. 7, No. 3, pp. 425–439.
3. Cousin, A., and Laurent, J.-P., 2008. Comparison results for exchangeable credit risk portfolios, *Insurance: Mathematics and Economics*, Vol. 42, No. 3, pp. 1118–1127.
4. Frey, R., and J. McNeil, A. J., Dependent defaults in models of portfolio credit risk, <http://www.ma.hw.ac.uk/~mcneil/ftp/defaults-jor.pdf>
5. Gordon, Y., 1987. Elliptically contoured distributions. *Probability Theory and Related Fields*, Vol. 76, pp. 429–438.
6. Hillebrand, E., Sengupta, A. N., and Xu, J., 2011. Impact of correlation fluctuations on securitized structures, in *Handbook of Modeling High-Frequency Data in Finance* (eds. F. G. Viens, M. C. Mariani, and I. Florescu), John Wiley & Sons, Inc., Hoboken, NJ.
7. Hillebrand, E., Sengupta, A. N., and Xu, J., 2012. Serially correlated defaults in subprime securitization, *Communications on Stochastic Analysis*, Vol. 6, No. 3, pp. 487–511.
8. Hull, J., and White, A., 2006. Valuing credit derivatives using an implied copula approach, *Journal of Derivatives*, Vol. 14, No 2, pp. 8–28.

9. Jarrow, R. A., and van Deventer, D. R., 2008. Synthetic CDO equity: short or long correlation, *Journal of Fixed Income*, Vol. 17, No. 4, pp. 31–41.
10. Joag-Dev, K., Perlman, M. D., and Pitt, L. D., 1983. Association of normal variables and Slepian's inequality, *The Annals of Probability*, Vol. 11, pp. 451–455.
11. Li, D., 2000, March. On default correlation: a copula function approach, *Journal of Fixed Income*, Vol. 9, No. 4, pp. 43–54. [http://www.defaultrisk.com/pp\\_corr\\_05.htm](http://www.defaultrisk.com/pp_corr_05.htm)
12. Meng, C., and Sengupta, A. N., 2011, June. CDO tranche sensitivities in the Gaussian copula model, *Communications on Stochastic Analysis*, Vol. 5, No. 2, pp. 387–403.
13. Müller, A., and Stoyan, D., 2002. *Comparison Methods for Stochastic Models and Risks*, John Wiley and Sons, Inc., Hoboken, NJ.
14. Plackett, R. L., 1954. Reduction formula for normal multivariate integrals, *Biometrika*, Vol. 41, pp. 351–360.
15. Slepian, D., 1962. The one-sided barrier problem for Gaussian noise, *Bell System Technical Journal*, Vol. 41, pp. 463–501.
16. CDOs are back: will they lead to another financial crisis?, *Wharton Business School*, <http://knowledge.wharton.upenn.edu/article.cfm?articleid=3230>

## Chapter Three

# A Nonlinear Lead Lag Dependence Analysis of Energy Futures: Oil, Coal, and Natural Gas<sup>1</sup>

Germán G. Creamer<sup>1</sup> and Bernardo Creamer<sup>2</sup>

<sup>1</sup>*School of Business, Stevens Institute of Technology, Hoboken, NJ, USA*

<sup>2</sup>*Universidad de las Américas, Quito, Ecuador*

### 3.1 Introduction

The main fossil fuels (oil, coal, and natural gas) have some common elements that affect their prices, such as emissions control or energy crises. Markets and political forces also affect fuel prices, especially, in the case of oil. It is also possible their prices are mutually determined or that one price depends on another one as they partially behave as substitute goods for the production of electricity.

Mohammadi (2011) finds that in the case of the United States, the oil and natural gas prices are globally and regionally determined, respectively, and coal prices are defined by long-term contracts. Mohammadi (2009), using cointegration analysis, exposes a strong relationship between electricity and coal prices and an insignificant relationship between electricity

---

<sup>1</sup>A preliminary version of this paper is included on the Proceedings of the AAAI 2014 Fall Symposium on Energy Market Predictions.

and oil and/or natural gas prices. Asche et al. (2003) and Bachmeier and Griffin (2006) find very weak linkages among oil, coal, and natural gas prices, using cointegration analysis, while crude oil and several refined product prices are integrated (Asche et al., 2003). Hartley et al. (2008) notice an indirect relationship between natural gas and oil prices. Even more, Aruga and Managi (2011) detect a weak market integration among a large group of energy products: West Texas Intermediate (WTI) oil, Brent oil, gasoline, heating oil, coal, natural gas, and ethanol futures prices.

Mjelde and Bessler (2009) observe that oil, coal, natural gas, and uranium markets are not fully cointegrated. Asche et al. (2006) indicate that the UK energy market between 1995 and 1998 was highly integrated where the demand was for energy rather than for a particular source of energy. Brown and Yucel (2008) show that oil and natural gas prices have been independent since 2000; however, when weather and inventories are taken into consideration in an error correction model, crude oil prices have an effect on natural gas prices. Similar results are obtained by Ramberg (2010), using cointegration analysis. Amavilah (1995) observes that oil prices influence uranium prices.

Most of the studies about fossil fuels mentioned earlier are based on cointegration analysis and Granger causality; however, none of these studies have used a nonlinear correlation measure such as the Brownian distance correlation proposed by Székely and Rizzo (2009). This distance correlation is a nonlinear multivariate dependence coefficient that can be used with random vectors of multiple dimensions. Among the different studies that use this distance correlation to finance problems, Creamer et al. (2013) evaluate the impact of corporate news networks on return and volatility; Grothe et al. (2014) measure the association of European bonds and stocks during the recent Euro crisis; Puliga et al. (2014) build networks of credit default swaps to forecast systemic risk, and Zhang et al. (2014) generate networks of the international shipping market to forecast systemic risk. This chapter conducts a lead–lag analysis of coal, oil, and gas with the Brownian distance correlation and compares its results with the well-known Granger causality test.

### 3.1.1 CAUSALITY ANALYSIS

Granger causality (Granger, 1969, 1980, 2001) is a very popular methodology used in economics, financial econometrics, and in many other areas of study, such as neuroscience, to evaluate the linear causal relationship

among two or more variables. According to the basic definition of Granger causality, the forecasting of the dependent variable  $Y_t$  with an autoregressive process using  $Y_{t-l}$  as its lag- $l$  value, should be compared with another autoregressive process using  $Y_{t-l}$  and the vector  $X_{t-l}$  of independent variables. So,  $X_{t-l}$  Granger causes  $Y_t$  when  $X_{t-l}$  happens before  $Y_t$ , and  $X_{t-l}$  has unique information to forecast  $Y_t$  that is not present in other variables.

Typically, Granger causality is tested using an autoregressive model with and without the vector  $X_{t-1}$ , such as in the following bivariate example:

$$\begin{aligned} Y_t &= \sum_{l=1}^L \alpha_l Y_{t-l} + \epsilon_1 \\ Y_t &= \sum_{l=1}^L \alpha_l Y_{t-l} + \sum_{l=1}^L \beta_l X_{t-l} + \epsilon_2 \end{aligned}$$

where the residual is a white noise series:

$$\epsilon_j \sim N(0, \sigma), j = 1, 2.$$

$X_{t-l}$  Granger causes  $Y_t$  if the null hypothesis  $H_0 : \beta_l = 0$  is rejected on the basis of the F test. The order of the autoregressive model is selected on the basis of the Akaike information criterion or the Bayesian information criterion.

Székely and Rizzo (2009) also proposed the Brownian distance covariance, which captures the covariance with respect to a stochastic process. Distance covariance ( $v(X, Y)$ ) between the random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  measures the distance between  $f_X f_Y$  and  $f_{X,Y}$  and is obtained as the square root of  $v^2(X, Y) = \|f_{X,Y}(t, s) - f_X(t)f_Y(s)\|^2$  where  $\|\cdot\|$  is the norm,  $t$  and  $s$  are vectors,  $f_X$  and  $f_Y$  are the characteristic functions of  $X$  and  $Y$ , respectively, and  $f_{X,Y}$  is the joint characteristic function of  $X$  and  $Y$ .

Empirically,  $v(X, Y)$  evaluates the null hypothesis of independence  $H_0 : f_X f_Y = f_{X,Y}$  versus the alternative hypothesis  $H_A : f_X f_Y \neq f_{X,Y}$ . In this chapter, this test is the distance covariance test of independence.

Likewise, distance variance ( $v(X)$ ) is the square root of  $v^2(X) = \|f_{X,X}(t, s) - f_X(t)f_X(s)\|^2$ .

Once distance covariance is defined, the distance correlation  $R(X, Y)$  is also defined in the following expression:

$$R^2 = \begin{cases} \frac{v^2(X, Y)}{\sqrt{v^2(X)v^2(Y)}}. & v^2(X)v^2(Y) > 0 \\ 0, & v^2(X)v^2(Y) = 0 \end{cases}$$

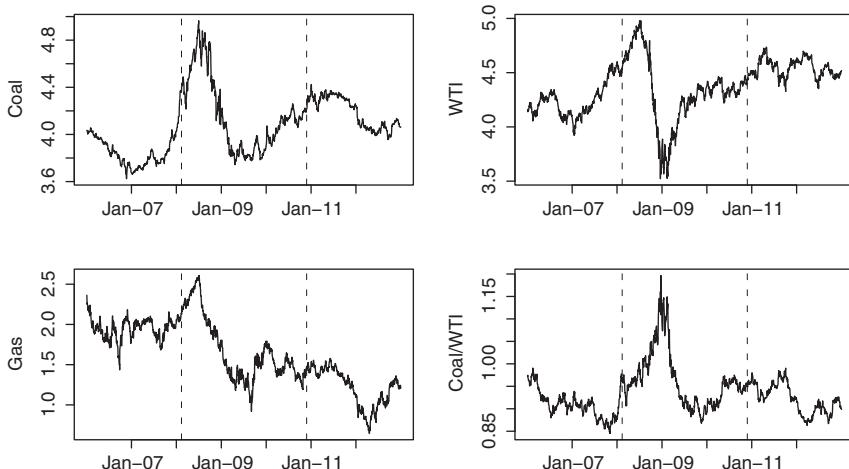
Distance correlation takes a value of 0 in case of independence and 1 when there is complete dependence.

### 3.2 Data

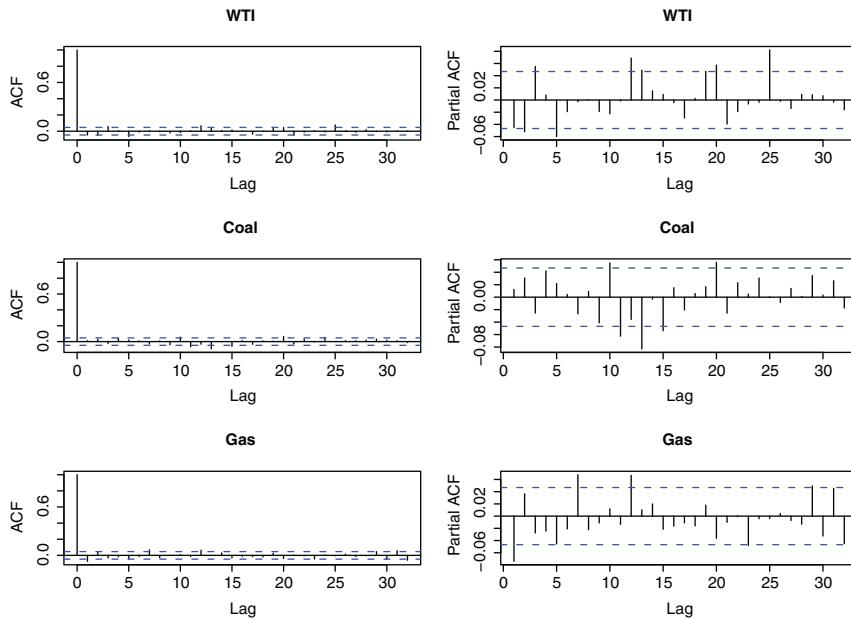
We used the daily time series of 1 month forward futures log prices of the fossil fuel series for the period 2006–2012: WTI oil, the Central Appalachian [bituminous] coal (Coal) and natural gas (Gas) from the New York Mercantile Exchange (see Figure 3.1). These series have some relevant autoregressive effects according to the autocorrelation function (ACF) and the partial ACF (see Figure 3.2); however, the emphasis of this chapter is on the lagged cross-correlation, which is explored in the next sections.

### 3.3 Estimation techniques

We evaluated the stationarity of the series, using the augmented Dickey–Fuller (ADF) test. We applied the Bai and Perron (1998)’s test to detect structural breaks of the coal/WTI log prices ratio, considering that these are the most dominant products of the causality analysis. The Bai–Perron test is particularly useful when the break date is unknown and there is



**FIGURE 3.1** Log prices by product. Horizontal lines represent structural breaks according to the Bai–Perron test of the Coal/WTI log prices ratio.



**FIGURE 3.2** ACF and partial ACF of log returns by product.

more than one break date. For the complete series and for each of the periods identified with the Bai–Perron test, we tested the nonlinearity of the series, using the White (Lee et al., 1993) and the Terasvirta test (Terasvirta et al., 1993). We also conducted a nonlinear lead–lag relationship analysis, using the Brownian distance correlation between each pair of variables and up to seven lags (1 week). We compared these results with the Granger causality test and evaluated the cointegration of the different pairs using the Johansen test (Johansen, 1988a, b) to decide if we had to use the VAR error correction model. In our analysis,  $\rightarrow$  denotes relationship. For instance,  $X \rightarrow Y$  indicates that X Granger causes Y when Granger causality is used or Y is dependent of X when the Brownian distance correlation is used. Therefore, the  $p$  value of every test evaluates only the effect of one variable into another one and is not affected by other time series.

## 3.4 Results

The Bai–Perron test applied to the Coal/WTI ratio series split the data into the following periods: January 3, 2006–January 17, 2008 (precrisis

**TABLE 3.1 Descriptive statistics of log prices.**

	2006–2012		
	Coal	WTI	Gas
Mean	4.08	4.37	1.62
SD	0.26	0.25	0.41
Skewness	0.79	-0.47	0.05
Kurtosis	0.34	0.48	-0.63

period), January 18, 2008–November 17, 2010 (financial crisis period), and November 18, 2010–December 31, 2012 (recovery period) (see Figure 3.1). We conducted our analysis in these different periods and in the complete series 2006–2012. The ADF test indicates that all log price series for the aforementioned periods are nonstationary and, as expected, the log return series are stationary. So, we used the log returns (the first difference of the log price) to conduct the causality tests.

The gas price distribution shows the highest volatility, although it is less skewed and flatter than the rest, according to Table 3.1 and Figure 3.1. The correlations between the price series significantly increase during the crisis period, as can be observed by the convergence of all the series in Figure 3.1 and in Table 3.2. In the precrisis period, the correlation between WTI and Coal is 0.64; however, after the crisis, this correlation falls to 0.08. The opposite happens with the correlation between Coal and Gas where this value increases from 0.29 to 0.82, while the correlation between Gas and WTI changes from 0.07 to 0.15. These cross-correlation changes indicate a high interrelationship among the three fossil fuel series; however, the long-term dynamic linkages are better captured by the lead-lag and Granger causality analysis included in Table 3.3.

**TABLE 3.2 Correlation matrix of log prices. The table does not include the diagonal values.**

	2006-12		Precrisis		Crisis		Posterisis	
	Coal	WTI	Coal	WTI	Coal	WTI	Coal	WTI
WTI	0.64		0.54		0.67		0.08	
Gas	0.16	-0.03	0.28	0.06	0.85	0.65	0.82	-0.15

**TABLE 3.3** Significance level of Granger causality ( $\dagger p \leq 0.05$ ,  $\ddagger p \leq 0.01$ ) and Brownian distance correlation (\*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ ) of log return series. Nonrelevant relationships are excluded. Yellow indicates nonlinearity according to either the White or Terasvirta test, and green means that both tests detect nonlinearity with a 5% significance level.

Periods	Lags/Effects	1	2	3	4	5	6	7
2006–12	WTI $\rightarrow$ Coal	** $\ddagger$	** $\ddagger$	** $\ddagger$	** $\dagger$	** $\ddagger$	* $\ddagger$	* $\ddagger$
	Gas $\rightarrow$ Coal	*						*
	Coal $\rightarrow$ WTI	**	**	**	**	**	**	*
	Coal $\rightarrow$ Gas		**	*		$\ddagger$	$\ddagger$	$\dagger$
	WTI $\rightarrow$ Gas	**						
Precrisis	WTI $\rightarrow$ Coal	**						
	Gas $\rightarrow$ Coal	*	*					
	Gas $\rightarrow$ WTI	* $\ddagger$	$\ddagger$	$\ddagger$	$\dagger$			
Crisis	WTI $\rightarrow$ Coal	** $\ddagger$	* $\ddagger$	* $\ddagger$	* $\ddagger$	* $\ddagger$	* $\ddagger$	$\ddagger$
	Gas $\rightarrow$ Coal							*
	Coal $\rightarrow$ WTI	**	*	**		**	**	**
	Coal $\rightarrow$ Gas		*	*		$\dagger$	$\dagger$	$\dagger$
Recovery	WTI $\rightarrow$ Gas	*						
	Gas $\rightarrow$ Coal							
	Coal $\rightarrow$ Gas	$\ddagger$	* $\ddagger$	$\ddagger$	$\dagger$	$\dagger$		

As none of the log price pairs are cointegrated in the different periods at the 5% significance level according to the Johansen test, we used a vector autoregressive (VAR) model of the log return series to run the Granger causality test with seven lags instead of using the vector autoregressive error correction model.

## 3.5 Discussion

During the complete period 2006–2012, WTI and Coal show a feedback relationship according to the Brownian distance, and only the WTI  $\rightarrow$  Coal relationship is maintained conforming to the Granger causality test with a 5% significance level (see Table 3.3). Coal also Granger causes Gas for the lags 5–7. In addition, the Brownian distance recognizes the following dependences (lags between parentheses): Gas (1, 7)  $\rightarrow$  Coal, Coal (2, 3)  $\rightarrow$  Gas, and WTI (1)  $\rightarrow$  Gas. Very similar relationships are observed during the crisis period (2008–2010). Both tests indicate that the Gas  $\rightarrow$

WTI dependence is relevant during the precrisis period, and the Brownian distance recognizes the importance of the relationship WTI (1) → Coal and Gas (1,2)→ Coal. During the recovery period, only the Coal → Gas relationship is relevant for both tests, especially for the Granger causality tests. Most of the additional relationships observed using the Brownian distance test, which were not recognized by the Granger causality test, were confirmed to be relevant nonlinear relationships according to the White and Terasvirta tests (see Table 3.3). Hence, the Brownian distance correlation recognizes an important number of dependences, and some of them are confirmed by the Granger causality test.

These identified nonlinear relationships may have an impact on the selection of inputs used to generate electricity in the United States. Electricity generated with coal reached its peak in 2007 and substantially decreased afterward. On the contrary, electricity generated with natural gas has been increasing since 1990, especially since 2009. Between the years 2000 and 2012, the proportions of electricity generated by coal and oil have decreased from 51.7% and 2.9% to 37.4% and 0.6%, respectively, while the proportion of electricity generated by natural gas almost doubled from 15.8% to 30.4% (see Table 3.4). The increase of the electricity generated

**TABLE 3.4** Net electricity generation: Participation of energy source. Ren. en. refers to renewable energy.

Year	Coal	Oil	Nat. Gas	Nuclear	Ren. En.	Other
2000	51.72%	2.92%	15.81%	19.83%	9.38%	0.35%
2001	50.96%	3.34%	17.10%	20.57%	7.70%	0.32%
2002	50.10%	2.45%	17.91%	20.22%	8.90%	0.42%
2003	50.83%	3.07%	16.74%	19.67%	9.15%	0.55%
2004	49.82%	3.05%	17.88%	19.86%	8.85%	0.53%
2005	49.64%	3.01%	18.77%	19.28%	8.82%	0.48%
2006	48.97%	1.58%	20.09%	19.37%	9.49%	0.51%
2007	48.51%	1.58%	21.57%	19.40%	8.49%	0.45%
2008	48.21%	1.12%	21.44%	19.57%	9.25%	0.42%
2009	44.45%	0.98%	23.31%	20.22%	10.57%	0.45%
2010	44.78%	0.90%	23.94%	19.56%	10.36%	0.45%
2011	42.27%	0.74%	24.72%	19.27%	12.52%	0.48%
2012	37.42%	0.56%	30.35%	18.97%	12.22%	0.47%

Source: US Energy Information Administration.

with natural gas is equivalent to the contraction of electricity generated with coal. This can be partially explained by the decline of natural gas log prices since December 2005 to April 2012 (see Figure 3.1). The US Clean Air Act restrictions on  $SO_2$  emissions and the relative reduction of natural gas prices led the power plants to partially substitute coal with natural gas as their main input. As more power plants have increased their consumption of natural gas, its price has also increased following similar trends of oil and coal. The linear and nonlinear lead–lag analysis also indicates that coal’s price has an important effect on natural gas price, especially during the crisis and recovery period. This particular case illustrates the nonlinear dynamic among the prices of the different commodities studied and the major interrelationship that exists among the three fossil fuel series. The main application of these nonlinear relationships is to improve the forecast of commodity prices.

### **3.6 Conclusions**

---

This chapter proposes the use of the Brownian distance correlation to conduct a lead–lag analysis of financial and economic time series. When this methodology is applied to asset prices, the nonlinear relationships identified may improve the price discovery process of these assets.

The Brownian distance correlation determines relationships similar to those identified by the linear Granger causality test, and it also uncovers additional nonlinear relationships among the log prices of oil, coal, and natural gas. This research can be extended to explore the lead–lag relationship between spot and future prices of complex assets such as commodities and foreign currencies applied to different markets.

### **Acknowledgments**

---

Authors thank participants of the 2014 Eastern Economics Association meeting, the AAAI 2014 Fall Symposium on Energy Market Predictions, Dr. Ionut Florescu, and two anonymous referees for their comments and suggestions. G.C. also thanks the Howe School Alliance for Technology Management for financial support provided to conduct this research. The opinions presented are the exclusive responsibility of the authors.

## References

- Amavilah, V. 1995. The capitalist world aggregate supply and demand model for natural uranium. *Energy Economics* **17**(3) 211–220.
- Aruga, K., S. Managi. 2011. Linkage among the U.S. energy futures markets. Paper prepared for the 34th IAEE International Conference Institutions, Efficiency and Evolving Energy Technologies, Stockholm.
- Asche, F., O. Gjølberg, T. Volker. 2003. Price relationships in the petroleum market: an analysis of crude oil and refined product prices. *Energy Economics* **25** 289–301.
- Asche, F., P. Osmundsen, M. Sandsmark. 2006. The UK market for natural gas, oil, and electricity: are the prices decouples? *Energy Journal* **27** 27–40.
- Bachmeier, L., J. M. Griffin. 2006. Testing for market integration: crude oil, coal, and natural gas. *Energy Journal* **27** 55–71.
- Bai, J., P. Perron. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* **66**(1) 47–78.
- Brown, S. P. A., M. K. Yucel. 2008. What drives natural gas prices? *Energy Journal* **29** 45–60.
- Creamer, G. G., Y. Ren, J. V. Nickerson. 2013. Impact of dynamic corporate news networks on asset return and volatility. *Social Computing/IEEE International Conference on Privacy, Security, Risk and Trust, 2010 IEEE International Conference on* September 8–14, 2013, pp. 809–814.
- Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**(3) 424–438.
- Granger, C. W. J. 1980. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control* **2** 329–352.
- Granger, C. W. J. 2001. *Essays in Econometrics: The Collected Papers of Clive W. J. Granger*. Cambridge University Press, Cambridge.
- Grothe, O., J. Schnieders, J. Segers. 2014. Measuring association and dependence between random vectors. *Journal of Multivariate Analysis* **123** 96–110.
- Hartley, P., K. B. Medlock III, J. E. Rosthal. 2008. The relationship of natural gas to oil prices. *Energy Journal* **29**(3) 47–65.

- Johansen, S. 1988a. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* **59**(6) 1551–1580.
- Johansen, S. 1988b. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**(2/3) 231–254.
- Lee, T.-H., H. White, C. W. J. Granger. 1993. Testing for neglected non-linearity in time series models. *Journal of Econometrics* **56** 269–290.
- Mjelde, J., D. Bessler. 2009. Market integration among electricity markets and their major fuel source markets. *Energy Economics* **31** 482–491.
- Mohammadi, H. 2009. Electricity prices and fuel costs: Long-run relations and short-run dynamics. *Energy Economics* **31** 503–509.
- Mohammadi, H. 2011. Long-run relations and short-run dynamics among coal, natural gas and oil prices. *Applied Economics* **43** 129–137.
- Puliga, M., G. Caldarelli, S. Battiston. 2014. Credit default swaps networks and systemic risk. *Scientific Reports* **4** 1–7.
- Ramberg, D. J. 2010. The relationship between crude oil and natural gas spot prices and its stability over time. Master of Science Thesis, Massachusetts Institute of Technology.
- Székely, G. J., M. L. Rizzo. 2009. Brownian distance covariance. *The Annals of Applied Statistics* **3**(4) 1236–1265.
- Terasvirta, T., C.-F. Lin, C W. J. Granger. 1993. Power of the neural network linearity test. *Journal of Time Series Analysis* **14**(2) 209–220.
- Zhang, X., B. Podobnik, D. Y. Kenett, H. E. Stanley. 2014. Systemic risk and causality dynamics of the world international shipping market. *Physica A* **415** 43–53.

## Chapter Four

# Portfolio Optimization: Applications in Quantum Computing

Michael Marzec

*Stevens Institute of Technology, Hoboken, NJ, USA*

### 4.1 Introduction

---

The traditional Markowitz mean-variance model of portfolio selection has provided the framework for portfolio asset selection for decades (Markowitz, 1952, Fabozzi et al., 2013). In this framework, an investor or a portfolio manager wishes to minimize the so-called *risk* with a particular asset mix. The measurement of risk in a portfolio, for this model, is rooted in the association of risk with the measure of variance, which is a measure of the degree of change or variability of the data compared with its expected value—its mean. With multiple assets, the expression of variance of the portfolio includes the covariance—how much the assets vary together. The mathematical expression is quadratic and the optimal selection of assets becomes a quadratic optimization problem. This is formulated as a nonlinear programming problem, which can be solved by suitable application of various operations research techniques (Hillier and Lieberman, 2010).

The efficient frontier is a curve, an area, or a surface that traces out the risk versus return values resulting from various combinations of assets

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

(Elton et al., 2007). As the number of asset combinations increases—in both the number of assets and their weightings in the portfolio—the selection of the optimal portfolio quickly becomes difficult. The optimization techniques traditionally used to solve this problem are so-called *classical* optimization techniques that rely on mathematically well-defined gradient-based or descent/directional indications that must be well defined and constrained (Gilli and Schumann, 2012).

More modern heuristic optimization techniques, such as stochastic local search, simulated annealing, threshold accepting, tabu search, genetic algorithms, particle swarm, and ant colony optimization provide a fast alternative to these classical optimization techniques (Gilli et al., 2011). These heuristic techniques provide ways to approach problems that are too hard to solve classically. Many classical and heuristic techniques will typically yield solutions that involve some combination of all of the assets under consideration. How can we make an appropriate selection of a subset of assets? In other words, in what way can we make a decision to select  $m$  assets out of a universe of  $n$ ?

With suitable crafting, this can be approached as a quadratic unconstrained binary optimization (QUBO) problem. This crafting is core to, and discussed further in, this chapter. The problem is quadratic because of mixed terms in the signature equation:  $V = X^T Q X$ . The final model is unconstrained because it is built without additional restricting equations. It is binary in nature because it is written as a binary decision problem (yes/no; include/exclude) of whether to include a particular stock in the final portfolio. The problem is one of optimization because the goal is to seek the best—optimal—combination of stocks.

Specifically, this is a *combinatorial* optimization problem because the formulation of the objective will seek a discrete set of assets represented in graph form: this problem can be investigated using a graph-theoretic approach (Boros et al., 2008; Jallo and Budai, 2010, Papadimitriou and Steiglitz, 1998).

Iterating through all of the combinations of assets and weightings can be computationally intensive. Heuristics provide some improvement over classical techniques but could either run too long, yield suboptimal results, or not return a solution at all. One paradigm of solving this type of problem is quantum computing (Choi, 2010; D-Wave, 2013b, 2013c). This paradigm holds promise toward many challenging computational problems that are either difficult to solve or not possible with current techniques.

The problem of making a binary (include/exclude) selection of assets for a portfolio can be solved in a rudimentary way using the maximum independent set (MIS) graph-theoretic approach to solving the QUBO in a quantum computing paradigm (D-Wave, 2013b). Can this technique be applied or extended to solve a weighted binary asset selection problem in a Markowitz mean-variance framework?

Overall, this investigation will help bridge the quantum computing paradigm with financial engineering research topics. It will provide insight into the problem domain of the quantum computer linked with associated financial engineering concepts. The fundamental goal of this chapter is to demonstrate how the framework can be used to solve financial problems. It will discuss the limitations of the environment with respect to financial modeling considerations. This will be accomplished by presenting the formulation of financial portfolio optimization in the context of this hardware paradigm. This chapter presents and discusses existing sample work (D-Wave, 2013b).

The research question covers three areas that are reflected in the literature: classical mean-variance portfolio theory; general operations research theory with specific consideration given to combinatorial optimization topics; and the hardware realization of adiabatic quantum computation.

The remainder of the chapter delves into a survey of background literature, the models used, experimental methodology, results obtained, discussion, and conclusion. The background literature section presents a variety of research and applied papers on the underlying topics. The model section relates the portfolio optimization representation to the graph-theoretic domain and into the underlying Ising problem domain of the target hardware. The methodology section presents the information relevant to the implementation undertaken in this study. The results section presents what came out of the investigation. The discussion section relates the results to the underlying domain, along with limitations and future areas for investigation. The conclusion finalizes the chapter.

---

## 4.2 Background

---

Considering the hybrid nature of topics in this chapter, the available literature is deep and varied, crossing several disciplines. The following is a survey of material that weaves a path through the background subjects of classical mean-variance portfolio theory, general operations research

theory, combinatorial optimization topics, and the hardware realization of adiabatic quantum computation.

### 4.2.1 PORTFOLIOS AND OPTIMIZATION

As related at the beginning of the “Introduction” section, the Markowitz paper on mean-variance portfolio theory historically laid the foundation for variance–covariance portfolio modeling over the last several decades (Markowitz, 1952, Fabozzi et al., 2013). Meanwhile, portfolio optimization has evolved from that work with different objective representations or constraints. For example, more recent models consider a conditional value-at-risk (VaR) objective with constraints (Krokhmal et al., 2001).

There is a lot of literature in the broad areas of portfolio theory, portfolio optimization, and the general topic of optimization theory and techniques. In particular, in the optimization arena, a recent review illustrates a lot of research activity over the 1998–2008 timeframe (Floudas and Gouraris, 2009). Similarly, finance-oriented discussions are presented in the study by Gilli et al., 2011 with specific consideration of heuristic models in finance with Gilli and Schumann, 2012.

Other (non–mean-variance) methods include heuristic optimization with differential evolution under consideration of risk preferences and loss aversion (Maringer, 2006). Risk parity, equal weighting, and minimum variance are discussed in the study by Chaves et al., 2010.

Some very particular formulations of the mean-variance problem are solved with modern and classical optimization techniques: using Lagrangian relaxation (Shaw et al., 2008); a hybrid Grey Relational Analysis approach (Huang et al., 2011); and a genetic algorithm formulation (Soleimani et al., 2009), as just a few examples.

Underlying all of these approaches is the desire to find improved methods to solve the optimization problem. Many of the problems are difficult to solve because their model exhibits several local optima or have discontinuities in the function expression (Gilli and Schumann, 2012). Reflecting back on the base mean-variance formulation of portfolio optimization is a problem that is quadratic in nature with squared terms of variance and mixed covariance terms that can exhibit some of these characteristics.

Finding the best overall solution out of these multiple local optima is the goal of global optimization. A sampling of current research papers specific to the global optimization of this quadratic formulation includes

equilibrium search techniques (Pardalos et al., 2008), barrier function formulations (Dang and Xu, 2000), an interior-point algorithm approach (Akrotirianakis and Rustem, 2005), and an unconstrained max-flow approach (Boros et al., 2008).

Some of the research considers constraints in the formulation whereas some are written in unconstrained representation. Constrained problems may sometimes be rewritten as unconstrained formulations using various techniques (Hillier and Lieberman, 2010; Gilli et al., 2011). The unconstrained representations are used in this investigation.

A link between MISs, cliques, and stock market data is discussed in the study by Boginski et al. (2005) where the cross-correlations between stocks are studied over time. The thesis by Jallo and Budai (2010) further elaborates on market graphs related to pure stock returns, liquidity-weighted returns, and volume measures.

The article by Charpin and Lacaze (2007) presents a binary optimization on portfolios that provides a method to constrain a portfolio to a specific size, that is, cardinality, and also that determines an optimal portfolio satisfying minimum weight conditions. They solved the problem with branch-and-bound optimization in the Lingo or CPLEX commercial application environments. The binary nature of the problem is similar to what is considered in this chapter, but the model has a different form and this chapter's model formulation is quickly translated to graph-theoretic representation. The paper by Bertsimas et al. (1999) uses mixed-integer programming to perform a constrained optimization using classical models that have binary selection characteristics.

#### 4.2.2 ALGORITHMIC COMPLEXITY

The topic of algorithmic complexity is an important motivation to, and marker of, the investigation of new techniques to solving these problems. Since the core of this chapter is the practical implementation of a new approach to solving hard problems, algorithmic complexity deserves at least a cursory overview. Algorithms, or sequences of steps in a well-defined computational procedure, are used to specify the way of solving various problems. Algorithms are used at some point in the description, implementation, or analysis of all of the methods cited in this chapter. To pick one algorithm over another, a measure of efficiency is used to compare their performance. This is typically the measurement of how quickly a particular algorithm finds an answer given a certain input size.

A few good references for this topic include Cormen et al. (2009), Kozen (1992), Papadimitriou and Steiglitz (1998).

Algorithms may be put into three classes: P, NP, and NP-complete. Problems that can be solved in polynomial time are in class P. Polynomial time solutions are found in a time proportional to the size of the problem input raised to some constant power. Problems that can only be verified, but not solved, in polynomial time are called nondeterministic polynomial, class NP. Finally, problems that do not have an algorithmic way of finding an exact solution are in the class NP-complete and are considered the hardest. Typically, algorithms applied to NP-complete problems find good enough solutions or approximations rather than the best or optimal solution. There is another class called NP-hard, which for the purpose of this chapter, means problems that are at least as hard as the hardest problems in NP. However, NP-hard problems do not need to be verifiable in polynomial time and, thus, do not actually need to be members of class NP.

This is important because some of the models considered in the chapter are in the classes NP-complete or NP-hard and the D-Wave system can be applied to these problems.

### 4.2.3 PERFORMANCE

As mentioned in Section 4.2.2, the performance categorization of algorithms allows us to measure usefulness, among other things. One of the reasons for investigating the D-Wave system is that it holds promise to help solve problems that are too complex, too slow or do not yield solutions under classical methods. In other words, it promises to yield performance improvements over the current paradigm.

Therefore, it would be useful to know how the D-Wave system compares to current, classical systems used to solve particular problems. The D-Wave system has been experimentally compared with three conventional software solvers IBM ILOG CPLEX, METSlib TABU search, and akmaxsat. Three problems were investigated from the NP-hard problem domain: QUBO, weighted maximum 2-satisfiability (W2SAT), and quadratic assignment. Various problem sizes and implementation situations are used in the comparison. Problem solution quality and success are also compared, in addition to timing comparisons. The results do indicate situations in which the D-Wave equipment may not be ideal or yield improvement. However, the results show that the hardware implementation

can be several thousand times faster, up to even 10,000 times faster, than current implementations (McGeoch and Wang, 2013).

#### 4.2.4 ISING MODEL

The Ising model was originally constructed to help understand the behavior of magnetic materials. Two main terms in the model represent the collection of individual field strengths at each molecule and the collection of interactions among neighboring molecules. The Ising representation discussed in Boixo (Boixo et al., 2012) and, in the form used here, in Choi (2008) is the energy function

$$E = \sum_{i \in V(G)} h_i s_i + \sum_{ij \in E(G)} J_{ij} s_i s_j$$

where the  $s_i \in \{-1, +1\}$  represents the spins of the molecules in a system that is in an applied magnetic field,  $h_i$  is the strength of a magnetic field at molecule  $i$ , and  $J_{ij}$  represents the strength of interaction between the neighboring molecule's spins  $i$  and  $j$ .

It was later realized that this simple model could be applied to many other situations in which consideration is given to a collection of individual properties and their interactions. As stated in the study by Bian and colleagues (Bian et al., 2010), there were apparently more than 12,000 papers published between 1969 and 1997 that used the Ising representation. Presumably, that number may be much higher now.

This Ising model shares a similar structure to the QUBO problem, which can be used to represent the portfolio optimization problem. Moreover, the D-Wave equipment is a hardware implementation of an optimization engine designed to solve the Ising problem, to which the QUBO form can be translated (D-Wave, 2013b, 2013c; Bian et al., 2010).

#### 4.2.5 ADIABATIC QUANTUM COMPUTING

Adiabatic quantum computing is a new computing paradigm that has potential to yield satisfactory results when used to solve some of the NP-complete or NP-hard problems. The quantum computing aspect speaks to the use of special purpose-built analog environments built to take advantage of quantum physics properties of the core materials. This is in contrast to our current digital computing paradigm. Adiabatic, in this sense, means that the underlying material's quantum nature is kept in its ground state as

the system evolves toward the solution. This contrasts with the classical cooling usage of the term that is a temperature-based annealing process (Farhi et al., 2000).

However, one of the heuristic methods that can be applied to solve difficult optimization problems is simulated annealing, which is based on that classical physical cooling process mentioned in the previous paragraph.

In both cases, the general idea is to iteratively evolve a problem by the introduction of a random component, through either temperature or quantum processes, to help improve the solution. In this sense, quantum annealing is conceptually similar and uses entropy of the quantum process to explore the objective. In the environment discussed in this chapter, the quantum annealing is realized by the D-Wave hardware system (Farhi et al., 2000; Bian et al., 2010; Hillier and Lieberman, 2010; Gilli et al., 2011).

The fundamental quantum nature of the D-Wave system is a hot topic of debate. Recent investigations seem to indicate that quantum signatures are present in the system (Boixo et al., 2012). No matter how this debate resolves, if the system can ultimately provide good cost-effective solutions to difficult problems, then it becomes useful.

In summary, the background literature presents a variety of models and solution methods that are beneficial to obtaining a portfolio mix that can be used to represent a range of financial objectives. In particular, the QUBO problem, mentioned earlier (Boros et al., 2008), is characteristic to portfolio optimization, the Ising problem in physics (Palubeckis, 2004, Chen and Zhang 2010, Boixo et al., 2012), and other discrete mathematics problems such as combinatorial optimization graph-theoretic representations (Choi, 2010). These topics combine to provide a unique portfolio problem formulation that will be investigated in this chapter. The model section discusses the specific representations used in this investigation.

---

### 4.3 The models

---

The core models are financial (portfolio), graph-theoretic combinatorial optimization, Ising/QUBO, and the resulting mixed model. Each is reviewed and discussed with consideration of how they are used for the current investigation.

### 4.3.1 FINANCIAL MODEL

The primary model is based on the Markowitz mean-variance paradigm. As such, the basic representation of portfolio risk as measured by the variance and covariance between stocks is considered.

The formula expression for the variance of a portfolio,  $\sigma_p^2$ , is

$$\sigma_p^2 = V = \sum_{j=1}^N (X_j^2 \sigma_j^2) + \sum_{j=1}^N \sum_{\substack{k=1 \\ k \neq j}}^N (X_j X_k \sigma_{jk})$$

Here the first term,  $\sum_{j=1}^N (X_j^2 \sigma_j^2)$ , is the sum of the *variances* on the individual assets,  $\sigma_j^2$ , multiplied by the square of the proportion invested in each,  $X_j^2$ . The second term,  $\sum_{j=1}^N \sum_{\substack{k=1 \\ k \neq j}}^N (X_j X_k \sigma_{jk})$ , is the effect of the *covariance* between the individual assets in the portfolio.

The matrix representation of the variance,  $V$ , equation is

$$V = X^T Q X$$

where  $X^T$  is the transpose vector of the proportional investment for each stock in the portfolio and  $Q$  is the variance–covariance matrix.

Traditionally, we may look to a simplified composite model to represent the portfolio risk and return, which can then be optimized. This would fundamentally look like

$$F = -R + V$$

where  $R$  represents the portfolio return and  $V$  represents the portfolio risk.<sup>1</sup> In this case, the optimization is a minimization: we wish to find

$$\min \{F\}$$

This optimization therefore minimizes the risk while maximizing the return. However, this optimization is usually influenced by constraints. This is further discussed, in depth, in the very tractable (Bartholomew-Biggs, 2005), which develops a composite function form with constraints, discusses the conversion to an unconstrained problem formulation, and builds up iterations on the problem formulation, finally considering global unconstrained optimization. Putting aside this model framework for a while, the transition to the quantum computing environment needs to be considered.

---

<sup>1</sup>An alternative composite function, not investigated in this chapter, is the risk–return ratio  $F = \frac{V}{R}$ .

### 4.3.2 GRAPH-THEORETIC COMBINATORIAL OPTIMIZATION MODELS

Graph-theoretic structures underlie the combinatorial optimization problem formulation. The combinatorial optimizations used are the MIS and weighted MIS (WMIS) representations. It is the MIS and WMIS representations that provide the bridge between the basic portfolio optimization problems and their formulations that may be solvable in the quantum computing environment.

The consideration, so far, has been a quadratic problem form but not a quadratic *binary* problem. However, papers by Choi and Bian provide the linking insight between the Ising/QUBO form and the MIS or WMIS representation that can be used to express problems that may be solvable in the quantum computing environment (Choi, 2008, 2010, Bian et al., 2010). As mentioned in these references, optimization and decision problems are related. The decision problem (looking for yes/no answer or confirmation to a question) for MIS is NP-complete and the optimization version (finding the minimum) of MIS is NP-hard. Finding a solution to such problems can be quite difficult, that is, take a lot of time. The quantum computing environment solves these sorts of difficult problems.

The initial adapted model used is the MIS. A graph,  $G(V,E)$ , of stocks is created with the stock symbol set as the vertices,  $V$ , and the edge connection,  $E$ , is determined by a selection criteria. The first selection criterion is risk represented as the *correlation* between stocks. Therefore, an edge is drawn between two vertices if the corresponding pair of stocks is correlated above some threshold.

The use of correlation here is different from the variance–covariance model and warrants a quick review. This selection criteria allows the optimization process to select what is (loosely) termed a *diversified* portfolio; here, the diversification metric is *correlation* whereas other metrics may be chosen, such as the percentage of risk born in the portfolio attributable to market movements, with the Dow-Jones representing the market, for instance. Such portfolio performance techniques and metrics are discussed in the study by Elton and colleagues (Elton et al. 2007). The Charpin and Lacaze article discusses one such representation in which they minimize  $V(R_p - R_b)$ , which is the variation of the differential return of the portfolio,  $R_p$ , with that of a benchmark,  $R_b$ , for example, something like the Dow-Jones index, etc. (Charpin and Lacaze, 2007). As such, other selection criteria or formulations could be chosen and would be suitable as future

investigations. It is important to note that with correlation as a selection parameter the edge selection is made when stocks are correlated above a certain threshold. With variation as a selection parameter, drawing edges with the selection parameter is not too intuitive: using the variation as the *vertex weight* and then optimizing for the *minimum* weight, that is, the negative of the maximum weight optimization function, would yield a set of stocks that have low variation.

The enhanced model considers the *weighted MIS*, WMIS. Here, as in the initial adapted model, a graph,  $G(V,E)$ , of stocks is created with the stock symbol set as the vertices,  $V$ , and the edge connection,  $E$ , is determined by a selection criteria. However, each vertex has an additional parameter: the weight. In this case, the optimization is to find the MIS that yields the maximum total weight (of vertices selected in the MIS). The weight, in this implementation, will represent the historical return of the stock, handicapped by its variance, represented by the vertex.

### 4.3.3 ISING AND QUBO MODELS

The Ising representation was briefly discussed in the earlier background section and is repeated here in the context of the influential models. The NP-hard Ising representation discussed in the studies (Boixo et al., 2012, Bian et al., 2010) and, in the form used here, in the study by Choi (2008) is the energy function

$$E = \sum_{i \in V(G)} h_i s_i + \sum_{ij \in E(G)} J_{ij} s_i s_j$$

In this model,  $s_i \in \{-1, +1\}$  represents the spins of molecules within a system. In this system, the strength of the applied magnetic field at molecule  $i$  is  $h_i$ . The strength of the interaction between the spins of neighboring molecules  $i$  and  $j$  is represented by the matrix term  $J_{ij}$ .

The QUBO form that represents the WMIS, discussed in the study by Choi (2010), is

$$Y(x_1, \dots, x_n) = \sum_{i \in V(G)} c_i x_i + \sum_{ij \in E(G)} J_{ij} x_i x_j$$

where  $c_i$  is the weight applied at vertex  $i$  and the  $J_{ij}$  is further discussed in the context of the translation, in Section 4.3.4.

### 4.3.4 MIXED MODELS

For the implementation considered here, the problem must be translated, that is, mapped, from WMIS into Ising form so that it can be implemented on the D-Wave system. The translation follows (D-Wave, 2013b), which is further elaborated in the study by Choi (2010).

The mapping generates a matrix  $J$ , where  $J_{ij}$  entries are essentially determined by constant\*<sup>\*</sup>marketGraph; that is, the market graph is multiplied by a constant. The constant is the parameter  $J = 1.1$ , that is anything greater than 1, discussed in the references: the condition on  $J_{ij}$  is that it satisfies  $J_{ij} > \min\{c_i, c_j\}$  and if the weights  $(c_i, c_j)$  are all equal-weight with value 1, then  $J = 1.1 > \min\{c_i, c_j\}$ . The market graph has entries with value 1, indicating the existence of an edge between the corresponding vertices, otherwise a value of 0. The  $h_i$  Ising vectors follow the mapping  $h_i = Jd_i - 2c_i$ , where the parameter  $J = 1.1$ ,  $d_i$  is a count of the edges connected to the  $i^{\text{th}}$  vertex and  $c_i$  is the weighting considered for that vertex.

The simplified model—the MIS model—is essentially the WMIS, with the  $c_i$  (node) weights set to 1, that is, equally weighted.

The WMIS model uses a combination of the stock's return and variance as the weights for each node in the market graph, that is, for each of the  $c_i$  in the  $h_i$  Ising vectors.

---

## 4.4 Methods

---

This proof of concept follows a practitioner's approach with the following steps: recognizing the potential use for the quantum computing environment to solve financial-oriented investigations, discussed earlier; reviewing the general models of relevance to both paradigms—quadratic and combinatorial optimization; locking in on analytically demonstrated commonalities between them—choosing a model representation, discussed earlier; choosing an implementation environment—MATLAB, in this case; gathering real stock data for model input—a subset of Dow Jones stocks; coding the implementation to call the underlying hardware, after preprocessing the model to fit the needed underlying representation; reviewing the results; and, finally, considering future research topics, which are presented in the “Discussion” section.

#### 4.4.1 MODEL IMPLEMENTATION

The model representation, discussed earlier, is realized through the motivating implementation (D-Wave, 2013b), which yields a simplified diversification of a stock portfolio by generating connections in a market graph between two stocks whose correlation is above a certain threshold. Then an MIS of the graph is found by rewriting it as an Ising problem. This Ising representation is solved with the hardware optimization engine. The reference implementation was written in Python, whereas this chapter looks at a MATLAB rendition. Also, the implementation will consider variations in risk-measure threshold and a WMIS implementation with the weightings represented by each stock's historical return. It is fully recognized that there may be alternatives to the weighting assignment, such as by VaR of the stock within the portfolio or, ultimately, with the actual weight of the stock in the portfolio; see the "Discussion" section.

Again, the primary objective is to use a known model (context) but to express it in the language necessary to run it on the D-Wave hardware optimization engine. The setting of this investigation is both theoretically analytic and pragmatic: portfolio theory mixes with quantum computing theory, the practicalities of plausible implementation and the actual mechanics of programming the problem to run in the environment.

#### 4.4.2 INPUT DATA

For the implementation, the data were retrieved from Yahoo!Finance and stored in arrays and structures. The next step calculated the log return for each of the historical daily returns: measure the day-to-day return  $r_i$  of the stock prices ( $S_i$ ) between day  $i - 1$  to day  $i$  as the logarithmic return:

$$r_i = \ln \left( \frac{S_i}{S_{i-1}} \right)$$

This is calculated over the sample space of historical periodic stock values. The log return values for each series were calculated and stored in new arrays.

#### 4.4.3 MEAN-VARIANCE CALCULATIONS

Next the variance–covariance matrix was determined over the collection period. The correlation matrix is used in the representative implementation, so that was calculated as well. It is understood that the values of the

variance–covariance matrix will change depending on the sample interval and the number of samples. In this case, the values were determined over the whole sample period, with daily samples starting from the beginning of 2012.

#### 4.4.4 IMPLEMENTING THE RISK MEASURE

The risk measures used in this chapter were the correlation and covariance value between pairs of stocks. Any other suitable risk measure could be chosen as the acceptance criteria for building the market graph, as long as it corresponds to a measure between pairs (for this use of the graph). The correlation is used as an edge selection threshold: when the correlation between vertices  $ij > \text{threshold}$ , then an edge is drawn between the two vertices in the graph. This generates the market graph: stocks are connected when their correlation is above some applicable threshold, say 0.5 for instance. Stocks that do not have edges to other stocks are considered *independent* (within the threshold). The optimization results in the maximum number of independent stocks that includes the best (single) picking from each of the groups of correlated stocks along with all of the independent (nonconnected) stocks.

For each model, the risk threshold value is varied as  $\text{risk} \in \{0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65\}$ . The step value of 0.05 was chosen *ad hoc* and could be changed to suit the investigation parameters. The boundary values of the range were chosen on the basis of observations of the correlation distribution, so that the resulting market graph would neither be fully connected nor would it be too sparse or disconnected.

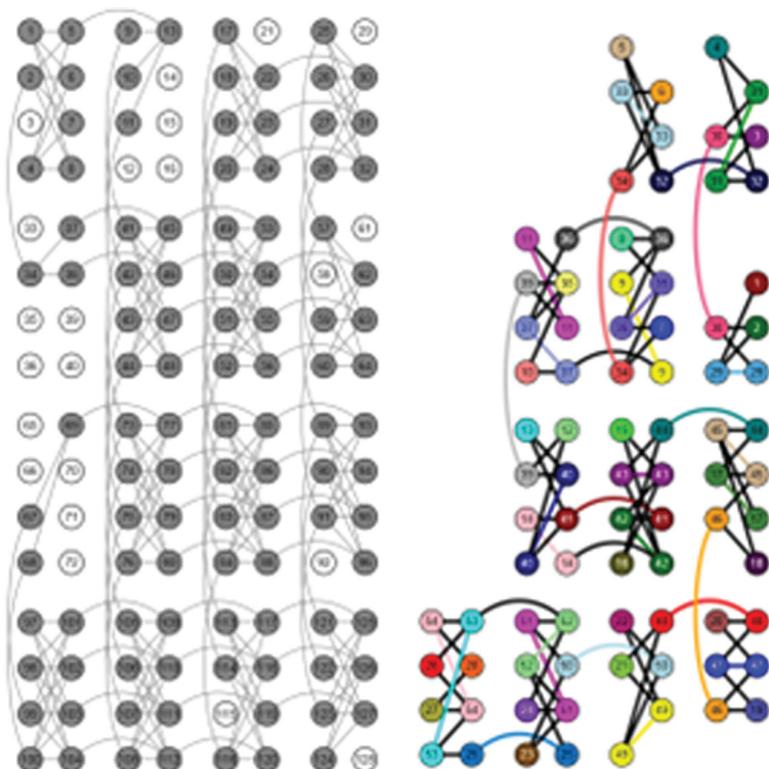
#### 4.4.5 IMPLEMENTATION MAPPING

To solve a problem on the D-Wave system, it must map into the structure of the hardware. The D-Wave system implementation, discussed in (Rose, 2008) and (D-Wave, 2013a), pragmatically cannot provide edges (links) between all of the pairs of qubits—see Figure 4.1 for the illustrative graph.

The fundamental hardware connectivity that has been implemented is described as a primal graph called *Chimera*. To solve a problem on the system, a problem must be expressed in a form that is connected by a graph that can be *embedded* into the hardware’s Chimera graph. Finding an embedding from non-Chimera into Chimera is in itself a hard problem.

If an embedding cannot be found, the problem cannot be directly solved in the system.

The hardware available at the time of the investigation was restricted to problems of 128 variables or less. Larger problems must be broken into subproblems. For this investigation, it was useful that D-Wave had already determined a subgraph suitable for use in problems with 17 variables or less. An embedding, labeled here as K17\_Q128\_EMBEDDING, that maps a 17-variable into 128-qubit, was already provided/solved by D-Wave. This embedding is adapted for the 10 stocks scenario used in this chapter—see the structure called K10\_Q128\_EMBEDDING in the Matlab code appendix, at the end of the chapter. The following graph (Figure 4.1), taken from the D-Wave materials, is illustrative of what the system's qubit connectivity might look like for a particular hardware implementation (in a gray scale) and a possible embedding graph fitting a particular problem (in color):



**FIGURE 4.1** Example embedding.

To implement a WMIS problem, the mapping  $h_i = Jd_i - 2c_i$  is used, as discussed earlier, in which the  $c_i$  term is the assigned weighting of the  $i^{\text{th}}$  vertex.

The optimization of the WMIS problem, as applied to the portfolio selection process and as formulated in this chapter, yields the maximum return within the minimum risk threshold for an equally weighted selection of stocks.

The Matlab code for the WMIS D-Wave application programming interface implementation is given in Appendix 4.A: WMIS Matlab Code. Administrative or utility scripts are not detailed here, such as implementing the actual down-load and pre-processing of data, for example, ordering.

---

## 4.5 Results

---

A few model variations were considered in the simulation runs. First, a simple model following the example presented in the D-Wave developer portal that uses the stock correlations as a risk measure indicating the level of market diversification. The next model used in the simulation was closer to the traditional mean-variance formulation using the variance-covariance matrix but was restricted by only considering the stock pair interactions with covariance; the node values remained equal weighted with unit value. The final version used covariance for the stock interactions, representing the market graph edge selection criteria with the addition of the market graph node weighting represented by the individual stock return handicapped by its variance.

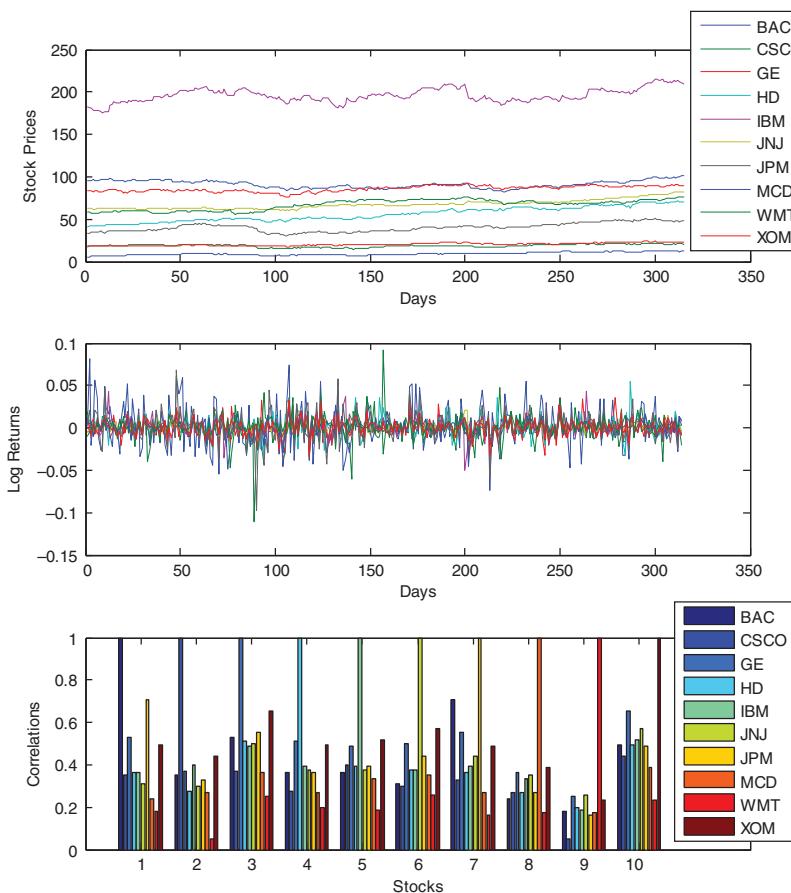
### 4.5.1 THE SIMPLE CORRELATION MODEL

A current run of the stock processing routines is summarized in the following set of graphs (Figure 4.2), with trading-day data from the beginning of 2012 to the most recent run of the script, which resulted in 314 samples.

The correlations bar graph indicates the pairwise correlation between each of the 10 stocks:  $10 \times 10 = 100$  correlations. The unit value correlations are the expected correlation values of the stock with itself. The result from the correlations and subsequent analysis or explanation is not the focus here: the correlation is used as representative of the risk measure. Since the representative set of stocks that were chosen from the Dow Jones index indicates little to moderate correlation, a correlation

## 4.5 Results

89



**FIGURE 4.2** Summary of input data.

threshold between 0.4 and 0.6 should be sufficient to generate a somewhat sparse market graph where the edge connectivity is determined by a fixed correlation threshold acceptance value, such as 0.5. The representative correlations are shown in Figure 4.3:

	Correlations between log-returns									
	BAC	CSCO	GE	HD	IBM	JNJ	JPM	MCD	WMT	XOM
BAC	1	0.3542	0.5287	0.3665	0.3618	0.3131	0.7055	0.2425	0.1826	0.4925
CSCO	0.3542	1	0.3671	0.2778	0.3971	0.3001	0.3307	0.2676	0.0527	0.4426
GE	0.5287	0.3671	1	0.5124	0.4884	0.5026	0.5545	0.3649	0.25	0.6561
HD	0.3665	0.2778	0.5124	1	0.396	0.3782	0.3652	0.2685	0.201	0.4916
IBM	0.3618	0.3971	0.4884	0.396	1	0.3779	0.3929	0.333	0.1835	0.5154
JNJ	0.3131	0.3001	0.5026	0.3782	0.3779	1	0.4414	0.3518	0.2598	0.5715
JPM	0.7055	0.3307	0.5545	0.3652	0.3929	0.4414	1	0.2667	0.1638	0.4856
MCD	0.2425	0.2676	0.3649	0.2685	0.333	0.3518	0.2667	1	0.1764	0.3896
WMT	0.1826	0.0527	0.25	0.201	0.1835	0.2598	0.1638	0.1764	1	0.2339
XOM	0.4925	0.4426	0.6561	0.4916	0.5154	0.5715	0.4856	0.3896	0.2339	1

**FIGURE 4.3** Log-return correlations.

The market graph connecting stocks greater than the initial risk threshold of 0.35 are given in Figure 4.4.

Market Graph for Minimum Risk Threshold 0.35										
	BAC	CSCO	GE	HD	IBM	JNJ	JPM	MCD	WMT	XOM
BAC		1	1	1	1	1	1			1
CSCO	1		1		1					1
GE	1	1		1	1	1	1	1		1
HD	1		1		1	1	1			1
IBM	1	1	1	1		1	1			1
JNJ			1	1	1		1	1		1
JPM	1		1	1	1	1				1
MCD				1		1				1
WMT					1					
XOM	1	1	1	1	1	1	1	1	1	

**FIGURE 4.4** Market graph for threshold 0.35.

This figure indicates that pairs of stocks are linked—connected by an edge when the pair's entry has value of 1—in the market graph when their respective risk measure is greater than 0.35.

The simulated resulting equally weighted portfolio from the MIS of this (0.35 threshold) market graph is BAC, IBM, JNJ, JPM, WMT, and XOM.

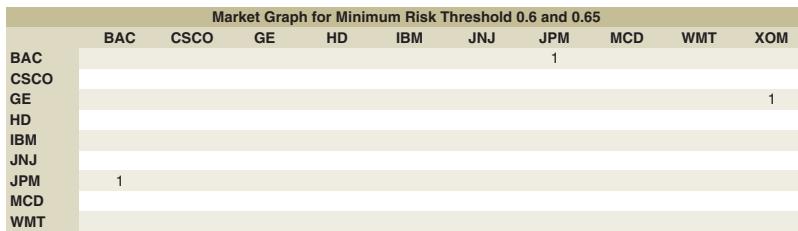
As indicated in the earlier bar graph depicting the correlations between stocks, as the threshold value is raised, the number of stocks matching that criterion is reduced (to a point where none of them are correlated at a high enough threshold). This means that the market graph typically becomes sparse as the threshold is increased. For example, here are the market graphs (Figure 4.5) for a few of the threshold levels:

Market Graph for Minimum Risk Threshold 0.4										
	BAC	CSCO	GE	HD	IBM	JNJ	JPM	MCD	WMT	XOM
BAC		1					1			1
CSCO			1							1
GE	1			1	1	1	1			1
HD				1						1
IBM				1						1
JNJ				1			1			1
JPM	1				1					1
MCD										
WMT										
XOM	1	1	1	1	1	1	1			

Market Graph for Minimum Risk Threshold 0.5										
	BAC	CSCO	GE	HD	IBM	JNJ	JPM	MCD	WMT	XOM
BAC		1					1			
CSCO			1							
GE	1			1		1	1			1
HD				1						
IBM										1
JNJ							1			1
JPM	1				1					
MCD										
WMT										
XOM										

**FIGURE 4.5** Market graphs for increasing thresholds.

**FIGURE 4.5 (Continued)**

Comparing the market graph representations, at the various thresholds, the sparseness does increase with increasing threshold. Conversely, threshold values less than a certain value result in fully connected market graphs.

With the range of risk  $\in \{0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65\}$ , it was found that the two levels  $\{0.6, 0.65\}$  resulted in the same market graph.

The final set of portfolios that correspond to each risk setting is shown in Figure 4.6:

Resulting Portfolios for Correlation-based MIS							
Minimum Risk Threshold:	0.35	0.4	0.45	0.5	0.55	0.6	0.65
Equal-weight Portfolio Per Threshold	BAC						
	CSCO						
				GE	GE	GE	GE
	HD						
	IBM						
	JNJ						
	JPM						
	MCD						
	WMT						
	XOM						

**FIGURE 4.6 Portfolios for correlation-based MIS model.**

### 4.5.2 THE RESTRICTED MINIMUM-RISK MODEL

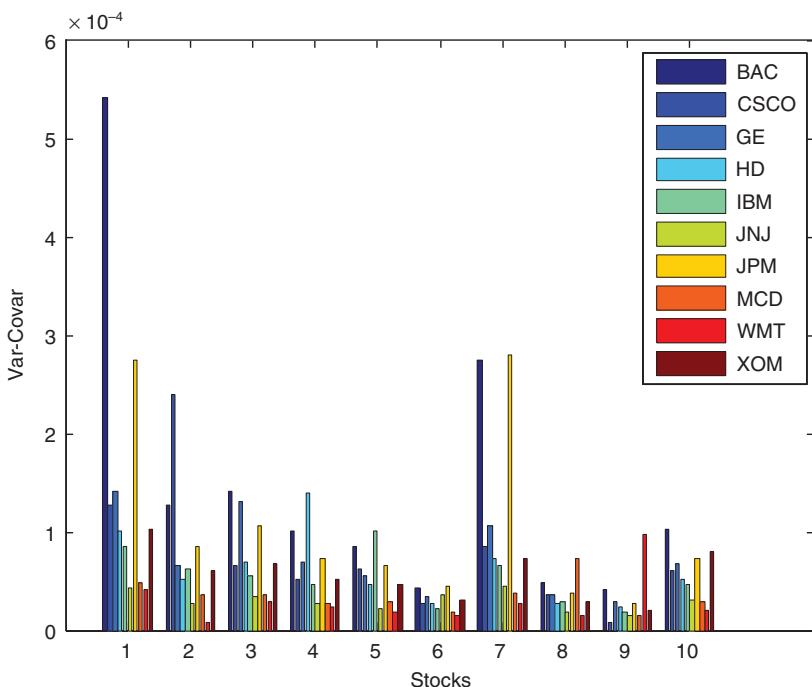
This model is closer to the mean-variance framework except that it has been restricted. The market graph edge selection criterion is the threshold range: Threshold = risk  $\in \{0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65\}$ . However, to scale it properly, the edge-selection condition checked is

$$\text{Covariance}(x_i, x_j) > \text{Threshold}^* \sigma_i^* \sigma_j$$

where the  $\sigma_i = \sqrt{\text{Variance}(x_i)}$ . The restricted part of this model is the use of unit-valued equal weightings on all of the nodes. This effectively

reduces a WMIS graph optimization to an MIS graph optimization. Also, the individual returns are not used. This could be considered a *minimum-risk portfolio*.

Using the same input data set, the bar graph showing the variance of each stock and the covariance between pairs is shown in Figure 4.7<sup>2</sup>:



**FIGURE 4.7** Variance-covariance of the input data.

The resulting market graphs for each of the thresholds were the same as in the correlation model. The final set of portfolios that correspond to each risk setting is shown in Figure 4.8:

<sup>2</sup>In the graph, the  $x$ -axis numbers correspond to the list of stock names, that is, BAC = 1, CSCO = 2, ...

Resulting Portfolios for Restricted Minimum-risk MIS							
Minimum Risk Threshold:	0.35	0.4	0.45	0.5	0.55	0.6	0.65
Equal-weight Portfolio Per Threshold	BAC						
	CSCO						
	HD						
	IBM						
	JNJ						
	JPM	JPM					
	MCD						
WMT	WMT	WMT	WMT	WMT	WMT	WMT	WMT
XOM							

**FIGURE 4.8** Portfolios for the restricted MIS model.

Comparing the restricted minimum-risk results with the earlier correlation-based model, it can be seen that the results were the same. This is expected, considering that the two models are related by the variance-covariance scaling, discussed earlier.

#### 4.5.2.1 Comparative classical minimum-risk result

Using **MATLAB** to run a “classical” *minimum variance* quadratic programming optimization (using the MATLAB function quadprog) on the data, with the restrictions that the weights sum to 1 and the weights are nonnegative results in the following comparative portfolio (Figure 4.9):

Minimum Variance - classical results:		
Stock	Portfolio Weight	Portfolio Basket
BAC	0.0000	
CSCO	0.0000	
GE	0.0000	
HD	0.0056	<b>HD</b>
IBM	0.0660	<b>IBM</b>
JNJ	0.5858	<b>JNJ</b>
JPM	0.0000	
MCD	0.1844	<b>MCD</b>
WMT	0.1582	<b>WMT</b>
XOM	0.0000	
Sum:		1.0000

**FIGURE 4.9** Classical results with minimum variance model.

The weights that are zero value in the minimum variance classical results were actually nonzero but much smaller than the significant digits shown here. Therefore, if very small weightings are allowed in the portfolio, then it might be possible to consider some combination of *all* of the stocks—this is usually not practical, especially once the cost of holding is built into the model.

The classical result is close to the MIS result for risk threshold 0.35, with the addition of BAC and XOM in the MIS results. Further investigation might reveal the subtleties in the difference; however, it is worth repeating that a key differentiator between the MIS method adopted here and a classical method is that the MIS weights are modeled as all equally weighted (unit value) whereas the classical method looks for the best choice of weights that result in the optimal solution.

### 4.5.3 THE WMIS MINIMUM-RISK, MAX RETURN MODEL

This model is even closer to the mean-variance framework. Again, it is somewhat restricted in the sense that it does not account for any mixed terms for the graph node weights. This could also be considered an adapted maximum return, minimum-risk model. Here, the market graph edge selection criterion is the threshold range, as used in the restricted minimum-risk model, with the same form. These results are representative of the full WMIS model.

The variance–covariance matrix and all of the resulting market graphs are the same as in the previous section. Here, the mapping  $h_i = Jd_i - 2c_i$  with  $c_i$  representing the node's weight,  $c_i = R_i - V_i$ . As such, the D-Wave Ising solutions are different for each of the risk thresholds, which results in a different set of optimal portfolios.

The final set of WMIS portfolios that correspond to each risk setting is shown in Figure 4.10:

Resulting Portfolios for MV WMIS with Returns							
Minimum Risk Threshold:	0.35	0.4	0.45	0.5	0.55	0.6	0.65
Equal-weight Portfolio Per Threshold	CSCO		BAC	BAC	BAC	BAC	
		CSCO	CSCO	CSCO	CSCO	CSCO	
	GE	GE		GE	GE	GE	
	HD		HD	HD	HD	HD	
			IBM	IBM	IBM	IBM	
			JNJ	JNJ	JNJ	JNJ	
	JPM						
MCD	MCD	MCD	MCD	MCD	MCD	MCD	
WMT	WMT	WMT	WMT	WMT	WMT	WMT	
XOM	XOM	XOM					

FIGURE 4.10 Portfolios under the WMIS model.

#### 4.5.3.1 Comparative classical minimum-risk, max return result

Using **MATLAB** to run a “classical” *maximum return–minimum variance* quadratic programming optimization (using the MATLAB function quadprog) on the data with the function  $F = -R + V$  (as discussed earlier), also with the restrictions that the weights sum to 1 and the weights are nonnegative results in the following comparative portfolio (see Figure 4.11):

Optimal Mean-Variance - classical results:		
Stock	Portfolio Weight	Portfolio Basket
BAC	0.7270	<b>BAC</b>
CSCO	0.0000	
GE	0.0000	
HD	0.2730	<b>HD</b>
IBM	0.0000	
JNJ	0.0000	
JPM	0.0000	
MCD	0.0000	
WMT	0.0000	
XOM	0.0000	
Sum:		1.0000

**FIGURE 4.11** Classical results with mean-variance model.

The previous classical result note regarding the weights also applies here. As also noted in the previous classical comparative results, the fact that the weights are allowed to vary continuously between 0 and 1 is a key differentiator to the WMIS results that have implicit (portfolio) equal-weighting.

## 4.6 Discussion

Simply put, the significant finding here is that a portfolio problem can be expressed, in a simple fashion, in a graph-theoretic paradigm, which may then be translated to the language required for implementation in a quantum computing environment. A binary decision on the inclusion/exclusion of particular stocks in the portfolio may be made resulting in a simplified optimal portfolio.

The interpretation of the model and results is quite specific, for the implementation considered here: the results pertain to a portfolio of stocks resulting from the optimization of the graph-theoretic MIS or WMIS. In this context, the results have meaning but might be challenged by current industry methods on the same set of data, especially in consideration of financial return.

In relation to previous research, this chapter has enhanced the D-Wave, 2013b and D-Wave, 2013c by investigating both in a single setting, while employing Matlab, and considering iterations over a range of values of the risk variable and incorporating the stock returns. It presented the quantum computation and financial-centric references together. As such, it provides a foundation upon which a financial engineering practitioner may further investigate the merger of these problem domains.

A couple of experimental observations are noteworthy. First, the stocks that become independent, that is, without edges connecting to/from them in the market graph due to the threshold selection, are always included in the final portfolio. Second, related to the first point, as the market graph grows sparse, the size of the portfolio tends to increase. Both of these characteristics are a result of the underlying *maximum independent set* combinatorial optimization.

Also, the minimization does not necessarily pick the optimal stock (as may be determined by other means, for example, the classical results, discussed above) from the remaining connected subgraphs when using equal weights on the nodes because the only noteworthy Ising parameter, in that case, is the value of  $h_i = Jd_i - 2$ , which is only dependent on the number of edges into that node. For  $h_i = Jd_i - 2c_i$  the node weighting,  $c_i$ , is *not* the weight of the stock in the portfolio, that is, the proportion of the total portfolio represented by each asset. The node weight is a characteristic of the fabricated market graph and is, in the model considered here, dependent on the stock's return and variance. In both of the MIS and WMIS models considered here, the optimization only indicates to include (yes) or not to (no) include the stock in our final portfolio, not the amount of each stock to include as a proportion of the overall pie—they are equally weighted, from that point of view. Determining how to encode the proportion into the formula is one of the areas for further investigation.

This point is a conjecture: ultimately, it appears that to get a result with different characteristics, a fundamentally different model would need to be considered. Most certainly, models and/or mappings that include the individual returns and portfolio weightings need to be investigated. Also,

a model that maps  $F$  more directly to the underlying Ising problem, that is, without the use of the MIS/WMIS representation, is worthy of further investigation.

#### 4.6.1 HARDWARE LIMITATIONS

Several limitations abound such as the following: the number of qubits that can be used is currently 128 (with the current D-Wave hardware available at the time of the initial investigation); need to present problems in Chimera representation, that is, reformulate non-Chimera into Chimera; break up problems larger than can be represented by Chimera into smaller problems; and finding an embedding of the problem into the hardware connectivity is a hard problem in itself. The exploration of a subgraph of Chimera for this chapter is not specifically undertaken, although it might be an interesting exercise for further investigation. Each of these is discussed in D-Wave, 2013a or some of the other D-Wave programmer's guides, plus also in the study of Ross (Rose, 2008). These limitations are workable but may fundamentally provide further complexity to the problem environment, such as: how would hundreds of stocks be considered, at once, with such constraints?

#### 4.6.2 MODEL LIMITATIONS

The WMIS model is not a direct rendition of the classical mean-variance model because the WMIS model, in the implementation used here, considers only equal (portfolio) weightings of stocks whereas the objective of  $\min \{F = -R + V\}$  is to find the best mix of weights that provide the optimal portfolio. To enable a full rendition of the mean-variance model requires an effort to map the non-Chimera problem into Chimera, which is beyond the scope of the current chapter, whereas the WMIS has already been demonstrated in this computational framework.

Also, the implementation of WMIS used here does not take into account the mixed terms contributing to the node weights, such as seen in the analytic expansion of  $\min \{F\}$ . A naïve mapping is used in this chapter that maps  $\min \{F = -R + V\}$  onto the WMIS, which values the market graph node weights as the composite values  $-R_i + V_i$ , with  $R_i$  representing the return of the  $i^{\text{th}}$  stock and  $V_i$  as its variance.

Thus, the node weight is the stock's return handicapped by its variance: a large variance reduces the return on the stock and thus lowers its chance

of being selected out of a connected set of stocks. This brings to light one of the characteristics of the MIS or WMIS model: the final optimization takes place over the remaining edge-connected set of nodes. The independent, nonconnected nodes will *always* be selected as part of the final set.

#### 4.6.3 IMPLEMENTATION LIMITATIONS

The D-Wave system is in constant use and under scheduling demand. Unfortunately, the implementation investigated in this chapter was only able to make use of the D-Wave simulator. Therefore, all results using the D-Wave simulator must be considered to be manifestations of the underlying software implementation. However, the mechanics and background material are still relevant and future implementations could simply replace the software simulator with a hardware connection.

As such, the results herein are not representative of any time reduction in finding a solution and no attempt is made to provide such a comparison. The performance characteristics of the D-Wave equipment were experimentally explored and reported in comparison to conventional software solvers (McGeoch and Wang, 2013). As mentioned in the “Background” section, the D-Wave hardware system shows speed-up thousands of times better than current software implementations, especially in QUBO, which is relevant to this financial model investigation.

#### 4.6.4 FUTURE RESEARCH

A future consideration is the change of risk measure (Gilli and Schumann, 2012), for instance, optimizing the portfolio’s VaR, which may not be solvable via classical methods if it does not yield to reformulation of the objective function.

*Weighted maximum independent set:* Other selection criteria or formulations could be chosen for the weight, such as VaR of the holding, and would be suitable as future investigations. It will be a useful future exercise to further characterize a version of this model and compare the results against current techniques.

Recommendations for future research include but are not limited to further analytical representation of the portfolio problem in the language of Boolean satisfiability; finding a model representation that determines the *weightings* of stocks in the resulting portfolio optimization; investigation of more complex representations of the selection criteria for the vertex weights

and edge determination, such as inclusion of min/max weight boundaries or inclusion of trading or carrying costs, etc.; extension of the problem to further financial domains, such as derivatives, fixed income instruments, or credit risk analysis. The average values used in this implementation were over the full sample period: perhaps other more typical sample periods and averaging (e.g., weighted or log decay) techniques would be appropriate for further investigation.

Since optimization methods such as heuristics do provide a way to solve similar problems using commercially available computers and applications, one of the most interesting aspects for future research would be to classify the boundary between what type and size of financial problem can be left to the current classical environments versus what type and size of problem should be solved in the new paradigm. It is also possible to reconsider the problem representations, investigated here, in QUBO form, which may be easier to understand and compare with other research with QUBOs. As well, a more general QUBO form may yield more flexibility compared to the specific MIS/WMIS representations considered in this chapter.

An investigation into the comparative financial performance of these models should be undertaken. That is, do the classical models and environments yield better or worse financial return, over certain timeframes, when compared with the implementation considered in this chapter?

Notably, the analysis herein has been with respect to historical returns. Indeed, the application of time-series models in this paradigm as a forecasting tool would be an interesting next step. Generally, if the stock returns were generated with a suitable forecasting model, for example, GARCH, then this environment would provide the best portfolio within the forecasting constraints.

The binary decision aspect of the technique is enticing to the extent that it provides a fairly simple means of determining a *yes* or *no* answer as to a stock's inclusion in a portfolio. The remaining, key, and unanswered question is how to determine the actual weightings of each holding within the portfolio: the method used here implicitly considered equal unit-weighting, that is, to purchase the same number of units of each stock to form the portfolio. Since this framework yields binary results, a method to determine the weightings (not to be confused with the weights attributed with each market graph node) would have to be concocted. A conjecture might be some sort of polynomial form involving the known characteristic variables, such as the number of edges connecting to a node, the variance

and return attributed to each node, the covariance edge values, etc. This leads to possible multilinear polynomial and posiform representations of a pseudo-Boolean function to represent the weights. Some insight into this area but in a wider optimization context can be seen in the study by Boros and Hammer, 2002. Yet, this is another area for future work.

## 4.7 Conclusion

---

This computational environment itself holds great promise toward solving hard financial engineering problems. However, the MIS and WMIS models, as investigated in this chapter, may have limited appeal. Further analytical and experimental investigation is required to improve upon the existing models or to characterize other viable alternative models.

## Acknowledgments

---

The author thanks D-Wave Systems Inc. for allowing to investigate this problem domain with the use of their simulation software. Zhengbing Bian (D-Wave) has been particularly valuable at helping me through the process of engaging D-Wave and in providing constructive feedback.

Thanks also go to supervising Professor David Starer, Stevens Institute of Technology, for his encouragement to take on this investigation.

## Appendix 4.A: WMIS Matlab Code

---

```
%%%% iteratedDWaveRuns_MeanVar_WMIS_compositeF
%
% PURPOSE: Perform iterations over the stock data using the D-Wave
% hardware optimization engine.
% Use Mean-Variance with WMIS optimizing F.
%
% Author: Michael Marzec, Stevens Institute of Technology
%
% the filename containing our stored data...
loadfile = 'myHistoricalStockData-20130408.mat';
%
% load the data from the file
load(loadfile, 'histData');
dataSize = size(histData(1).adjClse); % number of stocks retrieved
%
% Vector of our stock names...
% use Cell Array '{}'
Stocks = { ...
```

```

histData(1).symbol{1}; ...
histData(2).symbol{1}; ...
histData(3).symbol{1}; ...
histData(4).symbol{1}; ...
histData(5).symbol{1}; ...
histData(6).symbol{1}; ...
histData(7).symbol{1}; ...
histData(8).symbol{1}; ...
histData(9).symbol{1}; ...
histData(10).symbol{1} ...
};

% -----
% create a single Prices matrix from our structures
% -----
% For each column...
for k = 1:10
    % For each row...
    for i = 1:size(histData(k).adjClse)
        Prices(i,k) = histData(k).adjClse(i);
    end
end

% -----
% Need to calculate the log returns
% -----
logReturns = diff(log(Prices));

% Average log-return for each stock
avgLogReturns = mean(logReturns,1);

% -----
% Calculate the Variance-Covariance between pairs of stock log returns...
% -----
% Need the var/covar matrix...
covLogReturns = cov(logReturns);

% ...we'll use the Variance of the logReturns, as well...
varLogReturns = var(logReturns);

% ...also need the stdDev
stdDevLogReturns = sqrt(varLogReturns);

% a visualization of the Var-Covar...
bar(covLogReturns), xlabel('Stocks'), ylabel('Var-Covar'), legend(Stocks);

% -----
% Static variables/set-up for the D-Wave run
%
% Specify the embedding list...
% Adapted from D-Wave's K17_Q128_EMBEDDING
% -----
K10_Q128_EMBEDDING = cell(1,10);
K10_Q128_EMBEDDING{1} = [0 4 12 20 28];
K10_Q128_EMBEDDING{2} = [1 5 13 21 29];
K10_Q128_EMBEDDING{3} = [2 6 14 22 30];
K10_Q128_EMBEDDING{4} = [3 7 15 23 31];
K10_Q128_EMBEDDING{5} = [8 40 44 52 60];
K10_Q128_EMBEDDING{6} = [9 41 45 53 61];
K10_Q128_EMBEDDING{7} = [10 42 46 54 62];

```

```

K10_Q128_EMBEDDING{8} = [11 43 47 55 63];
K10_Q128_EMBEDDING{9} = [16 48 80 84 92];
K10_Q128_EMBEDDING{10} = [17 49 81 85 93];

% The J parameter used here; must be any real number > 1
% Note: this re-uses the normJ from the equal-weight example
% Actual condition for selecting normJ is: normJ > min{c_i, c_j}
normJ = 1.1;

% -----
% Here is the connectivity to the local or remote D-Wave solvers...
% -----
% Get a connection to the local solver...
conn=sapiLocalConnection();
% Get a list of solvers available locally...
solvers=sapiListSolvers(conn);
% Print the list of solvers available...
solvers

% create a handle to a particular solver
solver = sapiSolver(conn,'c4-sw_optimize');

% Create the embedding solver
embeddingSolver = sapiEmbeddingSolver(solver,K10_Q128_EMBEDDING);

% -----
% Re-used parameters from correlation-based runs:
% Iteration over various correlation threshold values
% from correlation = 0.35 to 0.65 by 0.05
% -----
%
% -----
% Generate a market graph...
% -----
isingSolutions = zeros(length(covLogReturns), 7);

% Init the optimumPortfolios array; this is a fast way - not the best.
optimumPortfolios = cell(10,7);

loopCount = 1;
for cThresh = 0.35 : 0.05 : 0.65

    % correlations are used to set up the graph
    marketGraph = covLogReturns;

    % iterate through our initial graph and determine Edges
    for m = 1:length(marketGraph)
        % determine connectivity using our threshold
        for n = 1:length(marketGraph)
            % connect or not connect pairs based on threshold
            if m==n
                marketGraph(m,n) = 0;
            elseif abs(marketGraph(m,n)) > cThresh*stdDevLogReturns(m)
                *stdDevLogReturns(n)
                marketGraph(m,n) = 1;
            else
                marketGraph(m,n) = 0;
            end
        end
    end
end

```

```

end
% print the market graph for each iteration
cThresh
marketGraph

% count the number of edges for each stock node
edgeCount = sum(marketGraph, 2);
% -----
% Calculate the Ising parameters...
% Note: Here, node weights are the (return - variance)
% values for each stock, ie node.
% -----
% the h_ij values
for p = 1:length(edgeCount)
    % This uses the variance handicapped returns expressed in F
    hValues(p) = normJ*edgeCount(p)-2*(avgLogReturns(p)-varLogReturns(p));
end

% the J_ij values
JValues = normJ*marketGraph;

% Solve the Chimera-structured Ising problem
answerIsing = sapiSolveIsing(embeddingSolver,hValues, JValues);

for t = 1:length(answerIsing.solutions)
    if answerIsing.solutions(t) == 1
        %optimumPortfolios(t,loopCount) = Stocks(t);
        optimumPortfolios(t,loopCount) = java.lang.String(Stocks(t));
    else
        optimumPortfolios(t,loopCount) = java.lang.String('');
    end
end

% add to a matrix of solutions
isingSolutions(:,loopCount) = answerIsing.solutions;
loopCount = loopCount + 1;
%gplot(marketGraph, [1 1; 2 4; 3 2; 4 8; 5 5; 6 10; 7 7; 8 3; 9 9; 10 6], '-*')
end

% print the optimum portfolios
optimumPortfolios
% -----
% THE END
% -----

```

## References

- Akrotirianakis, I., and B. Rustem. Globally convergent interior-point algorithm for nonlinear programming. *Journal of Optimization Theory and Applications*, 2005: v125n3: 497–521.
- Bartholomew-Biggs, M. *Nonlinear optimization with financial applications*. Boston, MA: Kluwer Academic Publishers, 2005.

- Bertsimas, D., C. Darnell, and R. Soucy. Portfolio construction through mixed-integer programming at Grantham, Mayo, Van Otterloo and Company. *Interfaces*, 1999; 29: 49–66.
- Bian, Z., F. Chudak, W. Macready, and G. Rose. *The Ising model: Teaching and old problem new tricks*. D-Wave Systems, 2010.
- Boginski, V., S. Butenko, and P. M. Pardalos. *Mining market data: A network approach*. Elsevier Ltd., 2005.
- Boixo, S., T. Albash, F. Spedalieri, N. Chancellor, and D. Lidar. *Experimental signature of programmable quantum annealing*. University of Southern California, 2012.
- Boros, E., and P. Hammer. Pseudo-boolean optimization. *Discrete Applications in Applied Mathematics*, 2002; 123: 155–225.
- Boros, E., P. L. Hammer, R. Sun, and G. Tavares. A max-flow approach to improved lower bounds for quadratic unconstrained binary optimization (QUBO). *Discrete Optimization*, 2008; 5(2): 501–529.
- Charpin, F., and D. Lacaze. Using binary variables to obtain small optimal portfolios. *The Journal of Portfolio Management*, 2007: Fall: 68–72.
- Chaves, D., J. Hsu, F. Li, and O. Shakernia. *Risk parity portfolio vs. other asset allocation heuristic portfolios*. Research Affiliates, LLC, 2010.
- Chen, W., and L. Zhang. Global optimality conditions for quadratic 0-1 optimization problems. *Journal of Global Optimization*, 2010; 46(2): 191–206.
- Choi, V. *Adiabatic quantum algorithms for the NP-complete maximum-weight independent set, exact cover and 3SAT problems*. Falls Church, VA: Department of Computer Science, Virginia Tech, 2010.
- Choi, V. *Minor-embedding in adiabatic quantum computation: I. The parameter setting problem*. D-Wave Systems Inc., 2008.
- Cormen, T., C. Leiserson, R. Rivest, and C. Stein. *Introduction to algorithms*. 3rd ed. Cambridge, MA: MIT Press, 2009.
- Dang, C., and L. Xu. A barrier function method for the nonconvex quadratic programming problem with box constraints. *Journal of Global Optimization*, 2000; 18: 165–188.
- D-Wave. Programming with QUBOs. Programmer’s Guide. Burnaby, British Columbia (BC), Canada: D-Wave Systems Inc., 2013a.
- D-Wave. Quantitative finance tutorial: optimizing portfolios. *D-Wave Developer Portal*. Burnaby, British Columbia (BC), Canada:

- D-Wave Systems Inc., 2013b. <http://www.dwavesys.com/en/dev-tutorial-finance.html> (accessed January 2013).
- D-Wave. Solving weighted maximum independent set using D-wave blackbox. *D-Wave Developer Portal*. Burnaby, British Columbia (BC), Canada: D-Wave Systems Inc., 2013c. <http://www.dwavesys.com/en/dev-tutorial-wmis.html>.
- Elton, E., M. Gruber, S. Brown, and W. Goetzmann. *Modern portfolio theory and investment analysis*. John Wiley & Sons, Inc., 2007.
- Fabozzi, F. J., H. M. Markowitz, P. N. Kolm, and F. Gupta. Mean-variance model for portfolio selection. In *Encyclopedia of Financial Models*, by Frank J. Fabozzi 2013. <http://onlinelibrary.wiley.com/doi/10.1002/97811182635.efm0003/full>.
- Farhi, E., J. Goldstone, S. Gutmann, and M. Sipser. Quantum computation by adiabatic evolution. 2000. <http://arxiv.org/pdf/quant-ph/0001106.pdf>.
- Floudas, C., and C. Gouraris. A review of recent advances in global optimization. *Journal of Global Optimization*, 2009; 45: 3–38.
- Gilli, M., and E. Schumann. Heuristic optimisation in financial modelling. *Annals of Operations Research*, 2012; 193: 129–158.
- Gilli, M., D. Maringer, and E. Schumann. *Numerical methods and optimization in finance*. Waltham, MA: Academic Press, Elsevier Inc., 2011.
- Hillier, F., and G. Lieberman. *Introduction to operations research*, 9th ed. New York: McGraw-Hill, 2010.
- Huang, K. Y., C.-J. Jane, and T.-C. Chang. An enhanced approach to optimizing the stock portfolio selection based on Modified Markowitz MV Method. *Journal of Convergence Information Technology*, 2011; 6(2): 226–239.
- Jallo, D., and D. Budai. *The Market Graph: A study of its characteristics, structure and dynamics*. Master's thesis. The Royal Institute of Technology, December 2010.
- Kozen, D.. *The design and analysis of algorithms*. New York: Springer-Verlag, 1992.
- Krokhmal, P., J. Palmquist, and S. Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *The Journal of Risk*, 2002; 4(2): 11–27 [http://www.ise.ufl.edu/uryasev/files/2011/11/kro\\_CVaR.pdf](http://www.ise.ufl.edu/uryasev/files/2011/11/kro_CVaR.pdf).

- Maringer, Dr. *Risk preferences and loss aversion in portfolio optimization*. Colchester, UK: CCFEA, University of Essex, 2006.
- Markowitz, H. Portfolio selection. *The Journal of Finance*, 1952: 7(1): 77–91.
- McGeoch, C., and C. Wang. Experimental evaluation of an adiabatic quantum system for combinatorial optimization. In: Conference on Computing Frontiers, ACM, 2013.
- Palubeckis, G. Multistart tabu search strategies for the unconstrained binary quadratic optimization problem. *Annals of Operations Research*, 2004: 131: 259–282.
- Papadimitriou, C., and K. Steiglitz. *Combinatorial optimization: Algorithms and complexity*. Mineola, NY: Dover Publications, 1998.
- Pardalos, P. M., O. A. Prokopyev, O. V. Shylo, and V. P. Shylo. Global equilibrium search applied to the unconstrained binary quadratic optimization problem. *Optimization Methods and Software*, 2008: 23(1): 129–140.
- Rose, G. The 128-qubit Rainier chip: II. Graph embedding. 2008. <http://dwave.wordpress.com/2008/10/23/the-128-qubit-rainier-processor-ii-graph-embedding/>.
- Shaw, D., S. Liu, and L. Kopman. Lagrangian relaxation procedure for cardinality-constrained portfolio optimization. *Optimization Methods and Software*, 2008: June: 411–420.
- Soleimani, H., H. R. Golmakani, and M. H. Salimi. Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization using genetic algorithm. *Expert Systems With Applications*, 2009: 5058–5063.

## Chapter Five

# Estimation Procedure for Regime Switching Stochastic Volatility Model and Its Applications

Ionut Florescu and Forrest Levin

*Financial Engineering Division, Stevens Institute of Technology,  
Hoboken, NJ, USA*

### 5.1 Introduction

Consider a continuous time stochastic signal whose variability is constant. However, the variability changes its value at random times. We may view this behavior of the variability as a regime switching model or a hidden Markov chain model. The purpose of the present work was to introduce a new methodology based on particle-filtering techniques to estimate the behavior of the variability. We then study applications.

The structure of this chapter is as follows. The remainder of this section gives an overview of the model and existing literature. Section 5.2 describes the main methodology employed in this chapter. Section 5.3 contains the results and conclusions obtained when using the estimating methodology with real data from two different areas: finance (Section 5.3.1) and geophysics (5.3.3). Section 5.4 concludes the chapter. Appendix 5.A.4 contains some of the proofs.

### 5.1.1 THE ORIGINAL MOTIVATION

The original motivation of the work comes from the world of Finance. Today, it is well accepted in the financial literature that the Black–Scholes model (Black and Scholes, 1973) is not complex enough to capture the dynamics of data sampled with high frequency (e.g., see Mariani et al., 2009). Various other models extending the original one have been proposed. We mention here stochastic volatility models (Shephard, 2005), jump diffusion models (Merton, 1992), and general Lévy models (Cont and Tankov, 2003).

In this chapter, we present a model where the variability of the signal is driven by a continuous time Markov chain. In the financial literature, the model we analyze here is called a regime switching volatility model, or a stochastic volatility model with volatility driven by a hidden Markov chain (Hamilton and Lin, 1996; Hamilton, 2005; Chib et al., 2004). While regime switching (in the drift) is prevalent in science and engineering, to our knowledge, the volatility switching model has not been applied outside finance literature.

### 5.1.2 THE MODEL AND THE PROBLEM

We assume as given a complete probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ . On this space, we are given a complete filtration  $\mathcal{F} = \{\mathcal{F}_t\}_t$  (see Protter, 2005, p. 3). We consider a continuous time signal  $S$ , which satisfies the following stochastic differential equation:

$$S_t = S_0 + \int_0^t rS_u du + \int_0^t S_u Y_u dW_u, \quad (5.1)$$

where  $r$  is known,  $W_t$  is a standard Brownian motion with respect to the filtration  $\mathcal{F}$  defined on the original probability space, and  $Y_t$  is modeled as a continuous time Markov chain. The Markov chain has finite state space  $\{a_1, \dots, a_p\}$ , transition probability matrix  $\Lambda = (\lambda_{ij})$ , and the transition times given by exponential random variables with parameter  $\lambda_i$ , that is,

$$\mathbf{P}(Y_s = i, \forall s \in (t, t+u] | Y_t = i) = e^{-\lambda_i u}.$$

In order for the stochastic differential equation (5.1) to make sense, we assume that both processes  $S$  and  $Y$  are adapted to the filtration  $\mathcal{F}$ . We note that the process  $Y$  is not directly observable, this process will be referred as the “hidden factor” or the “hidden Markov process.”

Our goal is to estimate of the number of states  $p$ , the values of the states  $\{a_1, \dots, a_p\}$ , as well as the transition rates  $\lambda_i$  and the transition probability matrix  $\Lambda$  based on a discrete sample  $\{S_{t_1}, S_{t_2}, \dots, S_{t_n}\}$  of the observed signal  $S_t$ . We note here, since we wish to apply the model to fields other than Finance, we do not use any other input other than a history of the process  $S$ . Specifically, we do not rely on options or futures or indeed any derivative data.

### 5.1.3 A BRIEF HISTORICAL NOTE

As already mentioned, the literature on the model presented in equation (5.1) can be found in the finance area. In finance, the problem of estimating parameters of the model (5.1) was studied in the 1990s; however, either the number of states and the actual state values are given a priori (Hamilton and Lin, 1996), or the estimation technique uses specific financial derivative data such as call and put options (Barndorff-Nielsen and Shephard, 2002).

We discuss in more detail two techniques that are related to our methodology. The first technique performing a full estimation technique using only the observed signal may be found in the studies by Cvitanić, Liptser, and Rozovskii (2006) and Cvitanić, Rozovskii, and Zaliapin (2006). The first article presents the theoretical results, while the second implements the theory and presents numerical results. Our methodology is different in the following aspects. While the model used is similar, the estimation methodology we propose is different. Specifically, the work cited uses a sequential Bayes methodology designed for the particular model under consideration. This allows the authors to obtain specific convergence rates for their algorithm. Furthermore, the analysis devoted to the number of states and the specific values of the states of the hidden Markov chain is somewhat *ad hoc* and is performed by visual inspection of the results given by the Bayesian filter. In our work, we use a more general particle-filtering method; however, we are not capable of obtaining precise convergence rates without making extra assumptions on the model. Our methodology is capable of recognizing the correct number of states using a mechanical approach easy to implement in a computer.

A second technique that we see related to our methodology is provided by Genon-Catalot, Jeantheau, and Larédo (2000). This article culminates a series of three earlier papers dedicated to parameter estimation for stochastic volatility models. In the cited article, the authors use a filtering methodology similar to our approach to a more general stochastic volatility model.

The process  $Y_t$  in the cited work is a stationary continuous time and continuous state space process. After using the particle filter, the authors use a method of moments estimation for the parameters. In our context when  $Y_t$  has a discrete distribution translating this method would need special care since there are multiple discrete values that provide the same moments up to many orders. This is why, we use a different estimation technique based on our specific discrete model.

For completeness, we mention earlier work dedicated to estimating parameters of stochastic volatility models from discrete data. Even though these papers are using models different from our work, they were a great source of inspiration. Nielsen and Vestergaard (2000), Sorensen (2003), Bladt and Sorensen (2007), and Ait-Sahalia and Kimmel (2007) are some of these important references for us.

## 5.2 The methodology

---

The theoretical foundation of the estimating methodology is extending fundamental work in Del Moral, Jacod, and Protter (2001). We note that because of the specific form of the process  $S_t$  in (5.1), we may consider the process  $X_t = \log S_t$ . This process has a simplified dynamic as an application of the Itô's lemma shows:

$$X_t = X_0 + \int_0^t \left( r - \frac{Y_u^2}{2} \right) du + \int_0^t Y_u dW_u. \quad (5.2)$$

The  $Y_t$  process remains unmodified under this transformation and the observations are now  $\{x_{t_1}, \dots, x_{t_T}\}$  with  $x_{t_i} = \log s_{t_i}$ .

We use a particle-filtering methodology on the basis of the study by Del Moral et al. (2001) to estimate the distribution of the volatility process  $\{Y_t\}_t$  at time  $t$ . We then use the theoretical distribution of  $Y_t$  to find parameter values that allows the best match with the estimated distribution. This estimation methodology is described in the next two sections.

### 5.2.1 OBTAINING FILTERED EMPIRICAL DISTRIBUTIONS

AT  $t_1, \dots, t_T$

Suppose that we are looking at the process evolution over the interval  $[t_{i-1}, t_i]$ . We are given the two endpoints  $x_{t_{i-1}}, x_{t_i}$ , and a previous estimate of the distribution of the volatility process  $Y_{t_{i-1}}$ , denoted by  $\Phi_{i-1}$ . We want

to estimate the distribution of the volatility process at  $t_i$ :  $Y_{t_i}$ . To proceed, we need to choose the following basic elements:

- A weighting function  $\varphi(\cdot)$ . This would be called a kernel function in other areas of statistics (any nonnegative function with  $L^1$  norm equal to 1), but for practical consideration it has to have most of the mass centered around zero. The kernel does not need to be symmetric, though in practical applications it always is. In practice, we normally do not know if we should penalize more if we under- or overestimate parameters; thus, we always choose a symmetric  $\varphi$ . In our applications, we typically use one of the triangular, Epanechnikov, or Gaussian kernels. The choice used has typically little effect on the estimation.
- A number  $n$  that represents the number of intermediate paths for the filter. This number is typically large.
- A number  $m$  that represents the number of intermediate points within the interval  $[t_{i-1}, t_i]$ . For theoretical convergence reasons, we need to have  $m \geq \sqrt[3]{n}$ .

With these basic elements, the filtering methodology consists in two steps: an evolution step and a selection step.

### 5.2.1.1 Evolution step

In this step of the algorithm, we create  $n$  paths from  $t_{i-1}$  to  $t_i$ , using the following process evolution. Take  $\Delta_{i,m} = \frac{t_i - t_{i-1}}{m}$  and draw a value  $Y_{i-1}$  from the previously estimated distribution  $\Phi_{i-1}$ . Each evolution path  $j$  is then constructed using a simple Euler scheme:

$$\begin{cases} \tilde{X}_{t_{i-1}+k\Delta_{i,m}}^j = \tilde{X}_{t_{i-1}+(k-1)\Delta_{i,m}}^j + \left( r - \frac{(\tilde{Y}_{t_{i-1}+(k-1)\Delta_{i,m}}^j)^2}{2} \right) \Delta t \\ \quad + \tilde{Y}_{t_{i-1}+(k-1)\Delta_{i,m}}^j \Delta W_k^j \\ \tilde{Y}_{t_{i-1}+k\Delta_{i,m}}^j = \tilde{Y}_{t_{i-1}+(k-1)\Delta_{i,m}}^j + \beta \Delta Z_k^j, \quad k \in \{1, \dots, m\} \end{cases} \quad (5.3)$$

where  $\Delta W_k^j$  and  $\Delta Z_k^j$  represent the increments of independent Brownian motions [i.e., they are i.i.d.  $N(0, \Delta_{i,m}^2)$ ].  $\beta$  is a parameter of the algorithm. At the end of each evolution path, we keep the pairs  $(\tilde{X}_{t_i}^j, \tilde{Y}_{t_i}^j)$ ,  $j = 1, \dots, n$ .

### 5.2.1.2 Selection step

The idea of this step is to give greater weights to the “better” values obtained. Specifically, we assign a weight to each final value  $\tilde{Y}_{t_i}^j$  equal to:

$$w_j = C(n) \sqrt[3]{n} \varphi \left( |\tilde{X}_{t_i}^j - x_{t_i}| \sqrt[3]{n} \right),$$

where  $C(n)$  is a normalizing constant that makes the weights sum to 1.

Our approximating distribution at the end of step  $i$  is the discrete distribution  $(\tilde{Y}_{t_i}^j, w_j)_{j \in \{1, \dots, n\}}$ . We denote this distribution  $\Phi_i$ . The selection step is equivalent to an importance sampling technique (Florescu and Tudor, 2013, Chapter 6).

### 5.2.2 OBTAINING THE PARAMETERS OF THE MARKOV CHAIN

Once we have the filtered distributions  $\Phi_1, \dots, \Phi_T$ , we calculate the means for each of them  $\bar{Y}_1, \dots, \bar{Y}_T$  and we take these as estimated realizations of the process  $Y_t$ . The next step is straightforward. The values  $\bar{Y}_1, \dots, \bar{Y}_T$  represent realizations from a mixture of Gaussian distributions with means precisely  $a_1, \dots, a_p$ , the states of the Markov chain. Thus, we may use one of the established methods to estimate the nature of the means, for example, the Expectation Maximization (EM) algorithm. However, the EM algorithm requires knowledge of the number of distributions present and it is slow so we prefer a simpler method. We follow the Minimum Error Thresholding method of Kittler and Illingworth as improved by Cho, Haralick, and Yi (1989). This method assumes that the best separating values for the mixture in the distribution are at  $T_1, \dots, T_{p-1}$ . Given this assumption, the method then calculates the likelihood of seeing the values  $\bar{Y}_1, \dots, \bar{Y}_T$ . In this way, the method estimates the likelihood function. The maximization may be performed using any nonlinear optimization method. Under the hypothesis of multivariate normal distribution for  $Y$  the function to be maximized has a global maximum. The result of the maximization is the set of optimal threshold values  $\hat{T}_1, \dots, \hat{T}_{p-1}$ . Once these are obtained, the algorithm uses a truncated normal methodology to obtain estimates of the means for each distribution in the mixture. These are our estimates for  $a_1, \dots, a_p$ . The value of  $p$  is chosen such that it corresponds to the largest maximum likelihood. Complete details of the methodology may be found in Levin (2010).

Finally, to obtain the rates  $\lambda_i$  and the transition probability matrix  $\Lambda$ , we use the estimated states  $\hat{a}_1, \dots, \hat{a}_p$  and the thresholds  $\hat{T}_1, \dots, \hat{T}_{p-1}$ . We

go back to the estimated sequence  $\bar{Y}_1, \dots, \bar{Y}_T$  for a second time and we link them by grouping the values that are within the estimated thresholds  $\hat{T}_1, \dots, \hat{T}_{p-1}$  and corresponding to the  $a_i$  values. This allows to identify the time periods when the Markov chain stays constant in state  $i$  as well as the destination of subsequent jumps. Using the times, we then estimate the rates  $\lambda_i$ , using a simple exponential distribution, and using the destination of the jumps, we estimate the transition probabilities and correspondingly the rates  $\lambda_{ij}$ .

In Appendix 5.A.4, we present convergence results, both theoretical and empirical. We feel that these results, while demonstrating the algorithm provides correct results, will detract from the principal merit of the paper that is the application to real data in the next section.

It is worth mentioning that the correct identification of the rates  $\lambda_i$  and  $\lambda_{ij}$  is the most difficult problem. The only hope for the estimation methodology to work correctly is to have enough samples to detect when the regime shift is happening. It is for this reason that the methodology is particularly suitable to analysis of data sampled with high frequency so that regime shifts are correctly identified.

## 5.3 Results obtained applying the model to real data

### 5.3.1 PART I: FINANCIAL APPLICATIONS

Volatility, defined as the squared root of the quadratic variation, has a crucial role in financial applications. It provides a way to measure the risk associated with a financial asset. This has been known for a long time in Mathematical Finance and indeed estimating volatility has a special place in this area. The estimation uses either historical data (realized volatility, integrated volatility, heteroskedastic models; Tsay, 2005; Shephard, 2005) or interpolating values computed using financial derivatives written on the stock (implied volatility, local volatility surfaces, etc.). However, despite the long history, in practice there are irreconcilable differences between the values obtained using historical values and on-the-spot derivative values. Traditionally, the difference is explained using the time paradigm—implied volatility reflect present risk values while the historical volatility represents risks from the past and including the present.

In the present work, we work under the following assumption. There are shifts in volatility value; these shifts are captured using the methodology

presented in this chapter. These shifts may be due to news, extraneous events, or trading patterns that change during the day.

The financial data we chose to illustrate the method is from March 2008. Bear Stearns began the week on Monday, March 10, 2008, with a market stock price of about \$70/share. The 85-year-old firm, the fifth largest Wall Street securities firm at the time, had never recorded a quarterly loss until the fourth quarter of the previous year (2007). They were anticipating a profit for the first quarter report for 2008. At the end of the week (Friday, March 14, 2008) the share price had dropped to \$30 and over the weekend Bear Stearns, JPMorgan and the Federal Reserve Bank of New York arranged for JPMorgan to Purchase Bear Stearns for \$2/share to avoid a bankruptcy filing (source for general information: Cohan, 2009).

How did an apparently very healthy firm get in this situation? We do not know and perhaps will never know without analyzing the internal balance of the company. However, this is a perfect setting for testing our model. By getting closer to the conclusion of that week, the market should become more and more volatile and consequently the volatility should perform a shift in magnitude. Will we be able to capture this shift (provided it exists)? Will we be able to determine an exact moment when the crisis had started? Were other companies affected by the collapse? Will we be able to determine all this from the analysis of stock data alone? All these are questions that we hoped to answer when we started the analysis of the minute data.

### 5.3.1.1 The events as they happened

At the time of the events mentioned, there had already been a large bankruptcy related to companies involved in mortgage instruments: the buy-out of Countrywide Financial, “the nation’s largest mortgage lender” by Bank of America early in January 2008 (Landon and Sorkin, 2008; Ross Sorkin, 2008; Dash, 2008, —*New York Times* articles). Many investment banks and other corporations, either large or small, and including government agencies were in the volatile subprime market. From the market facts and the media attention, both the corporate sector and the investing public were aware of the subprime mortgage loan crisis.

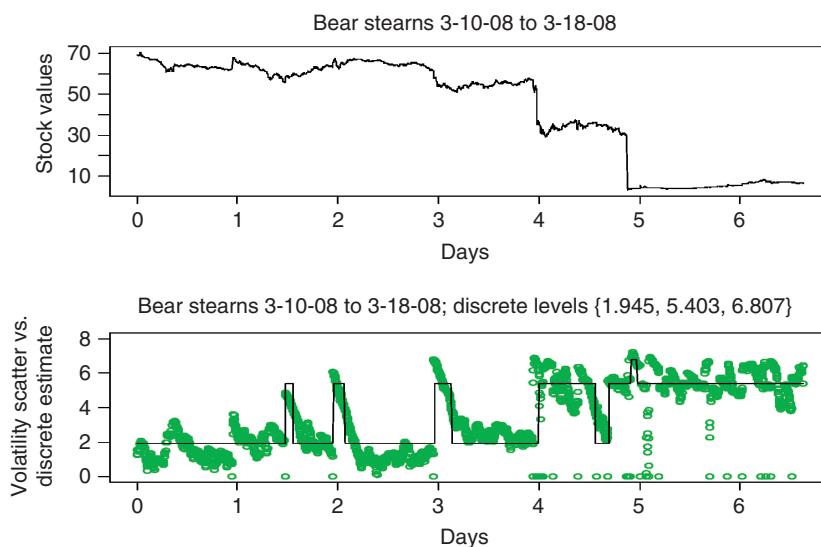
The people in charge of the business at Bear Stearns (including those working at the time for the company) knew that there was trouble with funding, overnight loans, and liquidity in general from the start of the week of the collapse (source: undisclosed former employee). Throughout

the investment industry, there were warning signals, in particular, home mortgage defaults (Geithner, 2008).

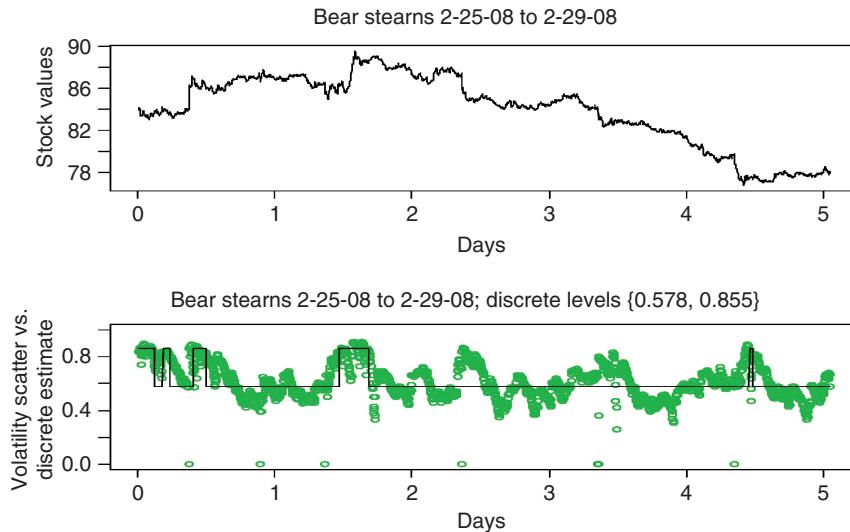
As late as Tuesday March 11, the stock market was heading up after the Federal Reserve opened up its lending policies to investment companies in addition to just commercial banks (which was the previous practice). On the evening of Thursday March 13, Bear Stearns executives were meeting with SEC regulators and Tim Geithner (president of NY Fed). The president of the Federal Reserves Bank of New York expressed the opinion that Bear Stearns would have to file for bankruptcy (source: U.S. Senate Committee on Banking, Housing and Urban Affairs, 2008). We assume that the market would react on Friday to this news so it is interesting to analyze the volatility behavior.

### 5.3.1.2 Data analysis and results

In Figure 5.1, we may observe the behavior of the Bear Stearns volatility during the week of March 10–14 and continuing with the following two days (trading did not stop after the weekend since the company still existed although under different management). The estimated states for the volatility of the Bear Stearns stock are 1.9, 5.4, and 6.8. All these values are outside the normal  $[0, 1]$  annual volatility values generally characteristic for the stock market. Furthermore, clearly the volatility levels shift upward



**FIGURE 5.1** Bear Stearns stock price/volatility March 10 to March 18, 2008.

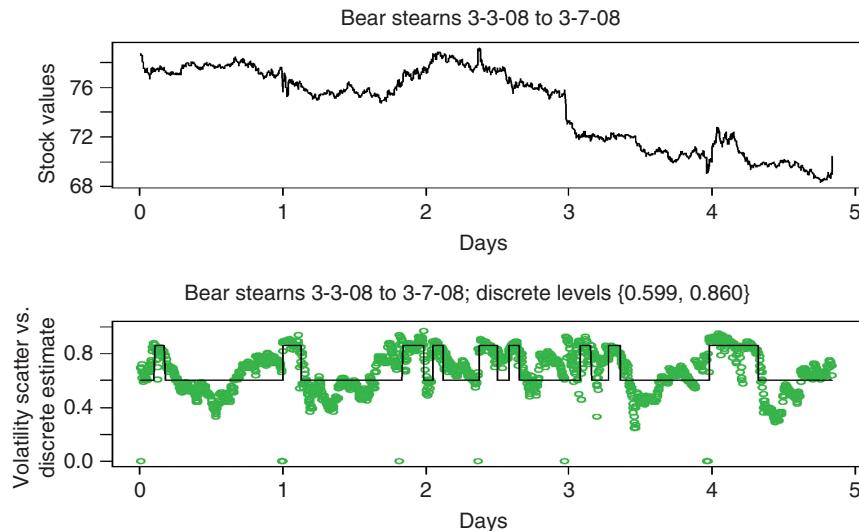


**FIGURE 5.2** Bear Stearns stock price/volatility 2 weeks before collapse.

toward the end of the week starting as early as the middle of Tuesday and certainly after the Thursday meeting.

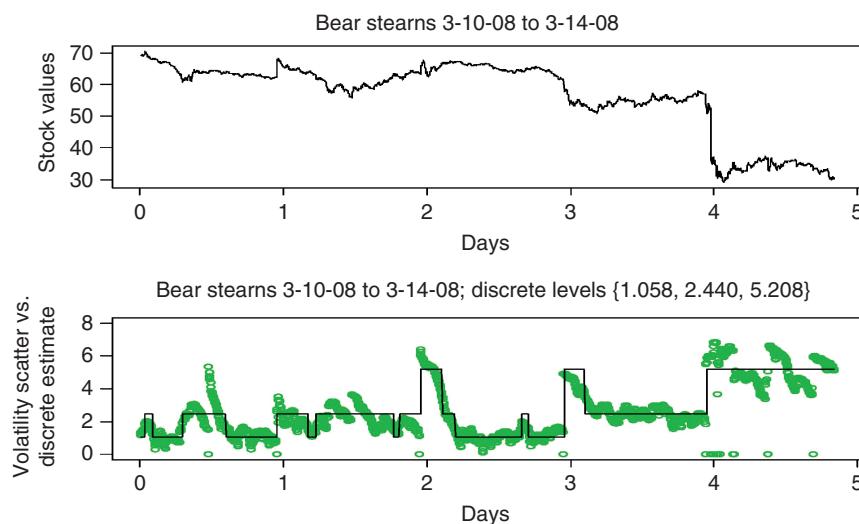
We recognized that the high volatility values may be due to the Monday (March 17) and Tuesday (March 18) following the buy-out announcement. During those days, the stock traded at \$3–\$4 range even though the announcement was that shares will be bought by JPMorgan at \$2/share. The explanation for the high trading activity came on March 24, 2008, when a class action lawsuit was filed on behalf of shareholders, challenging the terms of JPMorgan recently announced acquisition of Bear Stearns. That same day, a settlement was reached that raised JPMorgan Chase's offer to \$10 a share.

Recognizing the situation, we had eliminated the days after the buy-out and have redone the analysis separately for each of the 3 weeks preceding March 14 (see Figures 5.2, 5.3, and 5.4 on page 8). From these pictures, we see that the volatility levels ( $\sim 0.59$  and  $\sim 0.86$ ) for the 2 weeks preceding the critical week are similar. We see that the levels during the final week are all higher than 1 (1.1, 2.4, and 5.2) so the whole week behavior is abnormal. The second week does exhibit prolonged periods of higher volatility so we decided to pull the final 2 weeks (March 3–14, 2008) together for a final analysis in Figure 5.5. In this figure, we can finally see that abnormal levels were touched in the beginning and end of day 4 (Friday, March 7) and

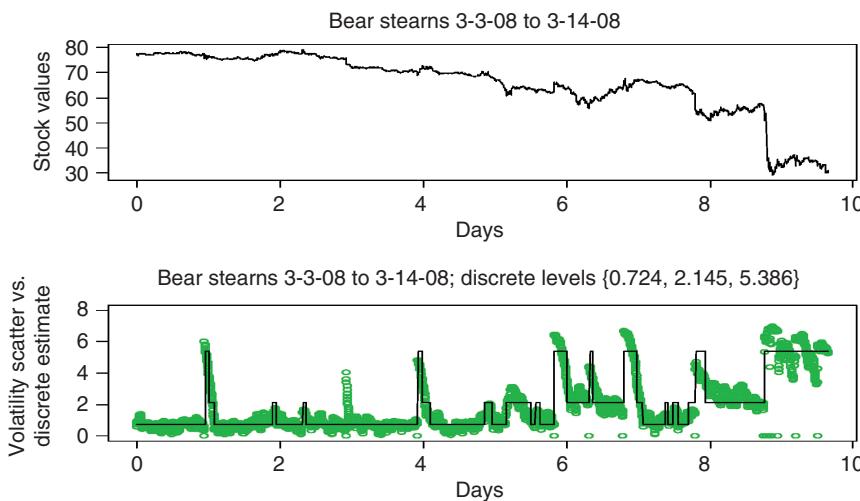


**FIGURE 5.3 Bear Stearns stock price/volatility 1 week before the collapse.**

certainly during the entire next week (as we had observed from the earlier analysis). Furthermore, we do observe lowered values for volatility during the Wednesday March 12, this is the day before the SEC meeting with Bear Stearns executives and we suspect from this behavior that the bankruptcy announcement came as a surprise to the general market participants.



**FIGURE 5.4 Bear Stearns stock price/volatility week of the collapse.**



**FIGURE 5.5 Bear Stearns stock price/volatility during the last 2 weeks.**

We have performed a thorough investigation, using data for several other equities from the investment sector (JPMorgan, Lehman Bros, Bank of America, Merrill Lynch) as well as noninvestment sector (IBM, Chevron, Exxon, FedEx) during the week of the collapse March 10–14. For lack of space, we present the plots and the detailed results for these stocks in the supplementary material Supplement B. The results that were evident from this analysis are commented next.

The volatility levels estimated for each stock in the investment sector during the week are extremely high (though not at the levels encountered for Bear Stearns). Beyond the actual values, all of them exhibit a common behavior during the week. All stocks went to high levels at the beginning of the trading day on Thursday until about the middle of the day and then again on Friday. The corresponding stock values do not exhibit any obvious pattern from which this volatility behavior may be easily obtained.

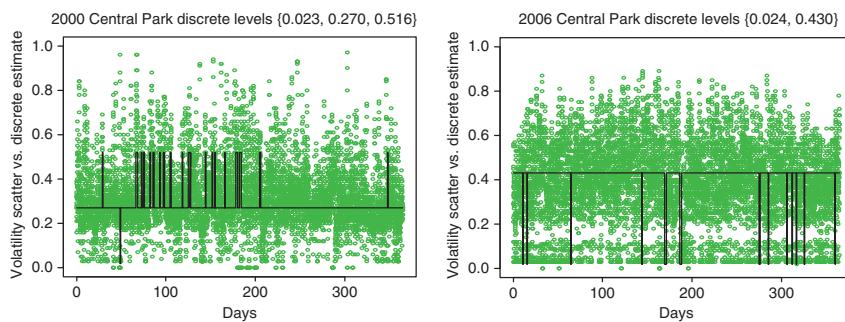
When looking at the noninvestment sector, we finally observe volatility levels that are characteristic of the respective stock. For example, the estimated levels for IBM were 0.36, 0.63, and 0.78 (the typical implied volatility level for IBM stock calculated from option data using the Black Scholes model is historically around 0.4). Furthermore, IBM does exhibit increased levels of volatility as early as Thursday morning and further increasing during Friday. The oil companies stock analyzed (Chevron and Exxon Mobil) are very similar in both volatility values and behavior. They

show only an increase in volatility starting Friday. This discrepancy in the volatility behavior between technology stock and others was explained to us by a trader who stated that in general technology, stocks react much faster to the market movement than the commodities stock. Among the noninvestment sector, we also analyze FedEx volatility behavior during the week. This equity behavior exhibits slight to no reaction to the Bear Stearns collapse. In conclusion, we believe that the model provided insight into market behavior during that week. First, we believe that the volatility levels for Bear Stearns equity were abnormal starting with Friday March 7, 2008, and continued throughout the week of March 10–14 until the collapse of the investment firm. Signs were present in other investment equity as well but none as in the Bear Stearns case. We see higher volatility levels on Thursday morning (before SEC meets with the Bear Stearns officials) exhibited by the entire investment sector as well as the highly traded technology stock. Commodity sector was affected on Friday following the remarks by the Federal Reserve Bank of New York that a bankruptcy is imminent. The event did not seem to affect the entire market only the highly traded equity.

### 5.3.2 PART II: PHYSICAL DATA APPLICATION. TEMPERATURE DATA

Global warming seems to be a “hot” topic these days so we wanted to see if we may adapt the model to study climate-related problems. Temperature record is the most important signal for global warming. The problem we faced was finding temperature data sampled with high enough frequency (which is needed to detect shifts in variability). The highest frequency sampled data we found was hourly temperature data gathered in Central Park, New York City, New York, USA, starting in 2000. *We realize that this data is obtained for a very large park located in a middle of arguably the largest metropolis on Earth and therefore our results are not to be extrapolated.* However, the results we found when estimating variability of this data were most interesting. We present two typical pictures of estimated volatility in Figure 5.6 on page 10 (the rest of images are in the supplementary material Supplement C).

It is remarkable that all the plots show evidence of a predominant variability value and the deviations from this value do not last very long. Accordingly, we conclude that in general the hourly data show a large degree of consistency in variability of temperature. Consequently,



**FIGURE 5.6** Hourly temperature data from 2000 and 2006.

Table 5.1 presents the value of the predominant variability each year. The last data available was from July 2010 and in order to have a consistent way of estimating we overlap some time period in the end to provide an entire year worth of data.

When looking at these values, one clear distinction stands out. For the years 2000–2005, the variability is remarkably close to 0.3. Beginning with 2006, there is a notable jump in variability (the rows are separated in the table for a better visibility) and the variability level stays remarkably close to 0.42. What exactly does this mean and what caused the shift in the temperature variability we do not know but we interpret this as some evidence of a certain change in the local climate in the New York City area.

**TABLE 5.1** Yearly predominant node volatility  
2000–10 Central Park, New York City.

Year	Predominant volatility level
2000	0.27
2001	0.269
2002	0.271
2003	0.276
2004	0.276
2005	0.293
2006	0.43
2007	0.423
2008	0.425
08/08–07/09	0.424
08/09–07/10	0.411

We would have liked to compare the results obtained in this way with data from earlier years. Unfortunately, before 2000 the data was not gathered with high enough frequency. The only other temperature data we found was gathered at the same location (Central Park, New York City) from 1976 to 1977, but this data was sampled every 3 hours and thus the numerical results obtained are not directly comparable. We may see the output in Figure 5.7 on page 11. This time the filter outputs only one volatility node as best fitting the data. The values for the two consecutive years are close too (0.49 vs. 0.48). However, as mentioned, we may not compare this value with the results obtained for the 2000–2008 since the sampling frequency is different. Having said that, if the stationary distribution of the stochastic volatility process is Gaussian (as it seems to be), then a simple transformation should give an indication of the hourly volatility value.

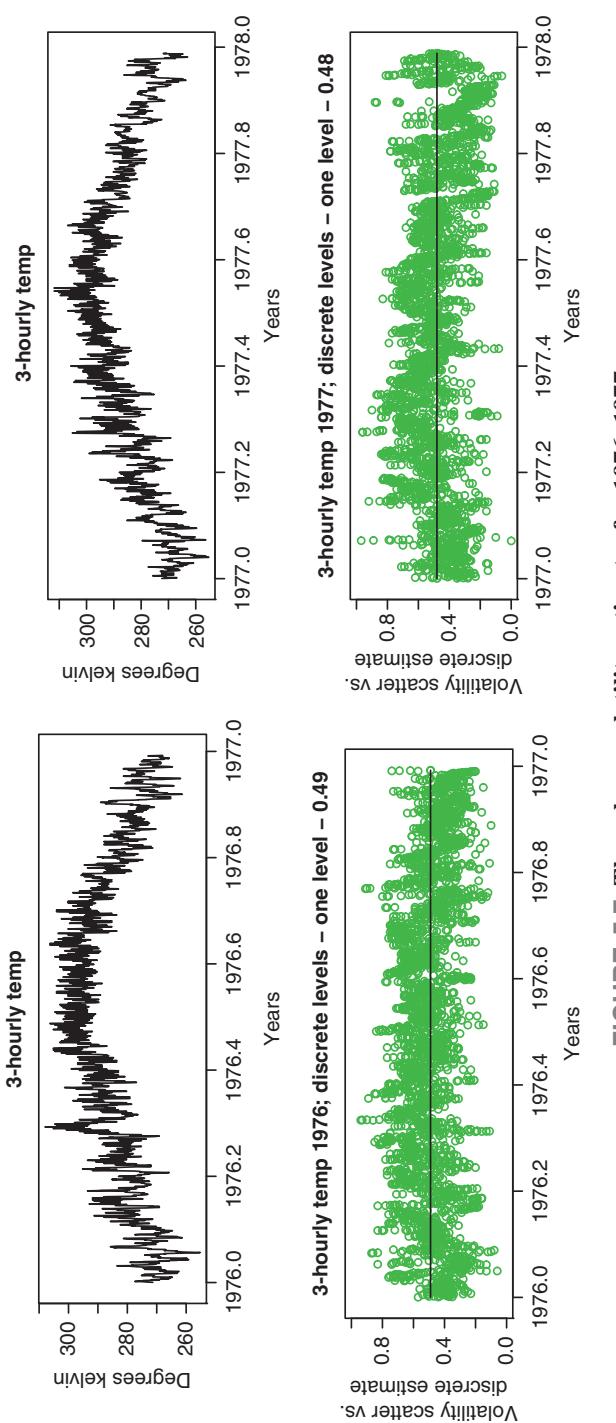
$$\sigma_{3h} = \sigma_h * \sqrt{3}$$

With this, we may guess the hourly values as  $0.48/\sqrt{3} = 0.277$  and  $0.49/\sqrt{3} = 0.283$ , which certainly look consistent with values we obtained from 2000 to 2005.

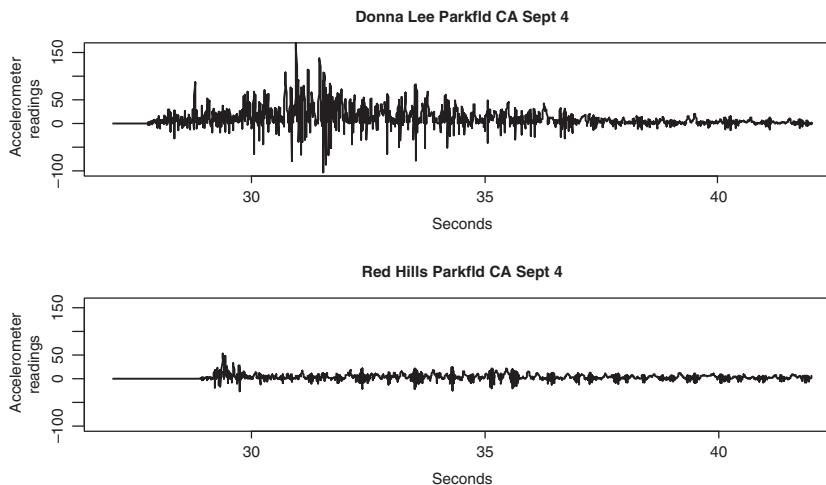
### 5.3.3 PART III: ANALYSIS OF SEISMOMETER READINGS DURING AN EARTHQUAKE

For the final practical application of the model presented in this chapter, we consider data collected for the Parkfield California earthquake of September 28, 2004: magnitude 6.0 on the Richter scale. Studies in the 1980s had predicted an earthquake for 1993 in this area near the San Andreas fault. As a result, a network of sensors was built and on December 20, 1994, an  $M = 5.0$  earthquake did occur (CSMIP, 2004; Borcherdt et al., 2004, 2006, and references within). We owe the data analyzed herein to the buildup of these array sensors.

The seismographic readings were programmed to trigger for a P wave (primary wave) of magnitude  $M > 3.0$ . The P wave is a compression wave analogous to sound waves. It is the fastest seismic of the seismic waves and consequently arrives first. The S waves (secondary waves) and the surface waves that are responsible for most of the destruction are slower. The Richter scale magnitude is a logarithmic (base 10) scale. This means that an  $M = 5.0$  event has 100 times the movement of a 3.0 event, a 6.0



**FIGURE 5.7** Three-hour volatility estimates for 1976–1977.



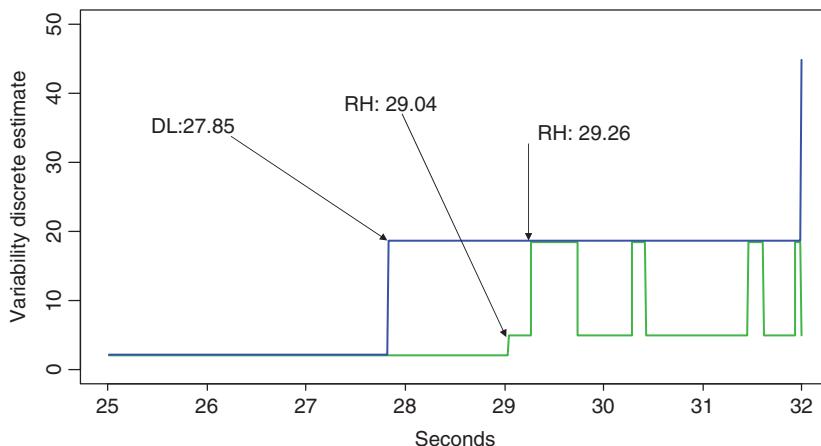
**FIGURE 5.8** Raw acceleration signal at the two stations.

event has 1000 times the movement of a 3.0 event. The signal used is the ground acceleration reading at 0.005-second intervals.

### 5.3.4 ANALYSIS OF THE EARTHQUAKE SIGNAL: BEGINNING

In the analysis we use data for the same earthquake (the 2004 Parkfield earthquake) obtained from two different research stations: Red Hills and Donna Lee stations. The two stations are in a mountainous region about at the same elevation and the distance between the stations is about 23.8 miles (or 38.3 km, distance calculated using their coordinates). Figure 5.8 presents the raw ground acceleration signal at the two stations. It is evident from the two pictures that Donna Lee was much closer to the epicenter. If we use the threshold arrival time (the time at which the raw acceleration passes magnitude 3) as the time at which the initial P wave energy arrived at each location, then the earthquake starts at Donna Lee at 17:27.849 and at Red Hills at 17:29.259.

We perform the variability estimates for a restricted region at the beginning of the signal and we plot the estimates obtained in Figure 5.9. While the Donna Lee signal presents a spike in variability at exactly the same time as the P wave is observed (29.85 s), for the Red Hills station there is a spike in variability 20 hundreds of a seconds earlier than the official earthquake starts. The second time in the image corresponds to the variability increase and this is remarkably close to the official P wave time

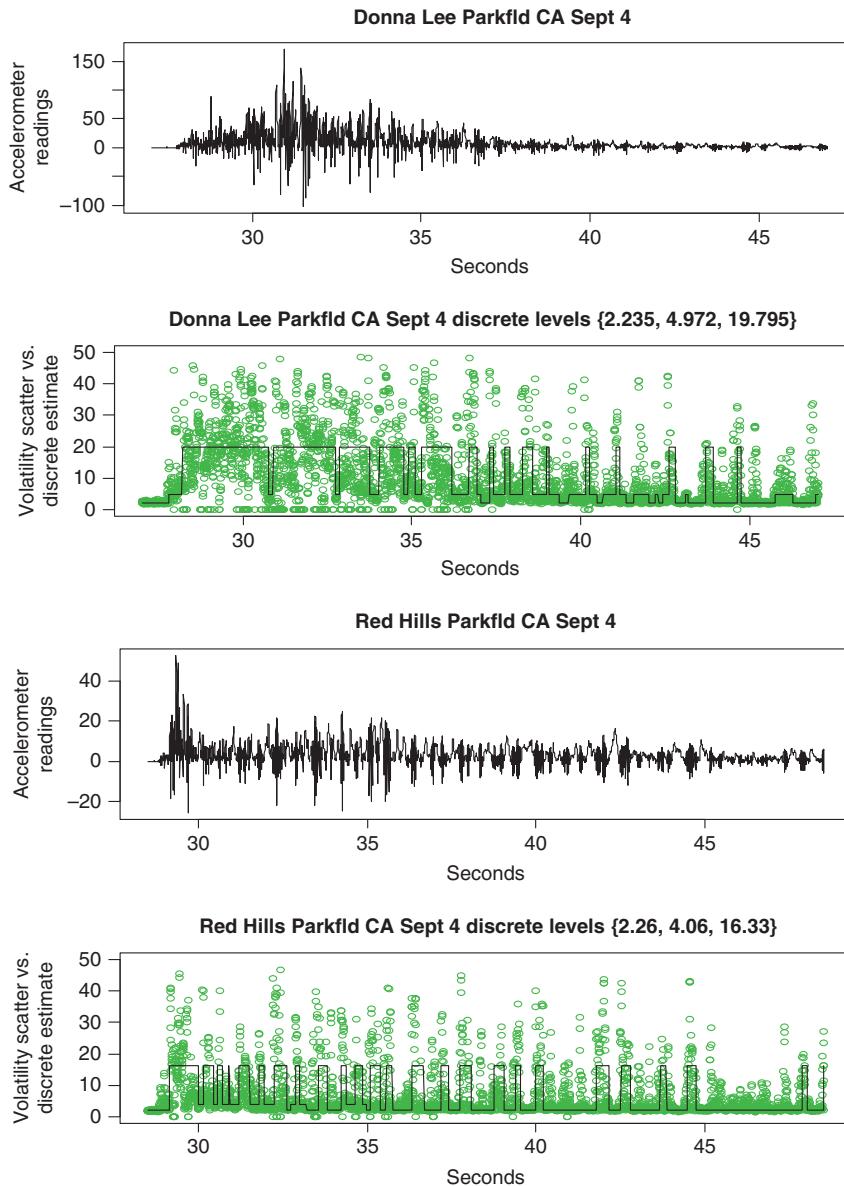


**FIGURE 5.9 Variability comparison. Beginning of an earthquake.**

(29.26 s). The difference may be due to the interference in signal traveling through the ground.

One of the methods for locating the actual origin or epicenter of the earthquake is by using the arrival times of the P waves. A simple triangulation method may be used if one knows the speed of travel of the waves through the ground and such speeds are charted and known. However, there are a number of factors affecting this speed that are not charted. Ground temperature and humidity, for instance, are factors that change during the day and their influence is unknown. Furthermore, the more dispersed the wave (further from the epicenter) the greater the interference (and influence) of these factors. Note that the signal at Red Hills does not look anything like the signal at a location closer to the epicenter. For this reason, we believe that the first time a shift in variance is detected may be an alternative way of providing triangulation results. The results obtained this way could be then compared with the more traditional methods either reinforcing or contradicting the results. In either case, we believe that the methodology may be useful.

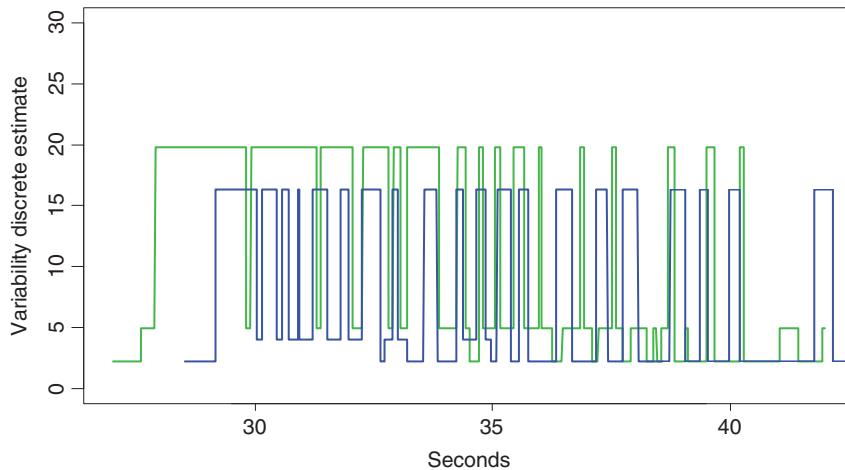
Calculating the focus or point of origin of the earthquake is very useful for the geophysical analysis and long-term estimates of earthquake hazard. Even though short-term prediction has not been successful, some correlation between smaller earthquakes and detectable activity shortly before an earthquake has been observed. The timing and amplitude of seismographic signals is also used for fundamental research about the earth's interior.



**FIGURE 5.10** Parkfield CA, Donna Lee Peak seismograph readings.

### 5.3.5 ANALYSIS: DURING THE EARTHQUAKE

In Figure 5.10, we plot the the variability analysis obtained using the entire acceleration signal at the two stations. Figure 5.11 superimposes the two estimates on the same timeline for the ease of comparison.



**FIGURE 5.11** Parkfield CA earthquake. Comparison of the variability signals, Donna Lee variability in green and Red Hills variability in blue.

We believe that the variability estimates may prove to be most useful for this period. First note that the shifts in variability are not evident from the accelerogram. People living through earthquakes will tell of the pulses that they feel during the earthquake. This shift in variability may indicate exactly that. Furthermore, by combining the magnitude and the frequency of these intervals, we may be able to measure the destructive power of an earthquake. This is done only approximately with the Richter scale. The Richter scale (adjusted logarithm base 10 of the highest perceived acceleration value at the epicenter), even though supposedly measuring the earthquake strength has, in fact, nothing to do with the actual destructive power. For example, the 2010 Haiti Earthquake, an extremely devastating earthquake (230,000 mortal cases, 3 million people affected), was recorded as 7.0 on the Richter scale. The only other recorded earthquake at 7.0 is the West Java earthquake of 2009—a very serious earthquake nonetheless one that cannot be compared with the severity of the Haiti earthquake (79 deaths displacing 210,000 people). We should also mention that Java is a much more densely populated region. The 2007 Tocopilla earthquake in Northern Chile recorded at 7.7 on the Richter scale caused two deaths and 150 minor injuries, 15,000 people were displaced. The economic damage of an earthquake is, in fact, measured by the Mercalli intensity scale, an entirely subjective scale that rates the earthquake from I—Instrumental (not felt) to XII—Cataclysmic. But, in fact the earthquake is rated, thus months

after it takes place and by looking at the total devastation it produces. On the contrary, looking at the estimated variability, we may be able to provide a much more accurate measure for this quantity immediately after the earthquake takes place.

Furthermore, if we look at the variability estimates (Figure 5.11) and the ground signal itself (Figure 5.8), it is pretty clear that this is the same earthquake even though one could not tell this from the signal itself. The variability patterns are parallel at the two locations. Given the relatively short distance between the two stations, one would expect the signal at Red Hills station to have a similar destructive power. This may not be the case based on this preliminary study.

### **5.3.6 ANALYSIS: END OF THE EARTHQUAKE SIGNAL, AFTERSHOCKS**

Figure 5.10 presents the variability estimates at the two stations. Note that after about 10 s at both locations, the predominant variability changes to a lower regime. The accelerometer signal shows lower values for the ground acceleration, but from that signal alone there does not seem to be a clear end point of the earthquake.

By looking at the variability on this last region, perhaps one could determine the probability of a future aftershock. An aftershock is a smaller earthquake that occurs after a previous large earthquake in the same area (the main shock). If an aftershock is larger than the main shock, the aftershock is redesignated as the main shock and the original main shock is redesignated as a foreshock. Aftershocks are mainly produced as the crust around the displaced fault plane adjusts to the effects of the main shock, and information about this perhaps may be gathered by looking at the variability pattern at the end of the earthquake signal.

---

## **5.4 Conclusion**

---

The work presents a methodology of estimating the unobservable variability of a signal. The method was used for three different areas and for each area provided insight into the working of the real-life signal. We believe that this is the main purpose of a model and that we have accomplished this purpose.

## Appendix 5.A: Theoretical results and empirical testing

In this section, we provide convergence results of the estimation algorithm presented in the chapter. The arguments presented are both: theoretical—by providing estimates for convergence rates, and empirical—by generating paths from a known process and comparing the estimates with the ground truth.

### 5.A.1 HOW DOES THE PARTICLE FILTER WORK?

To start this section, we note that the discrete process in (5.3) does not correspond to a discretization of the original continuous process in (5.2). Indeed, locally between any two times  $t_{i-1}, t_i$  the evolved process approximates the following continuous volatility process:

$$\begin{aligned} X_t &= X_{t_{i-1}} + \int_{t_{i-1}}^t \left( r - \frac{Y_u^2}{2} \right) du + \int_{t_{i-1}}^t Y_u dW_u \\ Y_t &= Y_{t_{i-1}} + \beta(Z_t - Z_{t_{i-1}}) \end{aligned} \quad (5.A.1)$$

The approximation in (5.3) of this auxiliary model is used to estimate the parameters of the Markov chain  $Y_t$ .

The convergence of the discretized filter  $Y_t$  in (5.3) to the continuous version from which it comes [equation (5.A.1)] is a well-studied problem. Del Moral, Jacod, and Protter [Del Moral et al. (2001); Del Moral (2004)] develop bounds on the expected error of the filter and the filter expected value. In the cited work, the authors work with a given observed process  $\{X_t \mid t \leq t_N\}$  a “noisy” measurement of  $\{Y_t \mid t \leq t_N\}$  the unobservable process or in our case the auxiliary volatility process.

Specifically, the following general situation is presented, where  $X_t$ ,  $Y_t$ ,  $a$ ,  $a'$ ,  $b$ , and  $b'$  are  $q$ -dimensional random vectors and valued functions.

$$\begin{cases} dX_t = a(X_t, Y_t)dt + b(X_t, Y_t)dW_t, \\ dY_t = a'(Y_t)dt + b'(Y_t)dZ_t \end{cases} \quad (5.A.2)$$

subject to the condition that the first and second derivatives of  $b$  and  $b'$  exist and are bounded. Furthermore,  $\|b\|^2$  and  $\|b'\|^2$  are not identically zero.

The distribution of the process being estimated  $\sigma(Y_t)$  is denoted with  $\Pi_N \sigma = \{\sigma_t \mid X_t, t \leq t_N\}$ . The output of the numerical filter at the  $N^{\text{th}}$

observation using  $n$  test paths is the estimated distribution of  $\sigma$  process, denoted with  $\hat{\Pi}_N^n \sigma = \{\hat{\sigma}_t \mid X_i, i \leq N\}$ . The work cited provides bounds for the expected error  $\mathbf{E}(\hat{\Pi}_N^n \sigma - \pi_N \sigma)$  with a function of  $N$ ,  $n$ , and  $\sigma$ .

The weighting function used in the general filtering algorithm  $\phi$  may be any arbitrary function integrating to 1 and the weights are assigned using:

$$\phi_n(v) = C(N) \phi(vn^{\frac{1}{2+q}}) \quad (5.A.3)$$

which is identical with what we use in our more specific algorithm ( $q = 1$ ), while  $n$  is the number of generated test paths and  $N$  is the total number of observations.

The cited results prove two convergence results of the filter to the auxiliary model as  $n \rightarrow \infty$ . Specifically, when  $q = 1$ :

$$\mathbf{E} [\hat{\Pi}_N^n \sigma - \pi_N \sigma] < \frac{C\sigma}{n^{\frac{1}{3}}} \quad (5.A.4)$$

$$\mathbf{P} (|\hat{\Pi}_N^n \sigma - \pi_N \sigma| \geq \delta) \leq Ce^{-n^{\frac{2}{3}} \frac{\delta}{(C\sigma)^2}} \quad (5.A.5)$$

as  $n \rightarrow \infty$  and  $C = C(N)$ , where  $n$  is the number of trial paths and  $N$  is the number of observation points.

In our case, the convergence proof needs to be modified because of the presence of the auxiliary process. The idea in our case is to show that the auxiliary process converges to the discretization of the true process generated from the Markov chain volatility. The only theoretical case considered is when the weighting function is a Gaussian (we take advantage of the normal–normal conjugate distributions). In the empirical testing section (5.A.4), we look at the effects of many other weighting functions. They have little effect on the estimation.

We state only the relevant results in the Appendix. For a more thorough analysis and the proof of the results, we refer to the original thesis (Levin, 2010).

## 5.A.2 THEORETICAL RESULTS ABOUT CONVERGENCE AND PARAMETER ESTIMATES

The objective of this section is to derive some simple and applicable convergence results for the hypothesized system (equation 5.2) and the filter (equation 5.3) used to analyze it.

### 5.A.2.1 First pass: initial volatility estimates

For a process  $X$  with volatility a Markov chain  $Y$  as used in this chapter, we use the following notations. For each volatility state  $i$  at time  $t$  when the estimator is obtained:

- $\sigma_i$  is the “true” underlying volatility state.
- $N$  is the number of consecutive observations during one uninterrupted stay in the respective state, up until time  $t$ .
- The total number of observations at the respective state  $i$  (throughout all the observed data available at time  $t$ ) is denoted with  $Pts$ .
- The number of test paths the filter generates is denoted with  $n$ .
- $\tilde{\sigma}_i$  is volatility of the discrete state system.
- $\sigma_{\text{est}}$  is the filter estimated volatility (which uses a continuous state space approximation).
- $\sigma_{\text{bound}}$  is the theoretical point at which extreme test values for  $\tilde{\sigma}_2$  are ignored or truncated. Its value is adapted to the filter performance and decreases as the filter converges.
- $\sigma_\phi$  is the standard deviation of the penalty function and is usually much smaller than  $\sigma_{\text{bound}}$  and  $\tilde{\sigma}_i$ .
- $\sigma_B$  is a derived factor in the theoretical description of the filter defined by:

$$\sigma_B^2 = \sigma_\phi^2 + \tilde{\sigma}_i^2 + \sigma_{\text{bound}}^2 \quad (5.A.6)$$

With these notations, the following result relates the mean of the estimator obtained using the auxiliary process with the mean of the estimator we would have obtained using the real process at a certain point in time.

**Theorem 5.1.** *For point volatility estimates after  $N$  points at a given state given  $\epsilon > 0$  there exists  $C$  large enough so that:*

$$P\left(|\tilde{\sigma}_i - \sigma_{\text{est}}| > C \sqrt{\frac{\sigma_B}{Nn\Delta t}}\right) < \epsilon \quad (5.A.7)$$

The next theorem relates the nodes estimated using all the data history.

**Theorem 5.2.** *For volatility node estimates after  $Pts$  total number observations at a given level and  $N$  observations per mean duration:*

$$P\left(|\tilde{\sigma}_i - \sigma_{\text{est}}| > C \times \sqrt{\frac{\sigma_B}{n \times Pts \times \Delta t} + \frac{\ln(N)}{N}}\right) < \epsilon \quad (5.A.8)$$

The next result relates the realized volatility of the discretized system to the parameter value.

**Theorem 5.3.**

$$P \left[ |\tilde{\sigma}_i - \sigma_i| > C \times \sqrt{\frac{2}{N} \sigma_i} \right] < \epsilon \quad (5.A.9)$$

For a two-state system  $\{a_1, a_2\}$  with transitions rates  $\lambda_1$  from state  $a_1$  to state  $a_2$  and  $\lambda_2$  the transition rate from state  $a_2$  to state  $a_1$ , we want to estimate the node levels  $\{a_1, a_2\}$  and the transition rates at each level. The node levels are determined from fitting a mixture of Gaussian distributions to the volatility estimates and using the means as node estimates. With enough points, the convergence specified in Theorems 5.1–5.3 allows us to distinguish and estimate nodes if the mean duration at the levels is long enough with enough transitions. From the convergence result for point estimates (ref to equation 5.A.7) we see an increase in accuracy as the number of paths and the number of observations per duration at a given level increase. This holds even more strongly for the node estimates (refer to equation 5.A.8) and also increases as the total number of points at a given level increases.

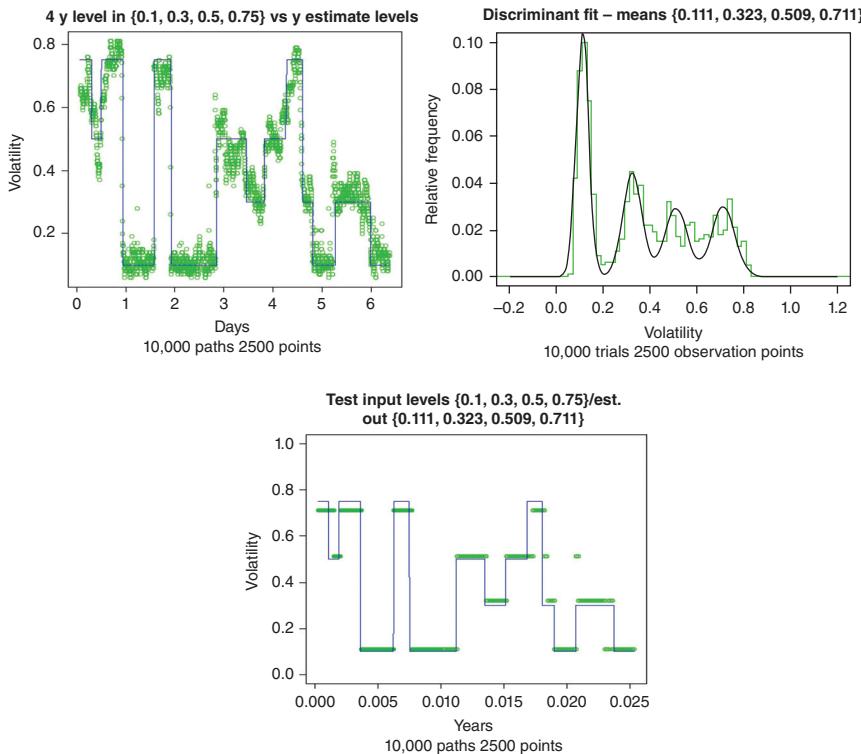
### 5.A.3 MARKOV CHAIN PARAMETER ESTIMATES

The conditional transition probability from state  $i$  to state  $j$  is given by:

$$p_{i,j} = \frac{\lambda_{i,j}}{\sum_{j=1}^n \lambda_{i,j}} \quad (5.A.10)$$

This formula allows estimates for transition probabilities to be computed using the empirical transition rates. Confidence intervals for these transition rates for individual cases may be found using the binomial approximation (for two states) or a multinomial distribution or Pearson's  $\chi^2$  test with more states.

In practice, the larger the window or frame of points used, the greater the percentage of correct estimates. One problem we encountered is the way in which isolated errors affect the estimates for mean duration at a certain state. If the algorithm erroneously detects a jump in volatility in the middle of a stay at a certain state, this effectively halves the duration time at the state. With large enough data, the problem goes away. However, this usually has rather absolute practical limitations. The quantitative relationship of probability of isolated errors to accuracy of mean durations estimates seems to be very complicated.



**FIGURE 5.12 Test data: 4 Nodes (0.1, 0.3, 0.5, and 0.75).**

#### 5.A.4 EMPIRICAL TESTING

In this section, we generate observations, using a known Markov chain (ground truth). Then we perform the estimation procedure and we compare the estimates with the known values.

We performed simulation results for 2, 3, and 4 node Markov chains. We simulated 2000 observation points (which translates into about a week of minute financial data), 8–10 thousand simulation points per path and about 50–200 observations per state duration. In Figure 5.12, we present the algorithm output for a 4-node estimation. The first image plots the real volatility (black line) and the estimated means (green points). The second image presents the histogram of the means (green) and the fitted gaussian mixture distribution on top. Finally, the third image displays the real volatility (black line) and the estimated points restricted to the node values. In general, the algorithm needs several observations at a new level,

before it detects a volatility shift. For the results obtaining extra test data, we refer the reader to the supplementary document Supplement A.

### 5.A.5 A LIST OF SUPPLEMENTARY DOCUMENTS

Because of the page constraint of this document, we were forced to leave out many results and images that we felt would have enhanced the work and demonstrated the extensive testing put into this work. Thus we have added three documents that are visible to the links provided as well as on the journal site.

**Supplement A: Selected test results using 2, 3, and 4 nodes** This pdf file presents test results using data generated from a process with known parameter values. The values we chose resemble the numerical values, frequencies, and distances between nodes as observed in our applications. Some values were chosen because of their difficulty in estimation (nodes close to each other) to illustrate the power of the algorithm.  
<http://www.math.stevens.edu/~ifloresc/SupplementA.pdf>

**Supplement B: Other stock volatility analysis** This pdf file presents volatility runs obtained from minute stock data during the week of the Bear Stearns collapse. We use financial stocks and commodity stock to illustrate the differences in the volatility patterns.  
<http://www.math.stevens.edu/~ifloresc/SupplementBStocks.pdf>

**Supplement C: Temperature variability results** This pdf file presents volatility runs obtained from the climate data in Central Park, New York City, from 2000 to 2010.  
<http://www.math.stevens.edu/~ifloresc/SupplementCClimate.pdf>

## References

- Aït-Sahalia, Y., and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics* 83, 413–452.
- Barndorff-Nielsen, O. E., and N. Shephard (2002). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society B* 64.
- Black, F., and M. Scholes (1973). The valuation of options and corporate liability. *Journal of Political Economy* 81, 637–654.
- Bladt, M., and M. Sørensen (2007). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Preprint.

- Borcherdt, R. D., M. J. S. Johnston, G. Glassmoyer, and C. Dietel (2004). Recordings of the 2004 Parkfield earthquake on the general earthquake observation system array. Technical report, U.S. Geological Survey. Open-file Report 2004-1376, October 11, 2004.
- Borcherdt, R. D., M. J. S. Johnston, G. Glassmoyer, and C. Dietel (2006, June). Recordings of the 2004 Parkfield earthquake on the general earthquake observation system array. *Bulletin of the Seismological Society of America* 96(4B), S73–S89.
- Chib, S., M. K. Pitt, and N. Shephard (2004, August). Likelihood based inference for diffusion driven models. Economics Papers 2004-W20, Economics Group, Nuffield College, University of Oxford.
- Cho, S., R. Haralick, and S. Yi (1989). Improvement of Kittler and Illingworth's Minimum Error Thresholding. *Pattern Recognition* 22(5), 609–617.
- Cohan, W. D. (2009). *House of Cards*. Thorndike Press.
- Cont, R., and P. Tankov (2003). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC.
- CSMIP (2004). Strong-motion data from the Parkfield earthquake of 20 december 1994. Technical report, California Strong Motion Instrumentation Program, Division of Mines, Department of Conservation. Open-file Report 2004–1376, October 11, 2004.
- Cvitanic, J., R. Liptser, and B. Rozovskii (2006). A filtering approach to tracking volatility from prices observed at random times. *Annals of Applied Probability* 16(3), 1633–1652.
- Cvitanic, J., B. Rozovskii, and I. Zaliapin (2006). Numerical estimation of volatility values from discretely observed diffusion data. *The Journal of Computational Finance* 9(4), 1–36.
- Dash, M. (2008, September). Bank agrees to buy troubled loan giant for 4 billion. *New York Times*. 9-25-08 archive resource; [www.nytimes.com](http://www.nytimes.com).
- Del Moral, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer.
- Del Moral, P., J. Jacod, and P. Protter (2001). The Monte-Carlo method for filtering with discrete time observations. *Probability Theory and Related Fields* 120, 346–368.
- Florescu, I., and C. Tudor (2013, November). *Handbook of Probability*. John Wiley & Sons, Inc.

- Geithner, T. (2008). *The Current Financial Challenges: Policy and Regulatory Implications*. New York: Council on Foreign Relations Corporate Conference 2008. March 6: Remarks by Timothy Geithner, President and CEO of the Federal Reserve Bank of New York.
- Genon-Catalot, V., T. Jeantheau, and C. Larédo (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* 6(6), 1051–1079.
- Hamilton, J. D. (2005). Regime-switching models. *Palgrave Dictionary of Economics*.
- Hamilton, J. D., and G. Lin (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11, 573–593.
- Landon and Sorkin (2008, March). J.P. Morgan acts to buy ailing bear stearns at huge discount. *New York Times*. 3-15-08 archive resource; www.nytimes.com.
- Levin, F. (2010). *Monte Carlo estimation of stochastic volatility for stock values and potential applications to temperature and seismographic data*. PhD thesis, Stevens Institute of Technology, Castle Point on the Hudson, Hoboken, NJ, USA.
- Mariani, M., I. Florescu, M. B. Varela, and E. Ncheuguim (2009, April). Long correlations and levy models applied to the study of memory effects in high frequency (tick) data. *Physica A* 388(8), 1659–1664.
- Merton, R. C. (1992). *Continuous-Time Finance (Macroeconomics and Finance)*. Wiley-Blackwell.
- Nielsen, J. N., and M. Vestergaard (2000). Estimation in continuous-time stochastic volatility models using nonlinear filters. *International Journal of Theoretical and Applied Finance* 3(2), 279–308.
- Protter, P. E. (2005). *Stochastic Integration and Differential Equations* (2nd ed.). Springer.
- Ross Sorkin, A. (2008, March). JP Morgan pays 2 dollars a share for Bear Stearns. *New York Times*. 3-17-08 archive resource; www.nytimes.com.
- Shephard, N. (Ed.) (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press.
- Sorensen, H. (2003). Simulated likelihood approximations for stochastic volatility models. *Scandinavian Journal of Statistics* 30(2), 257–276.

Tsay, R. (2005). *Analysis of Financial Time Series* (2nd ed.). John Wiley & Sons, Inc.

U.S. Senate Committee on Banking, Housing and Urban Affairs (2008).

*Actions by the New York Fed in Response to Liquidity Pressures in Financial Markets.* U.S. Senate Committee on Banking, Housing and Urban Affairs. April 3: Testimony given by Timothy Geithner, President and CEO of the Federal Reserve Bank of New York.

## Chapter Six

# Detecting Jumps in High-Frequency Prices Under Stochastic Volatility: A Review and a Data-Driven Approach

Ping-Chen Tsai<sup>1</sup> and Mark B. Shackleton<sup>2</sup>

<sup>1</sup>*Department of Finance, Southern Taiwan University of Science and Technology, Yongkang, Tainan City, Taiwan.*

<sup>2</sup>*Department of Accounting and Finance, Lancaster University Management School, Lancaster, England, United Kingdom.*

### 6.1 Introduction

Jumps in prices have been shown to be an indispensable characteristic of asset returns from a modeling perspective,<sup>1</sup> or when volatility models are calibrated with option data.<sup>2</sup> Over the past decade, nonparametric methods afforded by the advent of high-frequency data have made the direct identification of jumps possible and uncovered many more jumps

---

<sup>1</sup>[38, 40].

<sup>2</sup>[10, 22, 36].

than are suggested by parametric models [2; 15; 16; 62; 63].<sup>3</sup> Many of these tests are, however, executed at daily level and therefore rejecting the null of no jumps in effect indicates that at least one jump has occurred during a trading day. The exact timing, direction, and size, that is, the empirical distributions of individual jumps are unknown from the daily test results.

When testing for simultaneous arrivals of jumps across different assets,<sup>4</sup> or for the correlation between jumps in price and in volatility,<sup>5</sup> it is desirable that jumps in prices can be identified up to an intraday interval. Assuming that in a finite time span there are only a finite number of jumps in prices<sup>6</sup> [7], and [58] devise jump tests to be implemented for high-frequency returns. Although the two studies propose similar test statistics, they differ in estimating the spot volatility of returns and in controlling for spurious detection due to multiple comparisons. In the study by Andersen et al. [7], hereafter ABD test, the bandwidth chosen to estimate spot volatility may include high-frequency returns before and after the interval under consideration, whereas in the study by Lee and Mykland [58], hereafter LM test, only the most recent high-frequency returns are used. To correct for spurious detection, the ABD and LM tests, respectively, consider the finite-sample distribution and the asymptotic distribution for the maxima of their test statistics to determine the rejection region.

We give a detailed review on the ABD and LM tests in Section 6.2. In particular, we discuss the validity of the smooth-volatility assumption made by the LM test. If volatility does not change abruptly over the neighborhood of the return interval being tested, the ABD and LM tests would give identical results, *ceteris paribus*. Despite being constantly used to prove asymptotic theorems [55; 61], the smooth-volatility assumption may not be empirically valid in the context of high-frequency jump detection.<sup>7</sup> In ABD (2007), it is pointed out that because of stochastic volatility, returns standardized by ex ante volatility measures are necessarily fat-tailed and so a *forward-looking* test is doomed to be biased toward overrejection. The

<sup>3</sup>Barndorff-Nielsen and Shephard [16, 51 and 2] find that more than 10% of days in their samples are with jumps, whereas in [40] only a couple of jumps are expected to occur per annum.

<sup>4</sup>[24, 45].

<sup>5</sup>[12; 53; 54; 56].

<sup>6</sup>See the discussion in [74].

<sup>7</sup>The assumption may also contradict with the jump-driven volatility models in References [13, 76].

ABD test is thus implemented in a *backward-looking* manner. In the two studies, however, the ABD test is moderately oversized whereas the LM test is undersized.<sup>8</sup> We find that this unexpected result on the LM test is due to a printing error in the original paper of Lee and Mykland [58], and show via simulation that a corrected version of the LM test indeed gives more spurious detections than the ABD test.

The printing error in the LM test is pertinent to the normalizing constants of the maxima of test statistics, which have a limiting Gumbel distribution under the null hypothesis of i.i.d. normal variables. Since both the ABD test and the corrected LM test present overrejection, their critical regions can be modified. We note that the Gumbel distribution can be interpreted as the generalized extreme value (GEV) distribution with a shape parameter  $\xi$  approaching zero [32]. Accordingly, by admitting a strictly positive shape parameter  $\xi$  in the GEV distribution, namely a GEV distribution of Fréchet type, the overrejection issue can be resolved. This innovated methodology represents our main contribution to the literature. The non-zero-shape parameter  $\xi$  can thus be seen as bias correction for high-frequency jump tests under stochastic volatility.<sup>9</sup>

To formalize these ideas, in Section 6.3 we devise a data-driven procedure that generalizes the ABD and LM tests in estimating spot volatility and in selecting critical regions. An intermediate volatility measure between those of the two tests is expected to possess superior size properties than the LM test. We study the performance of our tests under a volatility model that consists of a two-factor structure, a compound Poisson process for jumps in prices, and correlated diffusions between prices and volatility. The model is directly estimated from our data, which are high-frequency Spyder (SPY) returns during January 2002 and April 2010. Simulation shows that when the sizes of tests are controlled at a given nominal level (1%), the *backward-looking* test has the highest power in detecting jumps in prices.

In the simulation, the shape parameter  $\xi$  of GEV distribution is estimated between 0.075 and 0.089 for the *backward-looking* test to have correct sizes. The value of  $\xi$  has to increase for the *forward-looking* tests to show no overrejection. In the empirical analysis, we use the calibrated

---

<sup>8</sup>Both [39, 69] document this result.

<sup>9</sup>Breidt and Davis [30] show that the extreme value distribution has the same scaling and a slightly larger centering constant under a standard stochastic volatility model than for i.i.d. Gaussian variables.

values of  $\xi$  from simulation to detect jumps in SPY returns. For three sampling frequencies, our test identifies jumps with daily intensities from 6.12% at the 10-min frequency to 28.7% at the 2-min frequency. In particular, at the 10-min frequency the estimated daily intensity ( $>6\%$ ) is close to the results in [45], in which a daily jump intensity around 5.6% is recorded for SPY data at a 11-min sampling frequency.

Our empirical results indicate that jump arrivals tend to cluster at the higher sampling frequencies, and that the empirical distributions of identified jumps, after the embedded intraday volatility pattern (IVP) is removed, do not appear to be Gaussian. The chapter contributes to the literature by designing a framework in which the jump tests are approximately exempt from size distortion and thus the testing procedure with the highest detecting power can be determined. Moreover, the shape parameter  $\xi$  of GEV distribution for the maxima of test statistics can be interpreted as the degree of deviation from the null of i.i.d. Gaussian. Our work hence complements those in [28; 39], which modify and improve the jump tests by accounting for IVP and proposing a bootstrap rejection region. Closest in spirit to our study is Corsi et al. [28], who use a threshold function scaled by spot volatility to detect jumps, with the optimal scaling coefficient also determined by simulation.

This chapter is organized as follows. Section 6.2 reviews the intraday jump tests that are constructed using realized volatility measures. Section 6.3 describes the Spyder data and introduces a generalized testing procedure. Simulation study is performed in Section 6.4, and empirical results are presented in Section 6.5. Section 6.6 concludes.

## 6.2 Review on the intraday jump tests

### 6.2.1 REALIZED VOLATILITY MEASURE AND THE BNS TESTS

In financial economics, the no-arbitrage axiom requires that asset prices dynamics follow a semimartingale process.<sup>10</sup> Specifically, the efficient log-price process  $X_t$  defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  can be decomposed as:

$$X_t = X_0 + A_t + M_t, t \geq 0, \quad (6.1)$$

---

<sup>10</sup>[37, 70].

where  $A_t$  is of finite variation and  $M_t$  is a local martingale [15].<sup>11</sup> A special case is the Itô semimartingale with Poisson jumps in prices:

$$X_t = X_0 + \int_0^t a_u du + \int_0^t \sigma_u dW_u + \sum_{k=1}^{N(t)} J_k, \quad (6.2)$$

where  $a = (a_t)_{t \geq 0}$  is a locally bounded predictable drift function,  $\sigma = (\sigma_t)_{t \geq 0}$  an adapted càdlàg volatility process,  $W = (W_t)_{t \geq 0}$  a standard Brownian motion,  $N(t)$  a counting process, and  $J_k$  the magnitude of price jumps [31]. Given the specification in (6.2), the total variation of price process is measured by the quadratic variation QV [71]:

$$[X]_t = \text{plim}_{M \rightarrow \infty} \sum_{i=1}^M (X_{t_i} - X_{t_{i-1}})^2, \quad (6.3)$$

for a partition of time  $0 = t_0 < t_1 < \dots < t_M = t$ . The QV then consists of the continuous  $[X]^c_t$  and the discontinuous component  $[X]^d_t$ :

$$[X]_t = [X]^c_t + [X]^d_t = \int_0^t \sigma_u^2 du + \sum_{0 \leq u \leq t} J_u^2. \quad (6.4)$$

In (6.4), the first term is also known as the integrated volatility (IV) while the second term is the QV due to jumps. Given the definition of QV in (6.3), and assuming a frictionless market structure, it is natural to estimate the QV of prices from 0 to  $t$  by the sum of squared returns, known as the realized variance measure  $RV$  [6; 14]:

$$RV_t = \sum_{i=1}^{\lfloor t/\Delta \rfloor} (X_{t,i\Delta} - X_{t,(i-1)\Delta})^2, \quad (6.5)$$

where  $\Delta$  is the length of a high-frequency interval and  $\lfloor t/\Delta \rfloor$  is the integer part of  $t/\Delta$ . We follow the convention in the literature and set  $t = 1$ . In (6.5),  $X_{t,i\Delta} - X_{t,(i-1)\Delta} = r_{t,i\Delta}$  is thus the  $i$ -th high-frequency return with length  $\Delta$  on day  $t$ . Denote the number of high-frequency returns per day by  $M = \lfloor 1/\Delta \rfloor$ ; it is seen that as  $\Delta$  shrinks to 0 or equivalently as  $M \rightarrow \infty$ ,  $RV$  converges to the QV of price process.<sup>12</sup>

<sup>11</sup>A process is a local martingale if there is an increasing sequence of stopping times such that the stopped process is a martingale [66].

<sup>12</sup>Unlike the conventional long-span asymptotic theorems, the convergence here is *in-fill asymptotic*.

Barndorff-Nielsen and Shephard [15; 16] propose to estimate the continuous part of  $QV$ , using the co-called bipower variation (BP):

$$\text{BP}_t = \mu_1^{-2} \left( \frac{M}{M-1} \right) \sum_{i=2}^M |r_{t,i}| |r_{t,i-1}| = \frac{\pi}{2} \left( \frac{M}{M-1} \right) \sum_{i=2}^M |r_{t,i}| |r_{t,i-1}|, \quad (6.6)$$

where  $\mu_a = E(|Z|^a)$ ,  $Z \sim N(0, 1)$  and  $a > 0$ , that is,  $\mu_a = 2^{0.5a}\Gamma(0.5(a+1))/\sqrt{\pi}$ . Barndorff-Nielsen and Shephard [15; 16] show that BP is a consistent estimator of IV in the presence of jumps, as the probability of a single jump influencing both returns  $r_{t,i}$  and  $r_{t,i-1}$  is asymptotically zero—provided no consecutive jumps occurred. Therefore, an estimator for the jump part of  $QV$  is the difference  $\text{RV} - \text{BP}$ . A test statistic can be constructed for the null hypothesis of no jumps, given the bivariate distribution of  $\text{RV}$  and  $\text{BP}$ ; in [51], the following statistic is recommended:

$$Z_{\text{BNS},t} = \text{RJ}_t / \sqrt{0.609 \frac{1}{M} \max \left( 1, \frac{\text{TP}_t}{\text{BP}_t^2} \right)} \sim N(0, 1), \quad (6.7)$$

where  $\text{RJ}_t = (\text{RV}_t - \text{BP}_t)/\text{RV}_t$  and  $\text{TP}_t$  is the tripower quarticity estimator for the daily integrated quarticity. The null hypothesis of no jumps for day  $t$  is rejected if  $Z_{\text{BNS},t}$  is larger than the critical value from the standard normal distribution at some significance level. In the simulation study of Barndorff-Nielsen and Shephard [16], although the test has a mild degree of overrejection, its power in detecting jumps is relatively low.<sup>13</sup> Nevertheless, empirical results from References [16; 51] find many more jumps in prices than are suggested by the parametric models in [40]. Further theoretical results on BP and on multipower variation are given in [19; 78].<sup>14</sup>

## 6.2.2 THE ABD AND LM TESTS

To identify jumps in high-frequency prices, ABD (2007) considers a high-frequency return  $r_{t,j}$  and show that under the null of no jumps, the

<sup>13</sup>Around 20% at 5% level when there is a single jump per day accounting for 5% of total price variation.

<sup>14</sup>Jump tests constructed using realized measures include, among others, the threshold-based estimator of [62, 63], the ratio of power variations in [2], the threshold-attached BP in [34], the quantile-based estimator of [31], and the bootstrap technique in [67].

scaled high-frequency return  $\Delta^{-1/2}r_{t,j}$  has a vanishing mean and a variance approaching the integrated variance IV:

$$\lim_{\Delta \rightarrow 0} E[\Delta^{-1/2}r_{t,j}] = 0, \lim_{\Delta \rightarrow 0} \text{var}[\Delta^{-1/2}r_{t,j}] = \int_{t-1}^t \sigma_u^2 du, \quad (6.8)$$

where  $\Delta = 1/M$ . ABD (2007) suggests that although the exact distribution of  $\Delta^{-1/2}r_{t,j}$  is unknown, under the null of no jumps it is approximately Gaussian. Assuming that spot volatility is constant within a trading day, the ABD test detects a jump in prices if:

$$|r_{t,j}| / \sqrt{\text{BP}_t/M} > \Phi_{1-\beta/2}, \quad (6.9)$$

where  $\Phi_{1-\beta/2}$  is the critical value from a standard normal distribution for significance level  $\beta$ . Equivalently, under the null of no jumps the high-frequency return  $r_{t,j}$  standardized by its spot volatility  $\sqrt{\text{BP}_t/M}$  is a standard normal variable. The rationale of the ABD test is thus consistent with the comment in [17] that “... non-Gaussian increments and jumps are synonymous.”

In the ABD test, the spot volatility of return  $r_{t,j}$ , denoted by  $\sigma_{t,j}$ , is estimated by  $\sqrt{\text{BP}_t/M}$  and so the test cannot be performed until the end of trading hours when all intraday returns are realized. In [58], on the contrary, the detection of jumps is done on a real-time basis. For a high-frequency return  $r_{t,j}$ , its spot volatility  $\sigma_{t,j}$  is estimated using past returns up to interval  $(j-1)$  on day  $t$ :

$$\sigma_{t,j} = \sqrt{\frac{1}{k-2} \sum_{i=j-k+2}^{j-1} |r_{t,i}| |r_{t,i-1}|}. \quad (6.10)$$

Specifically, in (6.10), the first  $(k-1)$  returns from a local window of size  $k$  are used to estimate  $\sigma_{t,j}$ . The sum in (6.10) is proportional to  $\text{BP}_t$  and so the window size  $k$  should be sufficiently large such that the asymptotic property of BP being a consistent, jump-robust estimator of IV holds. From their simulation results, Lee & Mykland [58] suggest  $k = \sqrt{MT}$ , where  $T$  is the number of days in the sample. For a year with 250 trading days, sampling at 5-min frequency for a day with 24 trading hours (exchange rate data) would give  $k = 268$  (rounded to integer).

The ABD and LM tests define their test statistics as the absolute standardized return  $|r_{t,j}|/\sigma_{t,j}$ , with the spot volatility  $\sigma_{t,j}$  being estimated differently. To select a critical region, ABD (2007) considers for a daily

significance level  $\alpha$  the distribution of the maximal observation from  $M$  intraday statistics,  $\max_{j=1..M}(|r_{t,j}|/\sigma_{t,j})$ . Under the null hypothesis, each of the  $M$  statistics is independent standard normal, and hence the distribution function of  $\max_{j=1..M}(|r_{t,j}|/\sigma_{t,j})$  is given by the probability of  $M$  independent binomial trials with no success:

$$\Pr \left( \max_{j=1..M} \frac{|r_{t,j}|}{\sigma_{t,j}} \leq u \right) = \Pr (\text{Bin}(M, \beta) = 0) = (1 - \beta)^M, \quad (6.11)$$

where  $\text{Bin}(M, \beta)$  is the binomial variable with success probability  $\beta$ , which is also the intraday significance level, that is,  $\beta = 2(1 - F_\Phi(u))$ , with  $F_\Phi(\cdot)$  the standard normal CDF. The all-acceptance probability  $(1 - \beta)^M$  in (6.11) is thus equal to  $1 - \alpha$  and so:

$$\beta = 1 - (1 - \alpha)^{1/M}. \quad (6.12)$$

Rasmussen [69] and Boudt et al. [28] point out that the adjustment between  $\alpha$  and  $\beta$  in (6.12) is the Šidák correction for tests with multiple comparison. A common alternative is the Bonferroni procedure, which gives  $\beta = \alpha/M$  [1]. In their simulation study, ABD (2007) finds that their test is moderately oversized, with an empirical size of 0.153% for 2-min returns given a 0.1% level. As a result, a more effective correction is needed.<sup>15</sup>

To solve the overrejection problem due to multiple testing, Lee and Mykland [58] consider the asymptotic distribution for the maximum of test statistics from a sample of size  $n = MT$ . Under the null of no jumps, with  $\sigma_{t,j}$  given in (6.10) the LM test statistics are proportional to standard normal and are asymptotically independent. Denote the largest observation of statistics by  $M_n$ , Lee and Mykland [58] use extreme value theory to show that there are two constants  $c_n$  and  $s_n$  such that  $M_n$  normalized by these constants has a limiting Gumbel distribution, as  $n \rightarrow \infty$ :

$$\Pr\{(M_n - c_n)/s_n \leq u\} \sim \text{Gumbel}(u) = \exp(-\exp(-u)), \quad (6.13)$$

where

$$\begin{aligned} c_n &= \sqrt{2 \log(n)}/c - [\log(\pi) + \log(\log(n))] / \left(2c\sqrt{2\log(n)}\right), \\ s_n &= 1 / \left(c\sqrt{2\log(n)}\right), \end{aligned} \quad (6.14)$$

<sup>15</sup>One relevant study is that of Bajgrowicz and Scaillet [9], who treat the problem as controlling the false discovery rate in multiple testing suggested by Benjamini and Hochberg [23].

**TABLE 6.1 Simulation on Šidák correction and Gumbel critical values, 1% level.**

Frequency	2 min	5 min	10 min
$M$	195	78	39
Šidák correction	0.0105	0.0104	0.0099
Gumbel: $c_n$ in (6.14)	0.0047	0.0039	0.0030
Gumbel: $c_n$ in (6.16)	0.0121	0.0103	0.0084

Table 6.1 reports the average rejection rates of Šidák correction and Gumbel critical values from 1000 samples of i.i.d.  $N(0, 1)$  variables at a 1% daily level  $\alpha$ . Each sample is of length 1 year, that is, 250 trading days; each day with 6.5 trading hours is sampled every 2, 5, and 10 min and thus  $M$  is 195, 78, and 39. The Šidák correction has standard normal critical value  $\Phi_{1-\beta/2}$ , where  $\beta = 1 - (1 - \alpha)^{1/M}$ . The Gumbel critical value  $-\log(-\log(1 - \alpha))$  in (6.15) is evaluated with  $c_n$  from (6.14) and (6.16), respectively.

in which  $c = \sqrt{2/\pi}$ . The LM test detects a jump in  $r_{t,j}$  if:

$$\left( \frac{|r_{t,j}|}{\sigma_{t,j}} - c_n \right) / s_n > -\log(-\log(1 - \alpha)), \quad (6.15)$$

given a significance level  $\alpha$ . In other words, if the test statistic after centered and scaled by  $c_n$  and  $s_n$  is larger than the critical value of Gumbel distribution, which is the asymptotic distribution for the largest observation from  $n$  i.i.d. Gaussian variables, we have confidence in rejecting the null of no jumps in  $r_{t,j}$ .

Here we note that in (6.14), the equation for  $c_n$  contains a printing error, and the correct expression is given in [3; p. 46]<sup>16</sup>:

$$c_n = \sqrt{2 \log(n)} - [\log(4\pi) + \log(\log(n))] / (2\sqrt{2\log(n)}), \quad (6.16)$$

Importantly, the missing factor “4” in  $c_n$  explains the very conservative, undersized results in [58]. To quantify the impact of this factor, we simulate 1000 samples of i.i.d.  $N(0, 1)$  variables and calculate over these samples the average rejection rate given a 1% daily level. Each sample is of length 1 year, that is, 250 trading days; each day with 6.5 trading hours is sampled every 2, 5, and 10 min and thus  $M$  is 195, 78, and 39. To permit direct comparison between the ABD and LM tests, we let  $k = M = n$ , that is, both tests estimate spot volatility with the same bandwidth  $M$ . In Table 6.1, the reported result suggests that the Šidák correction used in the ABD test

<sup>16</sup>See also [35, 43, 46, 77].

performs well. The correct expression of  $c_n$  gives rejection rate close to the nominal 1% level at the 5-min frequency; on the contrary, the Gumbel critical values with  $c_n$  in (6.14) are severely undersized across three frequencies.

We conclude this section by summarizing the many empirical works that have implemented the two tests. Table 6.2 lists some of these studies; it can be seen that much of attention has been paid to the effect of microstructure noise on the tests. This issue is either mitigated by sampling at a safe frequency or directly accounted for using noise-robust volatility measures. For example, Bos et al. [27] use the preaveraging estimator of Jacod et al. [52] to obtain consistent estimates of IV in the presence of noise and jumps. On the other hand, Boudt et al. [28] show that it is necessary for both the ABD and LM tests to be adjusted for the IVP to avoid size distortion. In terms of rejection region, most studies simply follow those given by the ABD and LM tests.

## 6.3 A data-driven testing procedure

### 6.3.1 SPY DATA AND MICROSTRUCTURE NOISE

The data studied in this chapter are the traded prices of SPY, an exchange traded fund designed to represent ownership of the S&P 500 Index. SPY trading has a high liquidity, which avoids sampling errors at high frequency and reduces the stale price effect [60]. SPY is shown to have significant contribution to the price discovery of the S&P 500 Index [50]. Many studies also analyze high-frequency SPY data.<sup>17</sup> From the NYSE TAQ database, we download the traded prices of SPY during the trading hours 09:30–16:00 EST from January 2, 2002, to April 30, 2010. The cleaning procedure described in [21] is used to remove incomplete trading days and unusual price records, leaving us a total of 2076 days of SPY price records.

The efficient price process observed at ultra-high frequency is subject to various forms of distortion caused by the structure of market, such as price discreteness and bid-ask bounce effect [11; 47]. Consequently, realized volatility measures constructed from returns sampled at or above some frequencies may be contaminated by these microstructure noises. In the literature, there are three main approaches in dealing with the effect

<sup>17</sup>[45, 48, 64].

**TABLE 6.2 List of studies on ABD and LM tests.**

Studies on intraday jump tests	Test performed	Microstructure noise	IVP	Data & overnight period	Critical region selection
[27]	LM ABD, LM	Preaveraged BP 5-min return	Adjusted Adjusted#	FX data 24h FX data excluding weekend	same as LM same as LM
[28]					
[39]	ABD, LM	Simulated	Noted/unadjusted	Equity excluding overnight period	bootstrap approximation
[41]	ABD	Staggered 5-min return	Noted/unadjusted	Futures data from 08:20 EST	same as ABD
[45]	ABD, LM	Simulated; 11-min return	Same as #	Equity data from 09:35 EST	Controlled at nominal level
[57]	ABD, LM	15-min return	Same as #	FX 24h; futures from 08:30 EST	same as LM
[59]	LM	15-min return	Controlled after detection	Equity including overnight period	same as LM
[69]	LM	Modeled and corrected	Unadjusted	Equity data	same as ABD
[75]	LM	Simulated; staggered return	Adjusted	Equity, futures & FX w/o overnight	same as LM

Table 6.2 lists nine studies that implement or modified the ABD and LM tests and summarize how issues such as microstructure noise, IVP, overnight period, and selection of critical region are addressed in these papers.

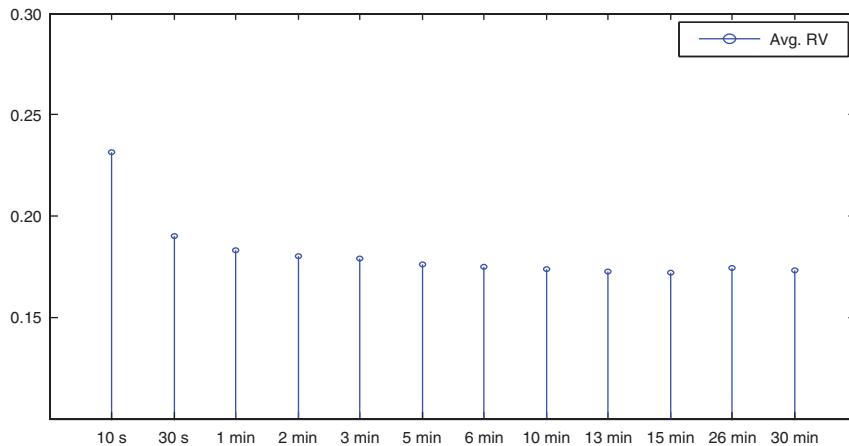


FIGURE 6.1 Volatility signature plot of RV, SPY.

of noises<sup>18</sup>; in this study, we follow the conventional approach of seeking a frequency at which RV is not influenced by noises [5]. In Figure 6.1, we plot RV (averaged over the data period) against sampling frequency; it can be seen that there is an upward bias in RV above the 1-min frequency. Accordingly, we choose to sample the SPY data at the 2-, 5-, and 10-min frequencies [65] and ABD (2007) also sample the S&P 500 Index and Index futures at the 2-min frequency.

Figure 6.1 plots the average RV of SPY in annual standard deviation against the sampling frequency of intraday prices. The average RV over our sample period increases above the 1-min frequency.

We summarize some empirical aspects of SPY data. First, the intraday SPY returns have a typical leptokurtic distribution and become more so at the higher frequencies [73]. A diurnal effect in the autocorrelations (ACFs) of absolute returns, equivalently an IVP, is visible with a peak at 10:00 EST [24]. Second, for daily volatility measures RV and BP, both of their logarithms are close to Gaussian but with fatter right tails. The slow decay of ACFs for log volatility measures is evident, yet augmented Dickey–Fuller test rejects the null hypothesis of unit root [8]. The XCFs of RV and BP with respect to daily open-to-close returns confirm a significant leverage effect. Third, consistent with [18], the jump component  $RV - BP$  has a

<sup>18</sup>The three approaches are the two-scaled realized variance of the study [79], the realized kernel estimator of the study [20], and the preaveraging estimator of the study [52].

modest degree of serial correlation. The XCFs of  $RV - BP$  with respect to open-to-close returns, on the contrary, do not suggest any signs for leverage effect [25]. For the ratio  $RJ = (RV - BP)/RV$ , our estimates range between 7.1% and 8.7% and are comparable to the estimate 7% in [51]. Fourth, we find that daily open-to-close returns standardized by  $RV$  or  $BP$  are approximately Gaussian but have thin tails, an observation also documented in [42]. Further details are given in a report available upon request.

### 6.3.2 A GENERALIZED TESTING PROCEDURE

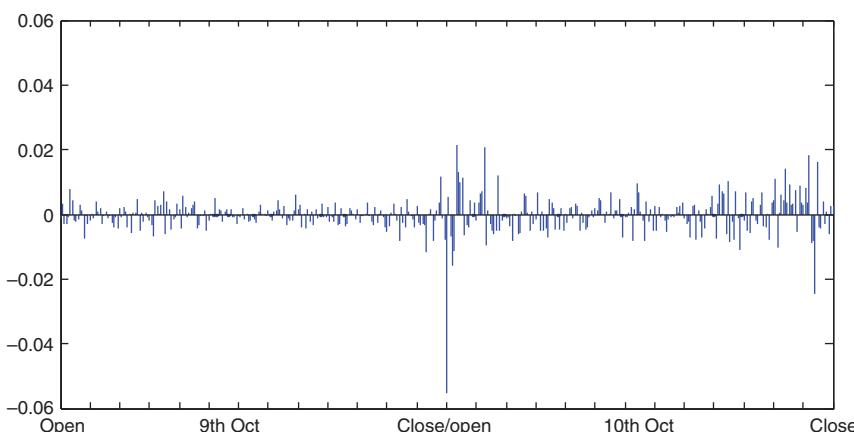
#### 6.3.2.1 Spot volatility estimation

As discussed in the “Introduction” section, the ABD test detects jumps retrospectively (ex post) whereas the LM test does it in a *forward-looking* manner (ex ante). The two tests would give identical results if the smooth-volatility assumption holds:

Figure 6.2 plots the 2-min returns of SPY during the 2-day period from October 9 to 10, 2008, with the overnight period posts a negative return of  $-5.54\%$ .

$$\sup_i \sup_{t_i \leq u \leq t_{i+1}} |\sigma(u) - \sigma(t_i)| = O_p(\Delta t^{0.5-\varepsilon}), \quad (6.17)$$

where  $\varepsilon > 0$ ,  $\Delta t = t_{i+1} - t_i$  (Assumption 1; [58]). For the studies in Table 6.2, few question the empirical validity of (6.17). The presence of overnight and weekend periods, however, could potentially undermine this assumption. As an illustration, in Figure 6.2, we plot the 2-min returns



**FIGURE 6.2** Two-min SPY returns on October 9 and 10, 2008.

**TABLE 6.3** Contingency table of overnight returns w.r.t. changes in BP, SPY.

Frequency	2-min		5-min		10-min	
	$r_{t-}^{\text{on}} < 0$	$r_{t-}^{\text{on}} > 0$	$r_{t-}^{\text{on}} < 0$	$r_{t-}^{\text{on}} > 0$	$r_{t-}^{\text{on}} < 0$	$r_{t-}^{\text{on}} > 0$
$\text{BP}_t < \text{BP}_{t-1}$	396	621	406	632	428	638
$\text{BP}_t > \text{BP}_{t-1}$	539	498	529	487	507	481
$\chi^2_{(3)}$ test	35.20	(1.1E-7)	34.84	(1.3E-7)	25.78	(1.1E-5)

Table 6.3 reports the contingency tables for the association between overnight returns and changes in BP. The overnight returns  $r_{t-}^{\text{on}}$  refer to those before day  $t$ . Twenty overnight returns of 0 are excluded in the table. Across three frequencies,  $\chi^2$  test statistics indicate that the association is significant, with  $p$  values in the parenthesis smaller than 0.1%.

of SPY during the 9th and 10th of October 2008. It can be seen that the overnight period posts a large negative return ( $-5.54\%$ ) and the data is much more volatile on the second day than the first day. Indeed, BP on October 10, 2008, is estimated to be at least 2.5 times larger than the previous day.<sup>19</sup> The 2-day period in Figure 6.2 hence represents an example in which spot volatility changes (increases) unexpectedly, with the overnight period being the change point. In Table 6.3, we construct contingency tables of overnight returns with respect to changes in BP and find that the null of no association is soundly rejected, providing further evidence against the smooth volatility assumption in (6.17).

The observation in Table 6.3 that BP tends to increase following negative overnight returns reinforces the leverage effect of volatility. Therefore, to obtain a suitable ex ante volatility measure  $E[\text{BP}_t | I_{t-1}]$ , we consider and modify the HAR model of Corsi [33] as:

$$\begin{aligned} \log(\text{BP})_t = & c + b_1 \log(\text{BP})_d + b_2 \log(\text{BP})_w + b_3 \log(\text{BP})_m + b_4 \log(\text{BP})_q \\ & + b_5 r_d^{\text{oc}} + b_6 r_w^{\text{oc}} + b_7 r_m^{\text{oc}} + b_8 r_d^{\text{on}} + b_9 r_w^{\text{on}} + b_{10} r_m^{\text{on}} + u_t \end{aligned} \quad (6.18)$$

$$u_t = e_t + \varphi_1 e_{t-1} + \varphi_2 e_{t-2}, e_t \sim N(0, \sigma_e^2). \quad (6.19)$$

In (6.18),  $\log(\text{BP})_l$ ,  $l = d, w, m, q$ , are the past daily, weekly, monthly, and quarterly log of BP, with  $\log(\text{BP})_w = (1/5) \sum_{i=1}^5 \log(\text{BP})_{t-i}$ , and  $r_l^{\text{oc}}$  and  $r_l^{\text{on}}$  are the corresponding variables for past open-to-close and overnight

<sup>19</sup>This ratio is estimated using 5-minute returns. At 2- and 10-minute frequencies the ratios are 4.8 and 4.3.

**TABLE 6.4 Standardized returns with different information flow of BP, SPY.**

Frequency	$r_t^{\text{oc}} / \sqrt{\text{BP}_{t-1}}$		$r_t^{\text{oc}} / \sqrt{E[\text{BP}_t   I_{t-1}]}$		$r_t^{\text{oc}} / \sqrt{\text{BP}_t}$	
	2 min	5 min	2 min	5 min	2 min	5 min
Mean	-0.0067	-0.0084	-0.0078	-0.0087	0.0486	0.0511
SD	1.1321	1.1853	1.0650	1.0960	1.0632	1.0937
Skewness	-0.4292	-0.4313	-0.3609	-0.3571	0.0183	0.0002
Kurtosis	4.3609	4.4987	3.6492	3.6468	2.7744	2.7045
JB Test	223.82	258.51	81.53	80.32	4.5174	7.5525
p value	<0.1%	<0.1%	<0.1%	<0.1%	0.1007*	0.0243

Table 6.4 reports the summary statistics of daily open-to-close returns  $r_t^{\text{oc}}$  of SPY standardized by different information flow of BP, with  $E[\text{BP}_t | I_{t-1}]$  obtained from the HAR model in (6.18) and (6.19). Note that in the calculation the sample means of  $r_t^{\text{oc}}$  are removed. The results for 10-min frequency are similar and thus are not reported to save space. The JB Test stands for Jarque–Bera test for normality.

returns. A residual analysis of  $u_t$  in Appendix 6.A leads to the MA (6.2) design in (6.19). The model thus has an approximate long memory structure with leverage effect.<sup>20</sup> In Table 6.4, we report the summary statistics of  $r_t^{\text{oc}}$  standardized by  $\text{BP}_{t-1}$ ,  $E[\text{BP}_t | I_{t-1}]$  and  $\text{BP}_t$ . The degree of nonnormality in  $r_t^{\text{oc}} / \sqrt{E[\text{BP}_t | I_{t-1}]}$  is evident but greatly reduced when compared with  $r_t^{\text{oc}} / \sqrt{\text{BP}_{t-1}}$ .

The fact that  $r_t^{\text{oc}} / \sqrt{E[\text{BP}_t | I_{t-1}]}$  is closer to normal than  $r_t^{\text{oc}} / \sqrt{\text{BP}_{t-1}}$  is relevant in jump detection. The leverage effect in volatility as captured by past returns in (6.18) is reflected in  $E[\text{BP}_t | I_{t-1}]$ , but this is ignored by the LM test in spot volatility estimation. In this study, we therefore introduce a third way in estimating spot volatility by interpolating  $E[\text{BP}_t | I_{t-1}]$  with realized high-frequency returns:

$$\begin{aligned} \text{Interpolated BP} = & \left( \frac{\pi}{2} \right) \left( \frac{M}{M-1} \right) \left\{ \sum_{i=2}^{j-1} |r_{t,i}| |r_{t,i-1}| \right. \\ & \left. + \left( \frac{2}{\pi} \right) \left( \frac{M-j+1}{M} \right) E[\text{BP}_t | I_{t-1}] \right\}, \quad (6.20) \end{aligned}$$

<sup>20</sup>The assumption that  $e_t \sim N(0, \sigma_e^2)$  is justified by the observation that  $\log(\text{BP})$  is close to normal. We adjust for Jensen's Inequality as in [68] when converting  $E[\log(\text{BP})_t | I_{t-1}]$  to  $E[\text{BP}_t | I_{t-1}]$ .

for  $3 \leq j \leq M$ . For the first two intervals in a day ( $j = 1, 2$ ), the spot volatility is estimated from  $E[\text{BP}_t | I_{t-1}]$ , because  $E[\text{BP}_t | I_{t-1}]$  is a less-biased measure than  $\text{BP}_{t-1}$  in obtaining spot volatility estimates. In the following, we will compare jump test results when the *forward-looking* (LM test), *backward-looking* (ABD test), and the *interpolated BP* are used to estimate spot volatility. It is obvious that the three measures converge as  $j \rightarrow M$ .

### 6.3.2.2 Selecting a critical region

In Table 6.1, we have seen that for i.i.d. normal variables, the Šidák correction and the Gumbel critical values give test results with correct sizes. For empirical data, the test statistics  $|r_{t,j}|/\sigma_{t,j}$  are likely to be dependent as spot volatility  $\sigma_{t,j}$  often involves estimation errors. In this case, the extreme value theory is still feasible, provided that the extreme observations are nearly independent or sufficiently distant in time [32]. This condition will be assumed to hold for our SPY data.

To identify a critical region that efficiently eradicates size distortion due to multiple comparisons and due to test statistics being dependent, we consider the GEV distribution, which nests the Gumbel distribution as a special case. Defined on the support  $\{u : 1 + \xi(u - \mathcal{L})/\mathcal{S} > 0\}$ , the GEV distribution is given as

$$\text{GEV}(u) = \exp \left\{ - \left[ 1 + \xi \left( \frac{u - \mathcal{L}}{\mathcal{S}} \right) \right]^{-1/\xi} \right\}, \quad (6.21)$$

where  $-\infty < \mathcal{L} < \infty$  is the location parameter,  $\mathcal{S} > 0$  is the scaled parameter, and  $-\infty < \xi < \infty$  is the shape parameter. The expression in (6.21) is given by the limiting distribution of the binomial variable in (6.11) as  $M \rightarrow \infty$ , that is, a Poisson distribution with intensity measure  $[1 + \xi((u - \mathcal{L})/\mathcal{S})]^{-1/\xi}$ . If  $\mathcal{L} = 0$ ,  $\mathcal{S} = 1$ , and  $\xi \rightarrow 0$ , the GEV distribution admits the Gumbel distribution in (6.13). As a result, the Gumbel critical value  $-\log(-\log(1 - \alpha))$  can be generalized by allowing a nonzero  $\xi$  to give the GEV critical value, and the decision rule (6.15) then becomes

$$\left( \frac{|r_{t,j}|}{\sigma_{t,j}} - c_n \right) / s_n > - (1/\xi) \left[ 1 - (-\log(1 - \alpha))^{-\xi} \right], \quad (6.22)$$

where  $c_n$  and  $s_n$  are given in (6.16). It can be shown that the GEV critical value is larger than the Gumbel value when  $\xi > 0$  and vice versa when

$\xi < 0$ . Since in ABD (2007) the *backward-looking* ABD test is moderately oversized, we expect a strictly positive  $\xi$  to be used in (6.22) to eliminate spurious detection.

Standard maximum likelihood method is available for the estimation of the GEV parameters including  $\xi$  [32]. However, as the data itself contains jumps, the estimated  $\xi$  from the maximums of test statistics  $|r_{t,j}|/\sigma_{t,j}$  cannot be directly used in the test.<sup>21</sup> What is needed is a  $\xi$  value such that the decision rule in (6.22) correctly rejects the null of no jumps given significance level  $\alpha$ . For this purpose, we therefore resort to simulation study to calibrate  $\xi$  at which the tests have correct sizes. As a result, the non-zero-shape parameter  $\xi$  in GEV distribution may be regarded as a measure of deviation from i.i.d. normal for the test statistics.<sup>22</sup>

## 6.4 Simulation study

### 6.4.1 MODEL SPECIFICATION

For the calibrated  $\xi$  to be credible, data must be simulated from a model that closely resembles the empirical SPY returns. From Section 6.3.1, such a model has a persistent but mean-reverting volatility, a jump component in prices, correlated return innovations with volatility, and a close-to-normal distribution for log volatility. We thus specify the model as:

$$\mathrm{d}\log(p(t)) = \left[ \mu_r - \frac{1}{2} V(t) \right] dt + \sqrt{V(t)} dW(t)^p + J(t)dN(t), \quad (6.23)$$

$$\log(V(t)) = \theta + \log(V_1(t)) + \log(V_2(t)), \quad (6.24)$$

$$\mathrm{d}\log(V_i(t)) = \kappa_i [-\log(V_i(t))] dt + \sigma_i dW(t)_i^v, \quad i = 1, 2. \quad (6.25)$$

In (6.23), the log-price process has drift  $\mu_r - V(t)/2$ , volatility  $\sqrt{V(t)}$  for the Brownian motion  $W(t)^p$  and jumps  $J(t)dN(t)$ , with Poisson arrivals  $N(t) \sim \text{Poi}(\lambda t)$  and normally distributed jump sizes  $J(t) \sim N(0, \sigma_j^2)$ . With a mean  $\theta$ , the log variance  $\log(V(t))$  in (6.24) has two components  $\log(V_1(t))$  and  $\log(V_2(t))$ , each following an Ornstein–Uhlenbeck (OU) process with  $\kappa_i$  the rates of mean reversion and  $\sigma_i$  the volatility for the Brownian motions

<sup>21</sup>In [26], the same point is taken and high-frequency jumps in prices are first identified before jump tails behavior is estimated.

<sup>22</sup>Our approach is related to Breidt and Davis [30], who show that for maximums from a stochastic volatility model, the Gumbel limit can be derived but with a slightly different centering constant.

$W(t)_i^v$  in (6.25). The two-factor structure has been widely used in the literature to describe the long-range dependence in volatility.<sup>23</sup> For generating the leverage effect in volatility, we require  $W(t)^p$  and  $W(t)_i^v$  to be negatively correlated:

$$dW(t)^p dW(t)_i^v = \rho_i dt, \quad \rho_i < 0, \quad i = 1, 2. \quad (6.26)$$

Given the two-factor structure, the jumps in prices and the two correlation coefficients  $\rho_i$ , the model is labeled SV2FJ\_2ρ and has 10 parameters:  $\{\mu_r, \theta, \kappa_1, \sigma_1, \kappa_2, \sigma_2, \lambda, \sigma_j, \rho_1, \rho_2\}$ .

To obtain the parameter estimates of the SV2FJ\_2ρ model, we first assume that  $\mu_r$  can be derived from the sample mean of open-to-close returns. For the parameters of the two-factor volatility structure  $\{\theta, \kappa_1, \sigma_1, \kappa_2, \sigma_2\}$ , we consider estimating an ARMA (2, 1) model for the daily observations of log(BP):

$$\begin{aligned} \log(\text{BP})_t = & \mu + \phi_1 (\log(\text{BP})_{t-1} - \mu) \\ & + \phi_2 (\log(\text{BP})_{t-2} - \mu) + \psi \varepsilon_{t-1} + \varepsilon_t, \end{aligned} \quad (6.27)$$

where  $(\phi_1, \phi_2)$  are the AR coefficients,  $\psi$  is the MA coefficient, and  $\varepsilon_t$  is serially uncorrelated with normal distribution  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ . It is well-established that an ARMA (2, 1) model can be decomposed into sum of two independent AR (1) processes if the two roots ( $\alpha_1$  and  $\alpha_2$ ) to the quadratic equation  $x^2\phi(x^{-1}) = 0$ , where  $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$  is the AR (2) polynomial, are both real-valued [49]. In this case,  $\alpha_1$  and  $\alpha_2$  are the AR coefficients of the individual AR (1) processes. In Appendix 6.B, we confirm that this condition indeed holds for log(BP) of SPY. Accordingly, we have the decomposition:

$$\log(\text{BP})_t = \mu + \log(V_1)_t + \log(V_2)_t, \quad (6.28)$$

where each of  $\log(V_i)_t$  is itself an AR (1) process:

$$\log(V_i)_t = \alpha_i (\log(V_i)_{t-1}) + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, \sigma_{\varepsilon,i}^2), \quad i = 1, 2. \quad (6.29)$$

Note that in (6.28), we follow Alizadeh et al. [4] and do not separate  $\mu$  between the two AR (1) processes. Hence, the discrete-time (6.28) and (6.29) have their continuous-time equivalents in (6.24) and (6.25), as an

---

<sup>23</sup>These studies include [4; 44; 68].

**TABLE 6.5** Estimation of ARMA (2, 1) and OU process parameters, log(BP) SPY.

	Sum of two AR (1) processes			Gaussian-OU processes		
	2 min	5 min	10 min	2 min	5 min	10 min
$\mu$	-9.6302	-9.7447	-9.7840	$\theta$	-9.6302	-9.7447
$\alpha_1$	0.9870	0.9860	0.9875	$\kappa_1$	0.0131	0.0141
$\sigma_{\varepsilon,1}$	0.1642	0.1703	0.1625	$\sigma_1$	0.1653	0.1715
$\alpha_2$	0.2413	0.2007	0.2012	$\kappa_2$	1.4217	1.6059
$\sigma_{\varepsilon,2}$	0.3445	0.3900	0.4419	$\sigma_2$	0.5986	0.7135

Table 6.5 reports the parameters of the OU processes in SV2FJ<sub>2ρ</sub> model, converted from the estimated ARMA (2, 1) model for log(BP) of SPY. The parameters reported are expressed for daily interval. The discrete-time parameters of ARMA (2, 1) model are converted into the continuous-time parameters of OU processes using the relation in (6.30). The details of ARMA (2, 1) estimation are given in Appendix 6.B. The mean-reversion parameters  $\kappa_1$  and  $\kappa_2$  correspond to half-life of around 52 and 0.5 days, respectively.

AR (1) process is the discrete-time counterpart of an OU process. We then have the conversion between the parameters:

$$\left( \mu, \alpha_i, \sigma_{\varepsilon,i}^2 / (1 - \alpha_i^2) \right) = (\theta, e^{-\kappa_i}, \sigma_i^2 / 2\kappa_i), \quad i = 1, 2. \quad (6.30)$$

In Table 6.5, we give the estimated discrete-time parameters and the corresponding values of  $\{\theta, \kappa_1, \sigma_1, \kappa_2, \sigma_2\}$ ; the continuous-time parameters in (6.30) are expressed for daily interval.

For the jump parameters  $\{\lambda, \sigma_j\}$ , since in (6.23) the Poisson arrival  $N(t)$  is independent from the zero-mean jump size  $J(t)$ , it can be shown that:

$$E[\text{RV}_t - \text{BP}_t] = E\left(\sum_{t-1 < s \leq t} J(s)^2\right) = E[N_t]E(J(s)^2) = E[N_t]\sigma_j^2. \quad (6.31)$$

Given the expected number of jumps per day  $E[N_t]$ , the value of  $\sigma_j$  can thus be calculated by setting  $E[\text{RV}_t - \text{BP}_t]$  equal to the sample mean of  $(\text{RV}_t - \text{BP}_t)$ . As a sensitivity analysis, we choose  $E[N_t]$  to be 0.08 and 0.40. Thus, when jumps are frequent, the variance of jump size becomes small, and vice versa, as shown in Table 6.6. For the two correlation parameters  $\{\rho_1, \rho_2\}$ , which determine the leverage effect, we define a quadratic loss

**TABLE 6.6** Parameter values of SV2FJ\_2 $\rho$  model, SPY.

Frequency	2 min	5 min	10 min
$\mu_r$	-0.0001	-0.0001	-0.0001
$\theta$	-9.6302	-9.7447	-9.7840
$\kappa_1$	0.0131	0.0141	0.0126
$\sigma_1$	0.1653	0.1715	0.1635
$\kappa_2$	1.4217	1.6059	1.6035
$\sigma_2$	0.5986	0.7135	0.8079
$\rho_1$	-0.4750	-0.5000	-0.5500
$\rho_2$	-0.1500	-0.1500	-0.1500
$E[RV_t - BP_t]$	9.1E-6	1.3E-5	1.6E-5
$E[N_t]$	0.08	0.08	0.08
$\sigma_j$	0.0107	0.0127	0.0141
$E[N_t]$	0.40	0.40	0.40
$\sigma_j$	0.0048	0.0057	0.0063

Table 6.6 collects the 10 parameter values of SV2FJ\_2 $\rho$  model in (6.23) to (6.26) calibrated from our SPY data. The parameter values are expressed for daily interval. Given the number of jumps per day  $E[N_t]$ , the jump size variance parameter  $\sigma_j$  is computed using  $E[RV_t - BP_t] = E[N_t]\sigma_j^2$  in (6.31); in the table the reported  $E[RV_t - BP_t]$  are sample means of positive ( $RV_t - BP_t$ ). The two correlation parameters  $\rho_i$ ,  $i = 1, 2$ , are obtained by minimizing the loss function  $\text{loss}(\rho_1, \rho_2)$  in (6.32).

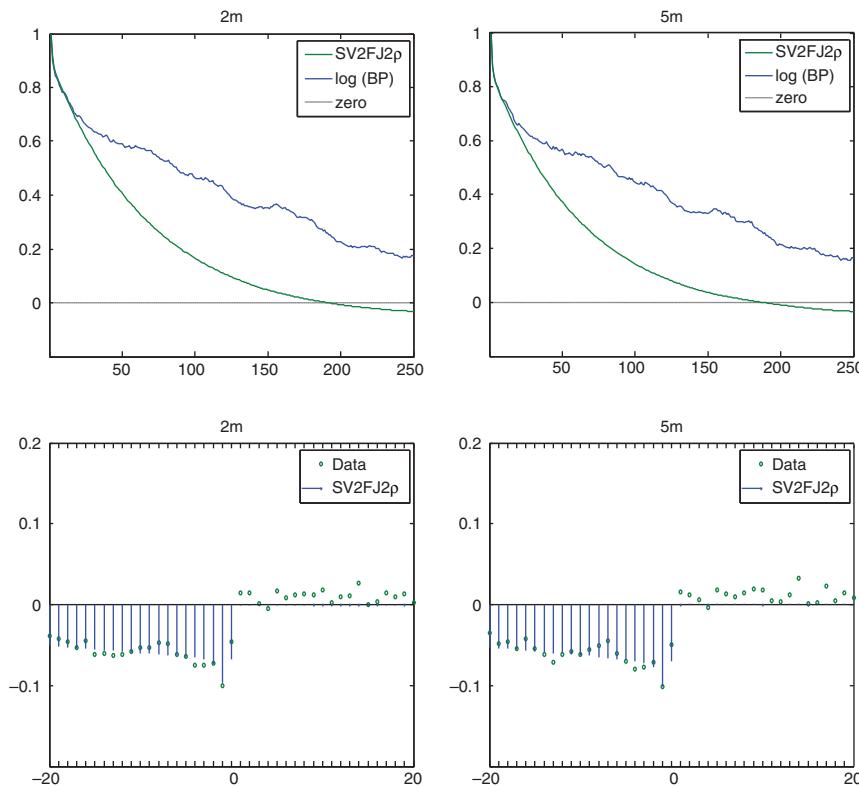
function in terms of the XCFs of  $\log(BP)$  with respect to past open-to-close returns:

$$\text{loss}(\rho_1, \rho_2) = \sum_{\text{lag}=0}^{-20} \left( XCF_{\log BP, \text{lag}} - XCF_{\log BP, \text{lag}}^{\text{sim.}} \right)^2, \quad (6.32)$$

where  $XCF_{\log BP}^{\text{sim.}}$  is the average XCFs of  $\log(BP)$  over 1000 simulations from SV2FJ\_2 $\rho$  model, in which  $\{\rho_1, \rho_2\}$  are selected from a grid of candidate values and the rest eight parameters are determined as described in Section 6.4.1.<sup>24</sup> Minimizing the loss function (6.32) then gives the optimal values for  $\{\rho_1, \rho_2\}$ ; see Appendix 6.C for details. Using this calibration procedure, we thus successfully obtain from our SPY data the 10 parameter values of SV2FJ\_2 $\rho$  model, which are presented in Table 6.6.

Figure 6.3 shows the ACFs, in the top, and XCFs (w.r.t. open-to-close returns), in the bottom, of  $\log(BP)$  for SPY and for the fitted SV2FJ\_2 $\rho$

<sup>24</sup>The leverage effect is driven by correlated diffusive movements and so jumps dynamics are irrelevant; we use  $E[N_t] = 0.40$  in constructing the loss function.



**FIGURE 6.3** The ACFs and XCFs of log(BP) from SPY and SV2FJ<sub>2ρ</sub> Model.

model. The simulated ACFs and XCFs are obtained as averages over 1000 simulations of the SV2FJ<sub>2ρ</sub> model using parameters listed in Table 6.6.

In Figure 6.3, we plot the averaged ACFs and XCFs (w.r.t. open-to-close returns) of log(BP) over 1000 simulations of SV2FJ<sub>2ρ</sub> model, together with their empirical counterparts of SPY. To save space, we only present the results when BP is constructed from the 2- and 5-min frequencies. At short lags the simulated ACFs approximate the empirical ones well, but at longer lags the simulated ACFs are not sufficiently persistent. The fit may be improved by estimating an ARMA (3, 2) model for log(BP) of SPY but at the cost of losing parsimony. On the other hand, in the bottom panel simulated XCFs of log(BP) match well with the empirical values, except at lag 1. The estimated values of  $\rho_1$ , which account for most of the leverage effect, center around  $-0.50$  and are similar to the estimates

in Eraker et al. [40].<sup>25</sup> We therefore consider the SV2FJ\_2 $\rho$  model and its parameter values a reasonable characterization of SPY data.

### 6.4.2 SIMULATION RESULTS

In the simulation, the length of data is 1275 days; after the first 25 days are discarded, the jump tests are applied to the last 250 days of simulated data. The in-sample period for the estimation of HAR model is therefore 1000 days. For three sampling frequencies and two jump dynamics, the SV2FJ\_2 $\rho$  model is simulated 1000 times and so the test size and power are obtained as averages over 1000 simulations. The jump tests are implemented at a 1% significance level; accordingly, the nominal size is 2.50 for a test period of 250 days. As in previous studies, we define test size as the number of false rejections and test power as the percentage of simulated jumps that are correctly identified.

We expect to observe the following outcomes in the simulation. First, under the same critical region, the *backward-looking* test should have the least overrejection, while the test using *interpolated BP* should outperform the *forward-looking* test. Second, we shall see that tests with Šidák correction and with Gumbel critical values are oversized. Hence, the shape parameter  $\xi$  of GEV distribution for the tests to have correct size should be strictly positive. Third, between the two scenarios of jump dynamics, the tests should have less power in detecting the smaller but frequent jumps. Fourth, a higher data frequency may enable higher testing power.

Table 6.7 reports the simulation results on the jump tests when  $E[N_t]$  is equal to 0.08. In the left two columns, the tests with Šidák correction are indeed seriously oversized, and as expected tests using *forward-looking BP* (middle panel) display the strongest degree of overrejection. The size distortion is more pronounced at the lower frequencies. In the central two columns, tests with Gumbel critical values show overrejection across three panels. Between the two corrections, tests with Gumbel critical values have more false detections at the higher frequencies than tests with Šidák correction, and vice versa at the lower frequencies. This observation is consistent with the results in Table 6.1. The right three columns of Table 6.7 report the calibrated values of  $\xi$  at which the tests have correct sizes. To obtain these values of  $\xi$ , we submit an array of candidate values to the

---

<sup>25</sup>The two AR (1) processes are not equally weighted; the persistent process, with  $\alpha_1$  above 0.986, accounts for nearly 90% of the total weight; see Appendix 6.B.

**TABLE 6.7** Simulation results of jump tests from SV2FJ\_2 $\rho$  model,  
 $E[N_t] = 0.08$ .

Frequency	Šidák correction		Gumbel c.v.		Correct size $\geq 2.50$		
	Size	Power	Size	Power	$\xi$	Size	Power
<i>Backward-looking BP</i>							
2 min	6.9920	79.32%	7.7340	79.48%	0.0890	2.4950	77.83%*
5 min	7.1980	75.61%	7.1300	75.59%	0.0870	2.4930	73.35%*
10 min	7.6700	70.07%	6.7680	69.67%	0.0840	2.5020	66.37%
<i>Forward-looking BP</i>							
2 min	15.8080	79.67%	17.2040	79.81%	0.1602	2.5070	76.70%
5 min	18.8430	76.97%	18.6920	76.96%	0.1782	2.4920	71.95%
10 min	24.1360	72.20%	21.9780	71.92%	0.2000	2.5100	65.02%
<i>Interpolated BP</i>							
2 min	11.7940	79.54%	12.8440	79.66%	0.1380	2.4750	76.96%
5 min	12.4430	76.41%	12.3230	76.38%	0.1410	2.4850	72.75%
10 min	13.7340	71.57%	12.4670	71.35%	0.1460	2.5050	66.61%*

Table 6.7 reports the simulation results of jump tests under the SV2FJ\_2 $\rho$  model with  $E[N_t] = 0.08$ . The tests are applied to the last 250 days in the simulated data at 1% level, giving a nominal size of 2.50. The top, middle, and bottom panels are the results when the *backward-looking*, *forward-looking*, and *interpolated* BP are used to estimate spot volatility. Each panel has results obtained with Šidák correction, with Gumbel critical value and when empirical sizes are held at around 2.50. The ABD test corresponds to the top-left block of the table, while the LM test is in the middle-middle block. The highest powers when empirical sizes are held at around 2.50 are marked with “\*”.

decision rule (6.22) and interpolate to identify  $\xi$  such that the test sizes are close to 2.50. The calibration process is tabulated in Appendix 6.D. The calibrated values of  $\xi$  are all positive, consistent with our conjecture that solving the oversized problem requires a strictly positive  $\xi$  in the GEV distribution. Moreover, the values of  $\xi$  for the *backward-looking* test (between 0.0840 and 0.089) are smaller than those for the *forward-looking* and *interpolated* tests. Again, this result is as expected. The value of  $\xi$  can thus be interpreted as a measure of deviation from i.i.d. normal variables for the jump test statistics.

Given the calibrated values of  $\xi$ , the tests are correctly sized and so direct comparison of test powers can be made. In the last column of Table 6.7, tests with the highest power are marked with an asterisk (\*). The *backward-looking* test scores the highest powers (77.83% and 73.35%) at the 2- and 5-min frequencies. At the 10-min frequency, the *interpolated*

**TABLE 6.8 Simulation (SV2FJ\_2 $\rho$  Model) and empirical results of BNS test, SPY.**

	$E[N_t] = 0.08$		$E[N_t] = 0.40$		SPY data	
	Size	Power	Size	Power	Count	Intensity
2 min	3.0800	67.43%	2.2350	44.86%	211	19.56%
5 min	3.6100	64.95%	2.6180	39.96%	116	10.75%
10 min	3.9740	58.52%	2.9340	33.05%	60	5.56%

Table 6.9 reports the simulation (under SV2FJ\_2 $\rho$  model) and empirical results of BNS test for SPY data. The BNS test is implemented with the Huang–Tauchen statistic in (6.8) at 1% significance level.

test has the highest power (66.61%). The tests attain higher power at the higher frequencies, consistent with the simulation results in the study [58]. The *interpolated* test achieves powers within 1% of those of *backward-looking* test at the higher frequencies. Considering that the *interpolated* test is implemented on a real-time basis, whereas the *backward-looking* test can be done only at market’s close, it is thus of value in adopting this *interpolated* procedure.<sup>26</sup>

The simulation results, when  $E[N_t]$  is equal to 0.40, exhibit qualitatively similar outcomes. To save space, we do not tabulate but list a few results. First, the tests tend to have larger sizes and lower powers at the lower frequencies. Second, tests executed with Šidák correction and Gumbel critical values are oversized. Third, the calibrated values of  $\xi$  at which the tests have correct sizes are strictly positive. Fourth, when test sizes are held at nominal levels, the *backward-looking* test has the highest power at the three frequencies. Last, the tests are less effective in spotting the frequent and small jumps; when  $E[N_t]$  is equal to 0.40, the test powers are much lower than those in Table 6.7.

For completeness, in Table 6.8 we report the simulation results of the BNS jump test, implemented with the Huang–Tauchen statistic in (6.7). Regardless of  $E[N_t]$ , the daily BNS test has much less spurious detection than is observed in the intraday tests. The power of BNS test declines from around 64% when  $E[N_t]$  is 0.08 to around 39% when  $E[N_t]$  is 0.40; in

<sup>26</sup>Timely knowledge on jump occurrences may be relevant in obtaining the conditional probability of jump arrival in the next interval, particularly when jumps in prices present some degree of clustering.

either cases, the BNS test does not show a higher power of the intraday tests, a result that is documented in ABD (2007). Our simulation results on BNS test in Table 6.8 are also in agreement with those in the study [16].

## 6.5 Empirical results

To apply our jump tests to SPY data, we reserve the first 4 years of data from 2002 to 2005 as an in-sample period for estimating the HAR model. The test period is thus from January 2, 2006, to April 30, 2010, totaling 1079 days. In light of the simulation results, we implement only the *backward-looking* and *interpolated* tests. The calibrated values of  $\xi$  from simulation will be used for detecting jumps in SPY data. Since there are two jumps dynamics in the simulation of SV2FJ\_2 $\rho$  model, we have two sets of  $\xi$  values and both will be used in the empirical analysis.

Boudt et al. [28] emphasize that IVP must be accounted for when implementing jump tests.<sup>27</sup> To be consistent with the information sets of the tests, the IVP for the first day of test period is given as [72]:

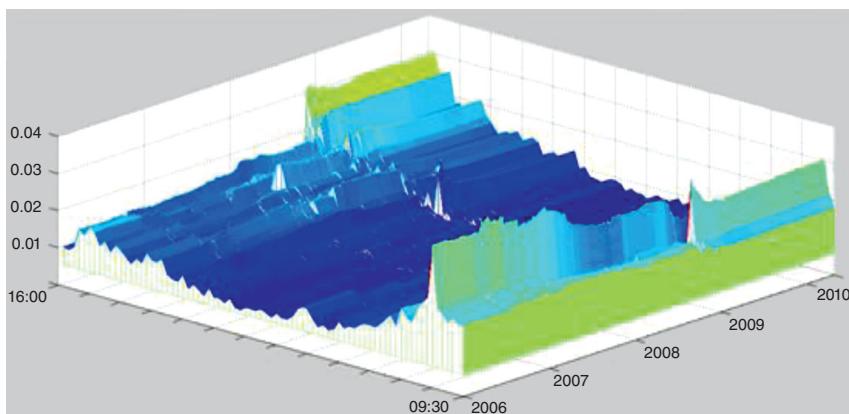
$$\eta_{1,j} = \sum_{t=1}^{T_{IS}} r_{t,j}^2 / \sum_{j=1}^M \sum_{t=1}^{T_{IS}} r_{t,j}^2, j = 1, \dots, M, \quad (6.33)$$

where  $t = 1, \dots, T_{IS}$  is the in-sample period during 2002 and 2005. When performing the *interpolated* test on the first day of test period, this estimated  $\eta_{1,j}$  is used to remove IVP.

Figure 6.4 plots the estimated IVP for 5-min SPY returns over the test period. The first pattern is estimated from the in-sample period during 2002 and 2005 using (6.33). The estimation window is then moved forward day by day, up to April 30, 2010.

For the *backward-looking* test, since it is performed at the market's close, the in-sample period is shifted 1 day forward and the IVP is updated accordingly. The procedure is then repeated for each day in the test period. Figure 6.4 plots the estimated IVP for 5-min SPY returns; a typical U-shape pattern evolves over the test period.

<sup>27</sup>Coles [32] also notes that extreme value theory is valid only for stationary processes and so the nonstationary intraday volatility patterns must be addressed before applying extreme value methods.



**FIGURE 6.4** Estimated IVP over test period, SPY.

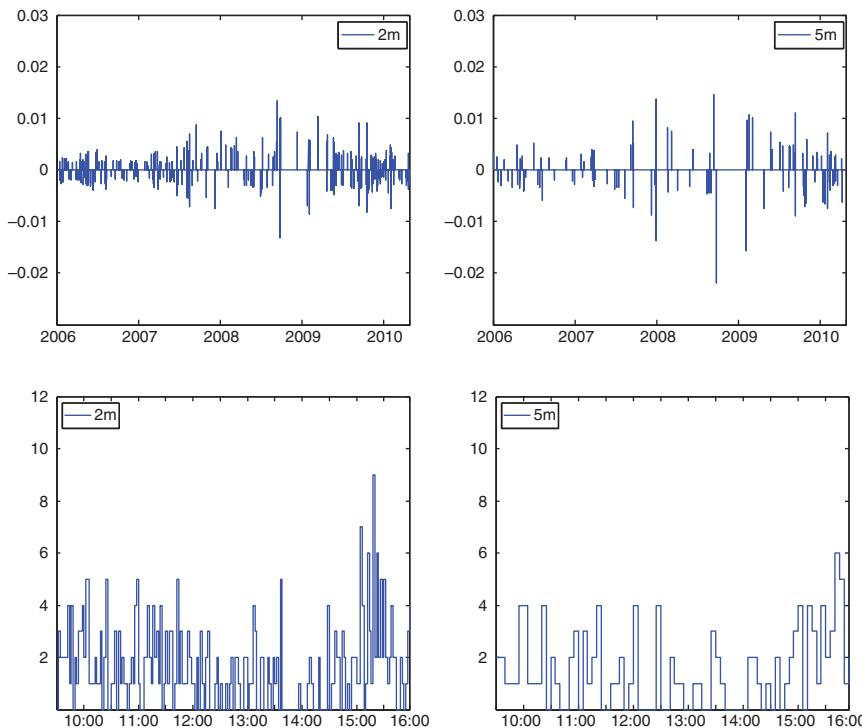
### 6.5.1 RESULTS ON THE BACKWARD-LOOKING TEST

In Table 6.9, we report the number of jumps given by the *backward-looking* test at 1% level. The results are very consistent with the top panels of Table 6.7 from simulation. The Šidák correction and Gumbel critical values lead to much more jump detections than the calibrated  $\xi$  values. The two sets of calibrated  $\xi$  values identify similar numbers of jumps in

**TABLE 6.9** Jump detection by Backward-Looking Test.

	Šidák correction			Gumbel c.v.		
	$\xi$	Count	$\lambda_{\text{Poi}}$	$\xi$	Count	$\lambda_{\text{Poi}}$
2-min	—	449	41.61%	0	462	42.82%
5-min	—	197	18.26%	0	196	18.16%
10-min	—	132	12.23%	0	125	11.58%
$\xi$ with $E[N_t] = 0.08$						
Calib. $\xi$	Count	$\lambda_{\text{Poi}}$	Calib. $\xi$	Count	$\lambda_{\text{Poi}}$	
2-min	0.0890	303	28.08%	0.0840	310	28.73%
5-min	0.0870	105	9.73%	0.0780	114	10.57%
10-min	0.0840	66	6.12%	0.0750	71	6.58%

Table 6.9 reports the jump test results on SPY data at 1% level when *backward-looking* BP is used to estimate spot volatility. The test period is from January 2, 2006, to April 30, 2010, a total of 1079 days.



**FIGURE 6.5** Detected jumps during the test period and intraday distributions, SPY.

SPY prices. For our test period of 1079 days, a 1% level has a nominal size 10.79. In Table 6.9, across three frequencies there are significantly more jumps identified than 10.79 when the calibrated  $\xi$  values are used in the test. The daily intensity  $\lambda_{\text{Poi}}$  ranges from 6.12% at the 10-min frequency to 28.73% at the 2-min frequency.<sup>28</sup> In Figure 6.5, we plot the identified jumps in SPY prices when  $\xi$  is calibrated with  $E[N_t] = 0.40$ . The jumps displayed are high-frequency returns with the IVP removed.<sup>29</sup> It can be seen that there are more jumps detected at the higher frequencies, and jump occurrences tend to cluster after 15:00 EST at the higher frequencies.<sup>30</sup>

<sup>28</sup> As our SV2FJ\_2 $\rho$  model underestimates volatility persistence at long lags, the calibrated  $\xi$  values may still be biased downward and so there are, at most, some spurious detections.

<sup>29</sup> To save space, we present only the results at the 2- and 5-minute frequencies.

<sup>30</sup> This is different from Boudt et al. [28, 29], who find that jump detections most often occur at 14:00 EST.

**TABLE 6.10 Summary statistics and interarrival times of detected jumps, SPY.**

Frequency	Detected jumps			Interarrival times		
	2 min	5 min	10 min	2 min	5 min	10 min
Mean	-2.2E-5	-2.7E-4	-5.0E-4	3.4577	9.4339	15.3256
SD	0.0037	0.0058	0.0079	5.1012	12.8664	14.0120
Dispersion	-	-	-	2.1765	1.8601	0.8359
Skewness	0.1935	-0.3068	-0.9144	4.3027	3.1233	1.7925
Kurtosis	3.7775	4.3036	4.6436	31.7083	16.4871	6.0830
Maximum	0.0134	0.0147	0.0146	51.2769	89.5897	65.0000
Minimum	-0.0131	-0.0220	-0.0277	0.0000	0.0000	0.0256
JB/KS Test	9.7425 (0.0147)	9.8601 (0.0163)	17.8852 (0.0047)	0.1129 (5.0E-4)	0.1515 (0.0056)	0.1055 (0.3180)
Q (6.20)	- -	- -	- -	57.1400 (<0.001)	26.6639 (0.1450)	17.7471 (0.6041)

Table 6.10 reports the summary statistics of detected jumps and their interarrival times for SPY data. The jumps are obtained at 1% level when  $\xi$  is calibrated with  $E[N_t] = 0.40$  in SV2FJ\_2 $\rho$  model. The reported statistics are computed with IVP removed. The interarrival times are expressed in daily unit. The dispersion is defined as  $\text{var}/(\text{mean})^2$  for the interarrival times. The JB test refers to the test for normality in the left panel, while the KS test is for the null of exponential distribution in the right panel.

Table 6.10 contains the summary statistics of detected jumps. The empirical distributions of jump returns are not Gaussian and the sample means of detected jumps are all negative. In the right panel, the interarrival times between jumps display overdispersion at the 2- and 5-min frequencies. A Monte Carlo Kolmogorov–Smirnov (KS) test for the interarrival times rejects the null of exponential distribution at 1% level at the higher frequencies. Thus, the detected jumps in SPY do not appear to follow Poisson arrivals but display some degree of clustering.

Our empirical results are comparable with many previous ones. In [45], the estimated daily jump intensity 5.6% for 11-min SPY returns at 1% level is very close to our estimates 6.12% and 6.58% for 10-min returns in Table 6.9. ABD (2007) obtains an empirical jump intensity of 9.26% at 0.001% level for 2-min S&P 500 futures returns, which is considered by them a very low value. Thus, the reported intensity 28% for 2-min SPY returns in Table 6.9 may be within a reasonable interval.

### 6.5.2 RESULTS ON THE INTERPOLATED TEST

To implement the *interpolated* test, we first estimate the HAR model in (6.18) and (6.19) for  $\log(\text{BP})$  from the in-sample period. We obtain the 1-day-ahead forecast  $E[\text{BP}_{t+1}|I_t]$  by rolling the estimation window forward. The test results are again consistent with our simulation. The Šidák correction and the Gumbel critical values give much more jump detections than those in Table 6.9. When the test is performed using calibrated  $\xi$  values, similar or marginally more jumps are obtained than the *backward-looking* test, with estimated intensity at most 3.3% higher than those of the *backward-looking* test. The overall results suggest that the real-time *interpolated* test has slightly inferior size and power properties to the *backward-looking* test.

Figure 6.5 plots the detected jumps in SPY returns, obtained when  $\xi$  is calibrated with  $E[N_t] = 0.40$  in SV2FJ\_2 $\rho$  model. The marks are high-frequency SPY returns with IVP removed. In the bottom panel, the plotted are frequency distributions of jump detections (counts) within a trading day.

## 6.6 Conclusion

By generalizing two intraday tests commonly used in the literature, in this study we implement a data-driven procedure in which the jump tests are exempt from size distortion and as a result direct comparison of test powers can be made. This improvement on the jump tests is achieved by exploiting the GEV distribution of the maximums of the test statistics, in which the shape parameter  $\xi$  is shown to be efficient in determining an optimal critical region of the tests. From the simulation results, it is seen that the stronger the effect of stochastic volatility on jump test statistics, the larger the shape parameter  $\xi$  are required to hold the test sizes at nominal level. Therefore, the shape parameter  $\xi$  of GEV distribution effectively measures the degree of deviation from i.i.d. Gaussian for the jump test statistics. This observation is one of our main contributions to the existing literature.

Our simulation study suggests that a *backward-looking* test as in ABD (2007) obtains the highest power when the test sizes are controlled at nominal level. On the other hand, the *interpolated* test proposed in this

study, which is a real-time testing procedure, can have powers close to the *backward-looking* test. The values of  $\xi$  calibrated from our SV2FJ\_2 $\rho$  model, which is specified and estimated on high-frequency SPY data, are then used in the empirical analysis. The jump test results on SPY data are consistent with our simulation study in many aspects and thus are considered plausible. Our estimates of (detected) jumps intensity are comparable with those reported in other studies. We also document that detected jumps in high-frequency SPY returns display some degree of clustering in arrivals and do not appear to have a Gaussian size distribution. These two critical observations may have to be incorporated in further studies on relevant topics in high-frequency finance.

---

## Acknowledgments

---

We thank the editor and two anonymous reviewers for comments on this chapter. The chapter is part of my PhD thesis and I thank my supervisor Stephen Taylor for his guidance and encouragement. I also benefit greatly from discussions with Jeff Fleming, Corsi Fulvio, Dudley Gilder, Ilze Kalnina, Masahito Kobayashi, Suzanne Lee, Roel Oomen, Lei Sun, Chi-Feng Tzeng, Giovanni Urga, Almut Veraart, and Martin Widdicks. Any remaining errors are our responsibility.

### Appendix 6.A: Least-square estimation of HAR-MA (2) model for log(BP) of SPY

HAR	<i>c</i>	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$	$b_9$	$b_{10}$
2 min	-0.4817 (0.1062)	0.3588 (0.0257)	0.3969 (0.0373)	0.0622 (0.0365)	0.1323 (0.0290)	-4.7941 (0.8255)	-10.6475 (2.3341)	-22.6403 (5.8768)	-8.1348 (1.3664)	-17.9600 (4.0129)	-20.8579 (9.4882)
5 min	-0.5378 (0.1208)	0.3093 (0.0260)	0.4097 (0.0391)	0.0834 (0.0397)	0.1425 (0.0319)	-4.6548 (0.9298)	-13.3684 (2.6251)	-25.1228 (6.6471)	-6.9036 (1.5382)	-21.2623 (4.5047)	-21.1967 (10.5808)
10 min	-0.5777 (0.1304)	0.2628 (0.0258)	0.3925 (0.0408)	0.1396 (0.0437)	0.1461 (0.0352)	-5.6899 (1.0191)	-15.2989 (2.8699)	-29.1153 (7.2799)	-5.8881 (1.6826)	-25.1451 (4.9190)	-23.4155 (11.5877)
	2 min	5 min	10 min		MA (2)	$\varphi_1$	$\varphi_2$		$\sigma_e$	log L.	Q (6.20)
$\sigma_u$	0.3946	0.4444	0.4862		2 min	-0.0219	0.0696		0.3924	903.80	37.336
Adj. $R^2$	0.8575	0.8254	0.8004			(0.0220)	(0.0220)		(0.0062)	—	(0.0030)
Q (6.20)	51.621	50.098	34.782		5 min	-0.0222	0.0614		0.4423	655.54	43.443
<i>p</i> value	(5.3E-8)	(1.0E-7)	(6.5E-5)			(0.0220)	(0.0225)		(0.0070)	—	(4.1E-4)
					10 min	-0.0211	0.0552		0.4841	468.27	26.969
						(0.0219)	(0.0217)		(0.0074)	—	(0.0585)

In the top and bottom-left panel, we report least-square estimation of HAR model in (6.18) for log(BP) of SPY. Most estimates of  $\{b_i\}$ ,  $i \leq 4$ , are significantly positive; all estimates of  $\{b_i\}$ ,  $i \geq 5$ , are significantly negative. Hence, the signs of  $b_i$  are consistent with the autoregressive aspect and leverage effect of volatility. The adjusted  $R^2$  ranges from 86% to 80% and on average increases by 2% when lagged returns are included in HAR. The Q (6.20) test rejects zero correlation in residuals  $u_t$  with very small  $p$  values. Examination of ACFs of HAR residuals  $u_t$  finds that across three frequencies, the ACFs at lag 2 exceed the 95% upper bound of an uncorrelated process. An MA (2) model in (6.19) is thus designed for the HAR residuals  $u_t$  and the estimation result is reported in the bottom-right panel. Across three frequencies, the estimates of  $\varphi_1$  are insignificant while  $\varphi_2$  is estimated between 0.05 and 0.07 with standard errors around 0.02. Estimates of  $\sigma_e$  are consistently smaller than those of  $\sigma_u$ . At the 10-min frequency, the Q (6.20) statistic cannot reject zero correlation in the HAR-MA (2) residuals at 5% level.

## Appendix 6.B: Estimation of ARMA (2, 1) model for log(BP) of SPY

	Time-domain estimation			Frequency-domain estimation			
	2 min	5 min	10 min		2 min	5 min	10 min
$\mu$	-9.6302 (0.1812)	-9.7447 (0.1895)	-9.7840 (0.2040)	$\alpha_1$	0.9870 (0.0044)	0.9860 (0.0047)	0.9875 (0.0042)
$\phi_1$	1.2251 (0.0446)	1.1824 (0.0412)	1.1849 (0.0387)	$\sigma_{\varepsilon,1}$	0.1642 (0.0125)	0.1703 (0.0133)	0.1625 (0.0133)
$\phi_2$	-0.2359 (0.0429)	-0.1946 (0.0395)	-0.1959 (0.0371)	$\alpha_2$	0.2413 (0.0448)	0.2007 (0.0411)	0.2012 (0.0371)
$\psi$	-0.7005 (0.0352)	-0.7061 (0.0319)	-0.7449 (0.0288)	$\sigma_{\varepsilon,2}$	0.3445 (0.0103)	0.3900 (0.0108)	0.4419 (0.0109)
$\sigma_\varepsilon$	0.4193 (0.0072)	0.4685 (0.0073)	0.5146 (0.0079)	-	-	-	-
log L.	766.48	536.26	341.28	log L.	768.71	538.25	343.53
Imp. $\alpha_1$	0.9858	0.9848	0.9863	$w_1$	89.23%	86.81%	83.93%
Imp. $\alpha_2$	0.2393	0.1976	0.1986	$w_2$	10.77%	13.19%	16.07%

In the left panel, we report the time-domain estimation of ARMA (2, 1) model in (6.27) for log(BP) of SPY. The two roots  $\alpha_1$  and  $\alpha_2$  of

$x^2\phi(x^{-1}) = 0$ , where  $\phi(L) = 1 - \phi_1L - \phi_2L^2$  is the AR (2) polynomial, are calculated at around (0.98, 0.20). In the right panel, frequency-domain estimation is done by maximizing the Whittle likelihood function:

$$-0.5 \sum_{\lambda} [\log(2\pi \cdot f(\lambda)) + I(\lambda)/f(\lambda)],$$

where  $\lambda = (1, 2, \dots, T-1)(2\pi/T)$  is the frequency ordinate,  $T = 2076$ ,  $I(\lambda)$  is the sample spectrum, and  $f(\lambda)$  is the theoretical spectral density:

$$f(\lambda) = \left( \frac{\sigma_{\varepsilon,1}^2}{2\pi} \right) \left| 1 - \alpha_1 e^{-i\lambda} \right|^{-2} + \left( \frac{\sigma_{\varepsilon,2}^2}{2\pi} \right) \left| 1 - \alpha_2 e^{-i\lambda} \right|^{-2}.$$

The weights of the two processes are given by  $w_i = \text{var}_i / \sum \text{var}_i$ ,  $i = 1, 2$ , where  $\text{var}_i = \sigma_{\varepsilon,i}^2 / (1 - \alpha_i^2)$  is the variance of individual process. The estimated  $\alpha_1$  and  $\alpha_2$  are very close to the implied values in the left panel. The persistent AR (1) process has weight ranging from 89% at the 2-min frequency to 84% at the 10-min frequency.

### Appendix 6.C: Minimized loss function $\text{loss}(\rho_1, \rho_2)$ for SV2FJ\_2 $\rho$ model, SPY

Frequency	$\rho_1$ (2 min)			$\rho_1$ (5 min)			$\rho_1$ (10 min)		
$\rho_2$	-0.450	-0.475	-0.500	-0.475	-0.500	-0.525	-0.525	-0.550	-0.575
-0.100	0.0019	0.0015	0.0015	0.0022	0.0018	0.0019	0.0026	0.0024	0.0023
-0.125	0.0017	0.0015	0.0014	0.0022	0.0018	0.0019	0.0025	0.0022	0.0025
-0.150	0.0016	<b>0.0013</b>	0.0014	0.0021	<b>0.0017</b>	0.0019	0.0023	<b>0.0021</b>	0.0022
-0.175	0.0018	0.0014	0.0016	0.0021	0.0019	0.0018	0.0023	0.0022	0.0022
-0.200	0.0017	0.0015	0.0017	0.0020	0.0019	0.0019	0.0024	0.0023	0.0024

In this table, we report the minimized loss function  $\text{loss}(\rho_1, \rho_2)$  for SV2FJ\_2 $\rho$  model in (6.32). For each frequency, local minimums of loss function can be found from a grid of candidate values on  $(\rho_1, \rho_2)$ . The minimized loss functions become smaller at the higher frequencies. We also calculate the minimized loss function when there is only one correlation parameter in the SV2FJ\_2 $\rho$  model, which is the one attached to the persistent volatility component; in that case the minimized loss functions are consistently larger than the reported values here across all three frequencies.

### Appendix 6.D.1: Calibration of $\xi$ under SV2FJ\_2 $\rho$ model at 2-min frequency, $E[N_t] = 0.08$

Backward-looking BP			Forward-looking BP			Interpolated BP		
Correction	Size	Power	Correction	Size	Power	Correction	Size	Power
Šidák	6.9920	79.32%	Šidák	15.8080	79.67%	Šidák	11.7940	79.54%
Bonferroni	6.9710	79.32%	Bonferroni	15.7610	79.67%	Bonferroni	11.7500	79.54%
$\xi$	Size	Power	$\xi$	Size	Power	$\xi$	Size	Power
0.0000	7.7340	79.48%	0.0000	17.2040	79.81%	0.0000	12.8440	79.66%
0.0100	6.9420	79.32%	0.0180	14.5680	79.55%	0.0200	10.5260	79.34%
0.0200	6.1800	79.16%	0.0360	12.1400	79.30%	0.0400	8.5610	79.03%
0.0300	5.5030	78.98%	0.0540	10.0340	78.96%	0.0600	6.8570	78.72%
0.0400	4.8840	78.82%	0.0720	8.2150	78.66%	0.0800	5.4270	78.26%
0.0500	4.2920	78.68%	0.0900	6.6660	78.36%	0.1000	4.2270	77.80%
0.0600	3.7940	78.41%	0.1080	5.3460	77.96%	0.1200	3.2500	77.31%
0.0700	3.2910	78.26%	0.1260	4.1990	77.53%	<b>0.1380</b>	<b>2.4750</b>	<b>76.96%</b>
0.0800	2.8560	78.07%	0.1440	3.2040	77.10%	0.1400	2.4090	76.91%
<b>0.0890</b>	<b>2.4950</b>	<b>77.83%</b>	<b>0.1602</b>	<b>2.5070</b>	<b>76.70%</b>	0.1600	1.7410	76.38%
0.0900	2.4580	77.82%	0.1620	2.4470	76.66%	0.1800	1.2580	75.83%
0.1000	2.1000	77.62%	0.1800	1.8630	76.22%	0.2000	0.8770	75.40%

In this table, we report the calibrated values of  $\xi$  at which the jump tests have empirical size close to 2.50 for data simulated from the SV2FJ\_2 $\rho$  model with parameters given in Table 6.6 and with  $E[N_t] = 0.08$ . The results are for 2-min sampling frequency. Results for tests implemented with Šidák and Bonferroni corrections are also given. The values of  $\xi$  are obtained from an array of candidate values with interpolation; these values along with the corresponding test size and power are highlighted in bold.

## Appendix 6.D.2: Calibration of $\xi$ under SV2FJ\_2 $\rho$ model at 2-min frequency, $E[N_t] = 0.40$

Backward-looking BP			Forward-looking BP			Interpolated BP		
Correction	Size	Power	Correction	Size	Power	Correction	Size	Power
Šidák	6.4540	58.47%	Šidák	14.4460	59.09%	Šidák	10.6350	58.64%
Bonferroni	6.4310	58.46%	Bonferroni	14.3980	59.08%	Bonferroni	10.5900	58.64%
$\xi$	Size	Power	$\xi$	Size	Power	$\xi$	Size	Power
0.0000	7.1220	58.75%	0.0000	15.7120	59.35%	0.0000	11.6290	58.88%
0.0100	6.3880	58.44%	0.0160	13.5450	58.91%	0.0200	9.5390	58.35%
0.0200	5.6940	58.16%	0.0320	11.5620	58.44%	0.0400	7.7630	57.76%
0.0300	5.0280	57.83%	0.0480	9.8270	57.96%	0.0600	6.2270	57.11%
0.0400	4.4570	57.53%	0.0640	8.2480	57.47%	0.0800	4.9480	56.42%
0.0500	3.9410	57.19%	0.0800	6.8620	56.94%	0.1000	3.8450	55.70%
0.0600	3.4510	56.86%	0.0960	5.6590	56.39%	0.1200	2.9580	54.96%
0.0700	3.0100	56.51%	0.1120	4.6090	55.81%	<b>0.1300</b>	<b>2.5190</b>	<b>54.59%</b>
0.0800	2.6140	56.18%	0.1280	3.7380	55.21%	0.1400	2.1880	54.21%
<b>0.0840</b>	<b>2.4920</b>	<b>56.03%</b>	0.1440	2.9800	54.60%	0.1600	1.5880	53.43%
0.0900	2.2850	55.81%	<b>0.1552</b>	<b>2.5180</b>	<b>54.13%</b>	0.1800	1.1470	52.56%
0.1000	1.9760	55.47%	0.1600	2.3440	53.94%	0.2000	0.8170	51.59%

In this table, we report for the 2-min frequency the calibrated values of  $\xi$  at which the jump tests have empirical sizes close to 2.50. The simulated SV2FJ\_2 $\rho$  model has parameters given in Table 6.6 with  $E[N_t] = 0.40$ .

### Appendix 6.D.3: Calibration of $\xi$ under SV2FJ\_2 $\rho$ model at 5-min frequency, $E[N_t] = 0.08$

Backward-looking BP			Forward-looking BP			Interpolated BP		
Correction	Size	Power	Correction	Size	Power	Correction	Size	Power
Šidák	7.1980	75.61%	Šidák	18.8430	76.97%	Šidák	12.4430	76.41%
Bonferroni	7.1780	75.60%	Bonferroni	18.7980	76.97%	Bonferroni	12.4070	76.41%
$\xi$	Size	Power	$\xi$	Size	Power	$\xi$	Size	Power
0.0000	7.1300	75.59%	0.0000	18.6920	76.96%	0.0000	12.3230	76.38%
0.0100	6.4690	75.37%	0.0180	15.9590	76.54%	0.0200	10.1750	75.94%
0.0200	5.7790	75.16%	0.0360	13.5650	76.09%	0.0400	8.3690	75.51%
0.0300	5.1650	74.91%	0.0540	11.4320	75.63%	0.0600	6.8020	74.98%
0.0400	4.5700	74.69%	0.0720	9.4490	75.10%	0.0800	5.4100	74.49%
0.0500	4.0710	74.45%	0.0900	7.7750	74.63%	0.1000	4.2770	73.85%
0.0600	3.5780	74.21%	0.1080	6.3760	74.10%	0.1200	3.3330	73.39%
0.0700	3.1460	73.94%	0.1260	5.1240	73.61%	0.1400	2.5450	72.79%
0.0800	2.7530	73.57%	0.1440	4.0540	73.04%	<b>0.1410</b>	<b>2.4850</b>	<b>72.75%</b>
<b>0.0870</b>	<b>2.4930</b>	<b>73.35%</b>	0.1620	3.1680	72.45%	0.1600	1.8870	72.09%
0.0900	2.4000	73.30%	<b>0.1782</b>	<b>2.4920</b>	<b>71.95%</b>	0.1800	1.3750	71.45%
0.1000	2.1000	72.98%	0.1800	2.4300	71.90%	0.2000	0.9610	70.60%

In this table, we report the calibrated values of  $\xi$  at which the jump tests have empirical size close to 2.50 for data simulated from the SV2FJ\_2 $\rho$  model with parameters given in Table 6.6 and  $E[N_t] = 0.08$ . The results are for 5-min sampling frequency.

### Appendix 6.D.4: Calibration of $\xi$ under SV2FJ\_2 $\rho$ Model at 5-min frequency, $E[N_t] = 0.40$

Backward-looking BP			Forward-looking BP			Interpolated BP		
Correction	Size	Power	Correction	Size	Power	Correction	Size	Power
Šidák	6.3760	51.28%	Šidák	16.9420	53.05%	Šidák	11.1970	52.21%
Bonferroni	6.3510	51.26%	Bonferroni	16.8860	53.04%	Bonferroni	11.1650	52.21%
$\xi$	Size	Power	$\xi$	Size	Power	$\xi$	Size	Power
0.0000	6.3190	51.24%	0.0000	16.7940	53.02%	0.0000	11.1010	52.19%
0.0100	5.7180	50.89%	0.0180	14.4260	52.42%	0.0200	9.1470	51.48%
0.0200	5.1250	50.52%	0.0360	12.2750	51.73%	0.0400	7.5060	50.74%
0.0300	4.5770	50.14%	0.0540	10.2480	51.05%	0.0600	6.0250	49.98%
0.0400	4.1020	49.72%	0.0720	8.4780	50.32%	0.0800	4.8170	49.17%
0.0500	3.6430	49.29%	0.0900	6.9760	49.62%	0.1000	3.7560	48.28%
0.0600	3.1890	48.84%	0.1080	5.6290	48.82%	0.1200	2.8780	47.34%
0.0700	2.8120	48.39%	0.1260	4.4910	47.99%	<b>0.1300</b>	<b>2.5210</b>	<b>46.83%</b>
<b>0.0780</b>	<b>2.4900</b>	<b>48.01%</b>	0.1440	3.5200	47.11%	0.1400	2.1940	46.34%
0.0800	2.4290	47.92%	0.1620	2.7830	46.24%	0.1600	1.6390	45.35%
0.0900	2.1220	47.48%	<b>0.1692</b>	<b>2.4930</b>	<b>45.90%</b>	0.1800	1.2130	44.32%
0.1000	1.8480	47.03%	0.1800	2.0850	45.34%	0.2000	0.8690	43.21%

In this table, we report for the 5-min frequency the calibrated values of  $\xi$  at which the jump tests have empirical sizes close to 2.50. The simulated SV2FJ\_2 $\rho$  model has parameters given in Table 6.6 with  $E[N_t] = 0.40$ .

### Appendix 6.D.5: Calibration of $\xi$ under SV2FJ\_2 $\rho$ model at 10-min frequency, $E[N_t] = 0.08$

Backward-looking BP			Forward-looking BP			Interpolated BP		
Correction	Size	Power	Correction	Size	Power	Correction	Size	Power
Šidák	7.6700	70.07%	Šidák	24.1360	72.20%	Šidák	13.7340	71.57%
Bonferroni	7.6450	70.06%	Bonferroni	24.0670	72.19%	Bonferroni	13.6900	71.57%
$\xi$	Size	Power	$\xi$	Size	Power	$\xi$	Size	Power
0.0000	6.7680	69.67%	0.0000	21.9780	71.92%	0.0000	12.4670	71.35%
0.0100	6.0960	69.34%	0.0200	18.7160	71.42%	0.0200	10.4630	70.77%
0.0200	5.4770	68.97%	0.0400	15.7640	70.90%	0.0400	8.5930	70.10%
0.0300	4.8830	68.58%	0.0600	13.0950	70.30%	0.0600	6.9890	69.60%
0.0400	4.3680	68.21%	0.0800	10.8260	69.70%	0.0800	5.6590	68.90%
0.0500	3.8570	67.84%	0.1000	8.7400	68.97%	0.1000	4.4870	68.14%
0.0600	3.4130	67.41%	0.1200	7.0170	68.24%	0.1200	3.5180	67.53%
0.0700	2.9930	67.03%	0.1400	5.5820	67.52%	0.1400	2.7230	66.78%
0.0800	2.6330	66.59%	0.1600	4.3930	66.78%	<b>0.1460</b>	<b>2.5050</b>	<b>66.61%</b>
<b>0.0840</b>	<b>2.5020</b>	<b>66.37%</b>	0.1800	3.3520	65.93%	0.1600	2.0560	65.99%
0.0900	2.2980	66.12%	<b>0.2000</b>	<b>2.5100</b>	<b>65.02%</b>	0.1800	1.5560	65.12%
0.1000	2.0180	65.68%	0.2020	2.4410	64.92%	0.2000	1.1480	64.05%

In this table, we report the calibrated values of  $\xi$  at which the jump tests have empirical size close to 2.50 for data simulated from the SV2FJ\_2 $\rho$  model with parameters given in Table 6.6 and  $E[N_t] = 0.08$ . The results are for 10-min sampling frequency.

### Appendix 6.D.6: Calibration of $\xi$ under SV2FJ\_2 $\rho$ model at 10-min frequency, $E[N_t] = 0.40$

Backward-looking BP			Forward-looking BP			Interpolated BP		
Correction	Size	Power	Correction	Size	Power	Correction	Size	Power
Šidák	6.7250	42.39%	Šidák	21.1080	46.15%	Šidák	11.7420	44.70%
Bonferroni	6.6990	42.38%	Bonferroni	21.0440	46.13%	Bonferroni	11.7040	44.69%
$\xi$	Size	Power	$\xi$	Size	Power	$\xi$	Size	Power
0.0000	5.9950	41.84%	0.0000	19.2810	45.66%	0.0000	10.6020	44.23%
0.0100	5.3680	41.37%	0.0200	16.3930	44.81%	0.0200	8.9360	43.35%
0.0200	4.8010	40.87%	0.0400	13.7940	43.90%	0.0400	7.3390	42.43%
0.0300	4.3600	40.40%	0.0600	11.4700	43.03%	0.0600	5.9430	41.54%
0.0400	3.9190	39.90%	0.0800	9.4700	42.07%	0.0800	4.7670	40.57%
0.0500	3.4690	39.45%	0.1000	7.6860	41.08%	0.1000	3.7250	39.54%
0.0600	3.0490	38.91%	0.1200	6.1390	40.14%	0.1200	2.9260	38.47%
0.0700	2.6920	38.36%	0.1400	4.9060	39.03%	<b>0.1320</b>	<b>2.4890</b>	<b>37.80%</b>
<b>0.0750</b>	<b>2.4910</b>	<b>38.07%</b>	0.1600	3.8290	37.95%	0.1400	2.2240	37.32%
0.0800	2.3380	37.81%	0.1800	2.9650	36.79%	0.1600	1.7030	36.16%
0.0900	2.0300	37.33%	<b>0.1920</b>	<b>2.5050</b>	<b>36.08%</b>	0.1800	1.2800	34.99%
0.1000	1.7590	36.79%	0.2000	2.2260	35.61%	0.2000	0.9350	33.81%

In this table, we report for the 10-min frequency the calibrated values of  $\xi$  at which the jump tests have empirical sizes close to 2.50. The simulated SV2FJ\_2 $\rho$  model has parameters given in Table 6.6 with  $E[N_t] = 0.40$ .

## References

1. Abdi, H. 2007, The Bonferroni and Šidák corrections for multiple comparisons, in Neil Salkind (Ed.), *Encyclopedia of Measurement and Statistics*. SAGE Publications.
2. Aït-Sahalia, Y., and J. Jacod. 2009, Testing for jumps in a discretely observed process, *The Annals of Statistics* 37, 184–222.
3. Aldous, D. 1989, *Probability Approximations via the Poisson Clumping Heuristic*. Springer.
4. Alizadeh, S., M.W. Brandt, and F.X. Diebold. 2002, Range-based estimation of stochastic volatility models, *Journal of Finance* 57, 1047–1091.
5. Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2000, Great realization, *Risk*, March 2000, 105–108.

6. Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens. 2001, The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43–76.
7. Andersen, T.G., T. Bollerslev, and D. Dobrev. 2007, No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: theory and testable distributional implications, *Journal of Econometrics* 138, 125–180.
8. Areal, N., and S.J. Taylor. 2002, The realized volatility of FTSE-100 future prices, *Journal of Futures Markets* 22, 627–648.
9. Bajgrowicz, P., and O. Scaillet. 2011, *Jumps in High-Frequency Data: Spurious Detections, Dynamics and News*, University of Geneva and Swiss Finance Institute preprint.
10. Bakshi, G., C. Cao, and Z. Chen. 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003–2049.
11. Bandi, F.M., and J.R. Russell. 2006, Separating microstructure noise from volatility, *Journal of Financial Economics* 79, 655–692.
12. Bandi, F.M., and R. Renò. 2012b, *Price and Volatility Co-jumps*, Johns Hopkins University and Università di Siena preprint.
13. Barndorff-Nielsen, O.E., and N. Shephard. 2001, Non-Gaussian Ornstein–Uhlenbeck based models and some of their uses in financial economics, *Journal of the Royal Statistical Society B*63, 167–241.
14. Barndorff-Nielsen, O.E., and N. Shephard. 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society B*64, 253–280.
15. Barndorff-Nielsen, O.E., and N. Shephard. 2004, Power and bi-power variation with stochastic volatility and jumps, *Journal of Financial Econometrics* 2: 1–48.
16. Barndorff-Nielsen, O.E., and N. Shephard. 2006, Econometrics of testing for jumps in financial economics using bi-power variation, *Journal of Financial Econometrics* 4, 1–30.
17. Barndorff-Nielsen, O.E., and N. Shephard. 2006, Impact of jumps on returns and realized variances: econometric analysis of time-deformed Lévy processes, *Journal of Econometrics* 131, 217–252.
18. Barndorff-Nielsen, O.E., and N. Shephard. 2007, Variation, jumps, market frictions and high frequency data in financial econometrics, in R. Blundell, P. Torsten, & W.K. Newey (Eds.), *Advances in Economics*

- and Econometrics: Theory and Applications, Ninth World Congress* (Ch. 10, Vol. 3, pp. 328–372). Cambridge University Press.
19. Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod, and N. Shephard. 2006, Limit theorems for bi-power variation in financial econometrics, *Econometric Theory* 22, 677–719.
  20. Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard. 2008, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, *Econometrica* 76, 1481–1536.
  21. Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard. 2009, Realized kernels in practice: trades and quotes, *Econometrics Journal* 12, C1–C32.
  22. Bates, D.S. 1996, Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options, *Review of Financial Studies* 9, 69–107.
  23. Benjamini, Y., and Y. Hochberg. 1995, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society B* 57, 289–300.
  24. Bollerslev T., G. Tauchen, and T.H. Law. 2008, Risk, jumps and diversification, *Journal of Econometrics* 144, 234–256.
  25. Bollerslev T., U. Kretschmer, C. Pigorsch, and G. Tauchen. 2009, A discrete-time model for daily S&P 500 returns and realized variations: jumps and leverage effects, *Journal of Econometrics* 150, 151–166.
  26. Bollerslev T., V. Todorov, and S. Z. Li. 2013, Jump tails, extreme dependencies, and the distribution of stock returns, *Journal of Econometrics* 172, 307–324.
  27. Bos, C.S., P. Janus, and S.J. Koopman. 2012, Spot variance path estimation and its application to high frequency jump testing, *Journal of Financial Econometrics* 10, 354–389.
  28. Boudt, K., C. Croux, and S. Laurent. 2011, Robust estimation of intraweek periodicity in volatility and jump detection, *Journal of Empirical Finance* 18, 353–367.
  29. Boudt, K., J. Cornelissen, C. Croux, and S. Laurent. 2011, Non-parametric tests for intraday jumps: impact of periodicity and microstructure noise, in L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Volatility Models and Their Applications*. Wiley-Interscience.
  30. Breidt, F.J., and R.A. Davis. 1998, Extremes of stochastic volatility models, *The Annals of Applied Probability* 8, 664–675.

31. Christensen, K., R.C.A. Oomen, and M. Podolskij. 2010, Realised quantile-based estimation of the integrated variance, *Journal of Econometrics* 159, 74–98.
32. Coles, S. 2001, *An Introduction to Statistical Modeling of Extreme Values*. Springer.
33. Corsi, F. 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.
34. Corsi, F., D. Pirino, and R. Renò. 2010, Threshold bi-power variation and the impact of jumps on volatility forecasting, *Journal of Econometrics* 159, 276–288.
35. Cramér, H. 1946, *Mathematical Methods of Statistics*. Princeton University Press.
36. Das, S.R., and R.K. Sundaram. 1999, Of smiles and smirks: a term structure perspective, *Journal of Financial and Quantitative Analysis* 34, 211–240.
37. Delbaen, F., and W. Schachermayer. 1994, A general version of the fundamental theorem of asset pricing, *Mathematische Annalen* 300, 463–520.
38. Duffie, D., J. Pan, and K. Singleton. 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.
39. Dumitru, A.-M., and G. Urga. 2012, Identifying jumps in financial assets: a comparison between nonparametric jump tests, *Journal of Business and Economic Statistics* 30, 242–255.
40. Eraker, B., M. Johannes, and N. Polson. 2003, The impact of jumps in volatility and returns, *Journal of Finance* 58, 1269–1300.
41. Evans, K.P. 2011, Intraday jumps and US macroeconomic news announcements, *Journal of Banking & Finance* 35, 2511–2527.
42. Fleming, J., and B.S. Paye. 2011, High-frequency returns, jumps and the mixture of normal hypothesis, *Journal of Econometrics* 160, 119–128.
43. Galambos, J. 1978, *The Asymptotic Theory of Extreme Order Statistics*. John Wiley & Sons, Inc.
44. Gallant, A.R., C.-T. Hsu, and G. Tauchen. 1999, Using daily range data to calibrate volatility diffusions and extract the forward integrated variance, *The Review of Economics and Statistics* 81, 617–631.

45. Gilder, D., M.B. Shackleton, and S.J. Taylor. 2013, *Cojumps in Stock Prices: Empirical Evidence*, Aston Business School and Lancaster University Management School preprint.
46. Hall, P. 1979, On the rate of convergence of normal extremes, *Journal of Applied Probability* 16, 433–439.
47. Hansen, P.R., and A. Lunde. 2006, Realized variance and market micro-structure noise, *Journal of Business & Economic Statistics* 24, 127–161.
48. Hansen, P.R., Z. Huang, and H.H. Shek. 2012, Realized GARCH: a joint model for returns and realized measures of volatility, *Journal of Applied Econometrics* 27, 877–906.
49. Harvey, A.C. 1993, *Time Series Models*. Pearson Education.
50. Hasbrouck, J. 2003, Intraday price formation in U.S. equity index markets, *Journal of Finance* 58, 2375–2400.
51. Huang, X., and G. Tauchen. 2005, The relative contribution of jumps to total price variance, *Journal of Financial Econometrics* 3, 456–499.
52. Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter. 2009, Microstructure noise in the continuous case: the pre-averaging approach, *Stochastic Processes and Their Applications* 119, 2249–2276.
53. Jacod, J., C. Klüppelberg, and G. Müller. 2012, Functional relationships between price and volatility jumps and their consequences for discretely observed data, *Journal of Applied Probability* 49, 901–914.
54. Jacod, J., and V. Todorov. 2010, Do price and volatility jump together? *The Annals of Applied Probability* 20, 1425–1469.
55. Kalnina, I. 2011, Subsampling high frequency data, *Journal of Econometrics* 161, 262–283.
56. Kobayashi, M. 2009, Testing for jumps in the stochastic volatility models, *Mathematics and Computers in Simulation* 79, 2597–2608.
57. Lahaye, J., S. Laurent, and C.J. Neely. 2011, Jumps, cojumps and macro announcements, *Journal of Applied Econometrics* 26, 893–921.
58. Lee, S.S., and P.A. Mykland. 2008, Jumps in financial markets: a new nonparametric test and jump dynamics, *The Review of Financial Studies* 21, 2535–2563.
59. Lee, S.S. 2011, Jumps and information flow in financial markets, *Review of Financial Studies* 25, 439–479.

60. Liu, C., and J.M. Maheu. 2009, Forecasting realized volatility: a Bayesian model-averaging approach, *Journal of Applied Econometrics* 24, 709–733.
61. Mykland, P.A., and L. Zhang. 2009, Inference for continuous semi-martingales observed at high frequency, *Econometrica* 77, 1403–1445.
62. Mancini, C. 2004, Estimation of the characteristics of the jumps of a general Poisson-diffusion model, *Scandinavian Actuarial Journal* 1, 42–52.
63. Mancini, C. 2009, Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps, *Scandinavian Journal of Statistics* 36, 270–296.
64. Patton, A.J., and K. Sheppard. 2011, *Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility*, Oxford-Man Institute of Quantitative Finance preprint.
65. Peters, R.T., and R.G. de Vilder. 2006, Testing the continuous semi-martingale hypothesis for the S&P 500, *Journal of Business and Economic Statistics* 24, 444–454.
66. Platen, E., and N. Bruti-Liberati. 2010, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Springer.
67. Podolskij, M., and D. Ziggel. 2010, New tests for jumps in semi-martingale models, *Statistical Inference for Stochastic Processes* 13, 15–41.
68. Pong S., M.B. Shackleton, S.J. Taylor, and X. Xu. 2004, Forecasting currency volatility: a comparison of implied volatilities and ARFIMA models, *Journal of Banking and Finance* 28, 2541–2563.
69. Rasmussen, T.B. 2009, *Jump Testing and the Speed of Market Adjustment*, Aarhus University School of Economics and Management preprint.
70. Rogers, L.C.G. 1997, Arbitrage with fractional Brownian motion, *Mathematical Finance* 7, 95–105.
71. Shreve, S.E. 2004, *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer.
72. Taylor, S.J., and X. Xu. 1997, The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance* 4, 317–340.
73. Taylor, S.J. 2005, *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press.

74. Taylor, S.J. 2010, *An Econometric Defence of Pure-Jump Price Dynamics*, Lancaster University Management School preprint.
75. Theodosiou, M., and F. Zikes. 2011, *A Comprehensive Comparison of Alternative Tests for Jumps in Asset Prices*, Central Bank of Cyprus and Imperial College London preprint.
76. Todorov, V., 2011, Econometric analysis of jump-driven stochastic volatility models, *Journal of Econometrics* 160, 12–21.
77. Vasudeva, R., and A.Y. Moridani. 2010, Limit distributions of the extremes of a random number of random variables in a stationary Gaussian sequence, *ProbStat Forum* 3, 78–90.
78. Veraart, A.E.D. 2010, Inference for the jump part of quadratic variation of Itô semi-martingales, *Econometric Theory* 26, 331–368.
79. Zhang, L., P.A. Mykland, and Y. Aït-Sahalia. 2005, A tale of two time scales: determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* 100, 1394–1411.

## Chapter Seven

# Hawkes Processes and Their Applications to High-Frequency Data Modeling

Baron Law and Frederi G. Viens

Purdue University, West Lafayette, IN, USA

### 7.1 Introduction

This short chapter introduces and surveys an emerging class of stochastic point processes used in modeling the evolution of high-frequency (HF) data on stock markets at a high level of quantitative detail.

The information contained in a stock market's *Limit Order Books* (*LOB*) is a multivariate time series that records the order arrival times and volumes at each price level of thousands of stocks trading on the exchange. An LOB exhibits a number of distinctive characteristics [1–3], including

1. irregular time interval between arrivals,
2. discrete state space of price ticks and volume lot sizes,
3. intraday seasonality (more activities around market open and close),
4. arrival clustering,
5. self-excitation from its own history,

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

6. cross (mutual) excitation from the history of other assets, and
7. long memory of excitation effect.

Consequently, classical time series models with fixed time intervals such as ARIMA and GARCH are not suitable to model HF financial data. A standard approach commonly used in practice is to resample the data in 5-min intervals [4, 5], thereby avoiding the time scale for liquid stocks where many of the characteristics listed earlier can be observed, but this may amount to discarding more than 99% of the data for such stocks. On the contrary, Poisson processes, which are widely used in the market microstructure literature [6, 7], fail to depict the aforementioned prevalent features of HF data.

This survey paper, on the current research in HF financial data modeling, concentrates on the use of so-called Hawkes processes, a family of point processes designed to model self- and cross-excitation. In Section 7.2, we offer an informal introduction to point processes while a more technical review can be found in Appendix 7.A. All the presented material about point processes is thoroughly covered in major textbooks including our references [8–12]. Sections 7.3 and 7.4 introduce Hawkes processes and their statistical inference; a brief history thereof is provided in Appendix 7.B. Section 7.5 presents the applications of Hawkes processes to HF data modeling.

---

## 7.2 Point processes

---

This section provides an informal introduction to point processes. A more rigorous treatment using random measures and martingale theory can be found in Appendix 7.A.

A point process is a random countable set of points  $\{x_1, x_2, \dots\}$  on a set  $X$ . The  $\{x_n\}$ 's are  $X$ -valued random variables; they can have a highly nontrivial dependence structure. For example,  $\{x_n\}$ 's can be locations of earthquakes on  $X = \mathbb{R}^2$ , and if self-excitation (property 5 above) is built into the model, which is the case with Hawkes processes as we will discuss, then  $\{x_n\}$ 's can exhibit random patterns of aftershocks surrounding major earthquakes.

Let  $N(A)$  be the number of points inside a region  $A \subseteq X$ . If  $N(A)$  is known for all subsets  $A$  of  $X$ , this essentially determines the locations of all the points and this representation is the principal tool to describe a point

process. A point process  $N$  is called *simple* if each location has at most one point.

When  $X$  is the positive half-line  $\mathbb{R}_+$ , this typically represents the time axis and the points are regarded as the times of event occurrences. In this case,  $N(t) = N((0, t])$  denotes the number of occurrences at or before time  $t$ . For our purpose of modeling HF data, we will mainly deal with point processes on  $\mathbb{R}_+$ .

When an event happens at  $t_n \in \mathbb{R}_+$ , it may carry an additional information  $y_n$  (its mark). For instance, in finance, each order arrival is associated with an order quantity (volume); in seismology, each earthquake is reported with a magnitude. A point process with marks is called a marked point process (MPP). Let  $Y$  be the mark space (i.e.,  $y_n \in Y$ ) and  $A \subseteq Y$ ; then  $N(t, A) = N((0, t] \times A)$  denotes the number of events that happened at or before time  $t$  such that the marks fall within the set  $A$ . The ground process  $N_g(t) = N((0, t] \times Y)$  is the count of all events in  $(0, t]$ , regardless of the marks.<sup>1</sup> The marks of an MPP are called *unpredictable* if  $y_n$  is independent of  $\{(t_i, y_i)\}_{i < n}$  and they are called *independent* if  $y_n$  is independent of  $\{(t_i, y_i)\}_{i \neq n}$ .<sup>2</sup> An MPP  $N$  is called simple if  $N_g$  is simple.

If a point process has multiple occurrences in the same location, we can treat it as a simple MPP with the mark being the number of points in each location. Hence, without loss of generality, most of the results to be presented will be based on simple point processes.

$N$  is called a multivariate point process if the mark space  $Y = \{1, \dots, d\}$ . In this case,  $N_i(\bullet) = N(\bullet \times \{i\})$  is called the marginal process for the points of type  $i$ . A simple  $d$ -variate point process is different from a  $d$ -dimensional ( $\mathbb{R}^d$ -valued) simple point process as the former cannot have any common jump times. In addition, if the multivariate point process also carries some extra information  $w_n \in W$  at each point, the mark space will become  $Y = \{1, \dots, d\} \times W$ . In this case, the marginal processes are  $N_i(t, A) = N((0, t] \times \{i\} \times A)$  and the marginal ground processes are  $N_i(t) = N((0, t] \times \{i\} \times Y)$ .

Many point processes can be modeled in terms of their stochastic intensities<sup>3</sup>  $\lambda(t)$ , which can be defined informally as the expected number

---

<sup>1</sup>We may refer to the ground process of an MPP without the subscript “g” if the meaning is clear from the input parameter.

<sup>2</sup>Notice that if marks are independent, future location  $t_{n+1}$  cannot depend on previous mark  $y_n$ .

<sup>3</sup>See Appendix 7.A for formal definition and existence condition of stochastic intensity.

of arrivals per unit of time, at time  $t$ , conditioned on all the information just before time  $t$ , that is,

$$\lambda(t) = \lim_{h \rightarrow 0^+} \frac{\mathbb{E}(N((t-h, t]) | \mathcal{F}_{t-h})}{h} \quad (7.1)$$

where  $\mathcal{F}_t$ , which represents the information contained in all the events happened up to and including time  $t$ , denotes the filtration of  $N$ . For a multivariate point process,  $\lambda_i(t)$  is the intensity of the marginal process  $N_i(t)$ .

The most well-known point processes are Poisson processes, where the intensities are deterministic functions of time. When one generalizes the intensities to become stochastic processes on their own, they play a role similar to stochastic volatility in diffusions. One important use of stochastic intensity is to allow it to change according to how events unfold over time. The Hawkes processes are directly exploring this feature for the purpose of modeling self-excitation, as we are about to see.

### 7.3 Hawkes processes

A Hawkes process [13] is a point process where its stochastic intensity has an autoregressive form. For a nonlinear multivariate marked Hawkes process, the intensity  $\lambda(t) = (\lambda_1(t), \dots, \lambda_d(t))$  is given by<sup>4,5</sup>

$$\begin{aligned} \lambda_i(t) &= \Phi_i \left( \sum_{j=1}^d \int_{(-\infty, t) \times Y} \gamma_{ij}(t-s, y) N_j(ds \times dy), t \right) \\ &= \Phi_i \left( \sum_{t_n < t} \gamma_{i, w_n}(t - t_n, y_n), t \right) \\ &\quad \Phi_i : \mathbb{R} \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+, \quad \gamma_{ij} : \mathbb{R}_+ \times Y \longrightarrow \mathbb{R}, \\ &\quad N_j : \mathcal{B}(\mathbb{R}_+ \times Y) \longrightarrow \mathbb{N} \end{aligned} \quad (7.2)$$

<sup>4</sup>Notice that some authors use  $\gamma_{ji}$ , so that the first index is the source type and the second index is the destination type.

<sup>5</sup>Hawkes process only specifies the intensity without any restriction on the mark distribution.

where  $w_n \in \{1, \dots, d\}$  denotes the type of  $t_n$  and  $\Phi_i$  is known as rate function. Consider the special case

$$\begin{aligned}\lambda_i(t) &= \mu_i(t) + \sum_{j=1}^d \int_{(-\infty, t) \times Y} \gamma_{ij}(t-s, y) N_j(ds \times dy) = \mu_i(t) \\ &\quad + \sum_{t_n < t} \gamma_{i,w_n}(t - t_n, y_n) \\ \mu_i : \mathbb{R}_+ &\longrightarrow \mathbb{R}_+, \quad \gamma_{ij} : \mathbb{R}_+ \times Y \longrightarrow \mathbb{R}_+, \\ N_j : \mathcal{B}(\mathbb{R}_+ \times Y) &\longrightarrow \mathbb{N}\end{aligned}\tag{7.3}$$

that is,  $\Phi_i(x, t) = \mu_i(t) + x$ . Such a Hawkes process determined by (7.3) is called linear, and  $\mu_i(t)$  is called the base or background rate. The function  $\gamma_{ij}$  is called (marked) decay/exciting/fertility kernel and often  $\gamma_{ij}(t, y)$  takes the separable form  $\gamma_{ij}(t)g_{ij}(y)$ , where  $g_{ij}$  is called mark impact kernel. Popular choices of decay kernel  $\gamma_{ij}(t)$  include exponential  $\alpha_{ij}e^{-\beta_{ij}t}$  [13], power law  $\alpha_{ij}(c_{ij} + t)^{-\beta_{ij}}$  [14], or Laguerre-type polynomial  $\sum_{k=0}^K \alpha_{ijk}t^k e^{-\beta_{ijk}t}$  [15].

If the decay function is exponential with  $\beta_{ij} = \beta_i$ , the intensity  $\lambda(t)$  and the vector  $(N(t), \lambda(t))$  are both Markov processes<sup>6</sup> [16, 17]. Moreover, provided that  $\mu_i(t) = \mu_i$ , then  $(\lambda_1(t), \dots, \lambda_d(t))$  satisfies the system of stochastic differential equations

$$d\lambda_i(t) = \beta_i(\mu_i - \lambda_i(t))dt + \sum_{j=1}^d \alpha_{ij}dN_j(t)\tag{7.4}$$

This specification has the simple interpretation that the events of  $N_j$ , which happened just before time  $t$ , increase the intensity  $\lambda_i(t)$  by  $\alpha_{ij} \geq 0$  and thus trigger further events. Yet if the intensity  $\lambda_i(t)$  is higher than  $\mu_i$ , the first term becomes negative ( $\beta_i > 0$ ) and draws the intensity back to the equilibrium level  $\mu_i$ . In other words, the intensity  $\lambda_i(t)$  is a mean-reverting process driven by its own point process. The Markov property and this intuitive interpretation may explain why the exponential decay kernels are so widely used.

For linear Hawkes processes,  $\mu_i(t)$ ,  $\gamma_{ij}(t)$ , and  $g_{ij}(t)$  must be nonnegative for all  $t$ , in order to ensure the positivity of  $\lambda_i(t)$ . As a result, unlike nonlinear Hawkes processes, linear Hawkes processes cannot model inhibitory effect (negative excitation). Nonetheless, the linear Hawkes processes are easier

---

<sup>6</sup> $N$  itself is not a Markov process as its intensity at time  $t$  depends on its full history before time  $t$ .

to handle, their properties are better understood and, most importantly, they have a branching structure representation, which is extremely useful in simulation, estimation, and interpretation of the models.

### 7.3.1 BRANCHING STRUCTURE REPRESENTATION

Linear Hawkes processes have a very elegant branching structure representation [18]. We describe here the version for the multivariate Hawkes processes with unpredictable marks [19].

There are  $d$  types of immigrants arriving according to Poisson processes with rates  $\mu_1, \dots, \mu_d$ . Each individual (descendant or immigrant) will carry an unpredictable mark when born or arrived. An individual of type  $j$  born at time  $t_n$  with mark  $y_n$  will give birth to an individual of type  $i$  according to a nonhomogeneous Poisson process with rate  $\gamma_{ij}(t - t_n, y_n)$ . All the nonhomogeneous Poisson processes are independent of one another.

Let  $N_i(t)$  be the total number of individuals of type  $i$  born/arrived at or before time  $t$  under the above scenario, then  $N(t) = (N_1(t), \dots, N_d(t))$  will follow the linear marked Hawkes process (7.3). This representation forms the basis of the Expectation Maximization (EM) algorithm in Section 7.4.2 and we will also see how it is used to measure the endogeneity of a point process in Section 7.5.4.

### 7.3.2 STATIONARITY

Considering a Hawkes process  $N$  with intensity (7.2) such that  $\Phi_i(x, t) = \Phi_i(x)$ ,  $N$  has an unique stationary version<sup>7</sup> if either of the following conditions is satisfied [18, 20]:

1.  $\Phi_i(x)$  is  $k_i$ -Lipschitz<sup>8</sup> and the spectral radius<sup>9</sup>  $\rho(A) < 1$  for the  $d \times d$  matrix  $A = [k_i \int_0^\infty |\gamma_{ij}(t)| dt]_{i,j}$ .
2.  $\Phi_i(x)$  is Lipschitz,  $\Phi_i(x) \leq M$ ,  $\int_0^\infty |\gamma_{ij}(t)| dt < \infty$  and  $\int_0^\infty t |\gamma_{ij}(t)| dt < \infty$ .

Technically speaking,  $N$  may have other nonstationary versions together with the stationary one; however, the nonstationary version will converge weakly to the stationary version when  $t \rightarrow \infty$  (see [21] for exact meaning). Since the Hawkes process starts at  $-\infty$ ,  $N((0, t])$  will have the stationary distribution for all  $t > 0$ .

<sup>7</sup>See Appendix 7.A for definition of stationarity of point processes.

<sup>8</sup> $f : \mathbb{R} \rightarrow \mathbb{R}$  is called  $k$ -Lipschitz ( $k > 0$ ) if  $|f(x) - f(y)| \leq k|x - y| \forall x, y \in \mathbb{R}$ .

<sup>9</sup> $\rho(A) = \max_i\{|\pi_i|\}$ ,  $\{\pi_i\}$  are eigenvalues of  $A$ .

For the case of an exponential decay kernel  $\alpha_{ij}e^{-\beta_{ij}t}$ , we have a simpler result. Let  $A = [\int_0^\infty \alpha_{ij}e^{-\beta_{ij}t} dt]_{i,j} = [\alpha_{ij}/\beta_{ij}]_{i,j}$ , then  $N$  has a unique stationary version under either of the following conditions [22]:

1.  $\Phi_i(x) = \mu_i + x$ ,  $\alpha_{ij} \geq 0$ ,  $\beta_{ij}, \mu_i > 0$ ,  $\rho(A) < 1$  (linear Hawkes process).
2.  $\Phi_i(x) = \max(\mu_i + x, \varepsilon_i)$ ,  $\alpha_{ij} \in \mathbb{R}$ ,  $\beta_{ij}, \mu_i > 0$ ,  $\varepsilon_i > 0$ ,  $\rho(A) < 1$  (T-Hawkes process).
3.  $\Phi_i(x) = \min(\mu_i + \exp(x), M_i)$ ,  $\alpha_{ij} \in \mathbb{R}$ ,  $\beta_{ij} > 0$ ,  $M_i > \mu_i > 0$  (E-Hawkes process).

For the univariate linear case with  $\mu = 0$ , if there exists  $r, R > 0, c \in (0, 1/2)$  such that  $\int_0^\infty \gamma(t)dt = 1$ ,  $\sup_{t \geq 0} t^{1+c}\gamma(t) \leq R$ ,  $\lim_{t \rightarrow \infty} t^{1+c}\gamma(t) = r$ , Brémaud and Massoulié [23] show that there exists a unique stationary nontrivial Hawkes process having such an intensity and he calls it the critical Hawkes process or Hawkes process without ancestors ( $\mu = 0$ ).

### 7.3.3 CONVERGENCE

In this section, we state the various results about the convergence of Hawkes processes. A properly scaled linear Hawkes process will converge weakly to a Brownian diffusion when the spectral radius of decay functions'  $L^1$ -norm is less than 1 [24]. When the spectral radius is close to 1 in a certain sense, it converges to the integrated Cox–Ingersoll–Ross (CIR) process [25]. For the nonlinear Hawkes processes, we have only the result for the univariate case and the sufficient conditions depend on the Lipschitz constant of  $\Phi$  [26].

#### 7.3.3.1 Law of large numbers for multivariate linear Hawkes processes

Assuming the model (7.3) without marks, if the spectral radius  $\rho(A) < 1$ , where  $A = [\int_0^\infty \gamma_{ij}(t)dt]_{i,j}$ , then [24]

$$\sup_{t \in [0,1]} \left\| \frac{N(nt)}{n} - t(I_d - A)^{-1}\mu \right\| \xrightarrow[n \rightarrow \infty]{\text{a.s./}L^2} 0^{10,11} \quad (7.5)$$

<sup>10</sup>A sequence of random variables  $X_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} X$  if  $\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1$ .

<sup>11</sup>A sequence of random variables  $X_n \xrightarrow[n \rightarrow \infty]{L^2} X$  if  $\lim_{n \rightarrow \infty} E(|X_n - X|^2) = 0$ .

where  $\mu = (\mu_1, \dots, \mu_d)$ . When  $d = 1$  and we take  $t = 1$ , it implies

$$\frac{N(T)}{T} \xrightarrow[T \rightarrow \infty]{\text{a.s./}L^2} \frac{\mu}{1 - \int_0^\infty \gamma(t)dt} \quad (7.6)$$

### 7.3.3.2 Functional central limit theorem for multivariate linear Hawkes processes

Assuming the model (7.3) without marks,  $N = (N_1, \dots, N_d)$ , if the spectral radius  $\rho(A) < 1$ , where  $A = [\int_0^\infty \gamma_{ij}(t)dt]_{i,j}$  and  $\int_0^\infty \sqrt{t}\gamma_{ij}(t)dt < \infty \forall i, j$ , then [24]

$$\sqrt{n} (N(\bullet n)/n - \bullet(I_d - A)^{-1}\mu) \xrightarrow[n \rightarrow \infty]{\text{weak}} (I_d - A)^{-1}\Sigma^{1/2}W(\bullet)^{12} \quad (7.7)$$

$$\Sigma = \text{diag}((I_d - A)^{-1}\mu)^{13}, \quad W \text{ is standard } d\text{-dimensional Brownian Motion}$$

(7.8)

### 7.3.3.3 Functional central limit theorem for univariate nonlinear Hawkes processes

Assuming the model (7.2) without marks and  $d = 1$ , if  $\gamma(t)$  is decreasing,  $\int_0^\infty t\gamma(t)dt < \infty$ ,  $\Phi(x, t) = \Phi(x)$  is increasing and  $k$ -Lipschitz,  $\int_0^\infty k\gamma(t)dt < 1$ , then [26]

$$\sqrt{n} (N(\bullet n)/n - \bullet v) \xrightarrow[n \rightarrow \infty]{\text{weak}} \sigma W(\bullet) \quad (7.9)$$

$$\begin{aligned} \sigma^2 &= \mathbb{E}((N([0, 1]) - v)^2) + 2 \sum_{n=1}^{\infty} \mathbb{E}((N([0, 1]) - v) \\ &\quad \times (N([n, n+1]) - v), \quad v = \mathbb{E}(N([0, 1])) \end{aligned} \quad (7.10)$$

<sup>12</sup>A sequence of probability measure  $P_n$  converges weakly to  $P$  if  $\int_{\Omega} fdP_n \rightarrow \int_{\Omega} fdP$  for all bounded continuous function  $f$ . A sequence of stochastic process  $X_n : \Omega \rightarrow D[0, 1]$  converges weakly (in distribution) to  $X$  if the law of  $X_n(P_n \circ X_n^{-1})$  converges weakly to law of  $X(P \circ X^{-1})$  in the sense of probability measure,  $D[0, 1]$  is the Skorokhod space of càdlàg (right continuous with left limits) functions (see [27, 28]).

<sup>13</sup> $v \in \mathbb{R}^d$ ,  $\text{diag}(v) = [a_{ij}]_{d \times d}$ ,  $a_{ii} = v_i$ ,  $a_{ij} = 0 \forall i \neq j$ .

### 7.3.3.4 Convergence of nearly unstable univariate linear Hawkes processes

Considering the linear model (7.3) without marks and  $d = 1$ ,  $N(T)/T$  converges to  $\mu/(1 - \int_0^\infty \gamma(t)dt)$ , when  $\int_0^\infty \gamma(t)dt < 1$  by (7.6), while it explodes when  $\int_0^\infty \gamma(t)dt = 1$ . However, Jaisson and Rosenbaum [25] find that the properly scaled Hawkes process converges to the integrated CIR process when one has a sequence of decay kernel  $\gamma^{(n)}(t)$  whose integral converges to 1 at the speed of  $n^{-1}$  (see 7.13). More precisely, let

$$\lambda^{(n)}(t) = \mu + \int_{(0,t)} \gamma^{(n)}(t-s)dN^{(n)}(s), \quad \mu > 0, \quad \gamma^{(n)}(t) = \alpha^{(n)}\gamma(t) \quad (7.11)$$

$$\begin{aligned} \gamma : \mathbb{R}_+ &\longrightarrow \mathbb{R}_+, & \int_0^\infty \gamma(t)dt = 1, & \int_0^\infty t\gamma(t)dt = m < \infty, \\ & \int_0^\infty |\gamma'(t)|dt < \infty, & \sup_{t \in [0,\infty)} |\gamma'(t)| < \infty \end{aligned} \quad (7.12)$$

$$\alpha^{(n)} \in [0, 1), \quad \lim_{n \rightarrow \infty} \alpha^{(n)} = 1, \quad \lim_{n \rightarrow \infty} n(1 - \alpha^{(n)}) = c > 0 \quad (7.13)$$

$$\psi^{(n)}(t) = \sum_{k=1}^{\infty} \gamma^{(n) \otimes k}(t), \quad \rho^{(n)}(t) = \frac{n\psi^{(n)}(nt)}{\int_0^\infty \psi^{(n)}(t)dt}, \quad |\rho^{(n)}(t)| \leq M \quad \forall n \quad \forall t \quad (7.14)$$

where  $f^{\otimes k}$  denotes the  $k$ -fold self-convolution of  $f$ . If the sequence of Hawkes process  $N^{(n)}$  has intensity  $\lambda^{(n)}$  satisfying (7.11–7.14), then the scaled intensity converges to the CIR process and the scaled Hawkes process converges to the integrated CIR process [25] as follows:

$$(1 - \alpha^{(n)})\lambda^{(n)}(n\bullet) \xrightarrow{\text{weak}} X(\bullet) \quad (7.15)$$

$$(1 - \alpha^{(n)})\frac{N^{(n)}(n\bullet)}{n} \xrightarrow[n \rightarrow \infty]{\text{weak}} \int_0^\bullet X(s)ds \quad (7.16)$$

$$dX_t = \frac{c}{m}(\mu - X_t)dt + \frac{\sqrt{c}}{m}\sqrt{X_t}dW_t, \quad X_0 = 0 \quad (7.17)$$

## 7.4 Statistical inference of Hawkes processes

### 7.4.1 SIMULATION

In this section, we give an overview of the algorithms that simulate Hawkes processes. Assume that we know all the parameters in the functional form

of  $\mu(t)$  and  $\gamma(t, y)$ , our goal is to simulate the points  $(t_1, y_1), (t_2, y_2), \dots$  on the interval  $[0, T]$ .

If the marks distribution depends only on  $t_n$ , we can simply generate  $y_n$  conditioned on the generated  $t_n$ . Next,  $t_{n+1}$  can be generated from the intensity  $\lambda(t)$  for  $t > t_n$ , which depends on  $\{(t_1, y_1), \dots, (t_n, y_n)\}$ . If the distribution of  $y_n$  also depends on  $\{(t_{n-1}, y_{n-1}), (t_{n-2}, y_{n-2}), \dots\}$ , the algorithms can be modified accordingly.

#### 7.4.1.1 Inverse CDF transform

The first simulation algorithm for Hawkes processes appears in Ozaki [29]. Suppose that the intensity is governed by the univariate Hawkes model in (7.3). Let  $t_n$  be the arrival time and  $\tau_n = t_n - t_{n-1}$  be the interarrival time. By (7.63),  $\lambda(t) = h_n(t - t_{n-1})$  for  $t \in (t_{n-1}, t_n]$ , where  $h_n(t) = g_n(t)/(1 - G_n(t^-))$  and  $g, G$  are the conditional pdf, cdf of  $\tau_n$  given  $\mathcal{F}_{t_{n-1}}$ . If  $G_n(t)$  is continuous,  $h_n(t)$  is simply the hazard function, and it can be shown that

$$G_n(\tau_n) = 1 - \exp \left( - \int_0^{\tau_n} h_n(s) ds \right) = 1 - \exp \left( - \int_{t_{n-1}}^{t_{n-1} + \tau_n} \lambda(s) ds \right) \quad (7.18)$$

Given  $t_{n-1}$ , we can generate  $t_n = t_{n-1} + \tau_n$  by inverse cdf transform  $\tau_n = G_n^{-1}(U)$ ,  $U \sim \text{Unif}(0, 1)$ . However, the inversion needs to be done numerically, so this method is largely superceded by Ogata's modified thinning that we now discuss.

#### 7.4.1.2 Ogata's modified thinning

Ogata [30] introduces the modified thinning method that does not require numerical inversion. The algorithm is based on the following theorem. Let  $N = (N_1, \dots, N_d)$  be a multivariate point process with intensity  $(\lambda_1, \dots, \lambda_d)$  such that  $\sum_{i=1}^d \lambda_i(t) \leq \lambda^*(t)$   $\forall t$  a.s. ( $\lambda^*(t)$  is an exogenously chosen deterministic rate function) and  $N^*$  is the univariate nonhomogeneous Poisson process with intensity  $\lambda^*(t)$ . If each point  $t_n$  in  $N^*$  is given a mark  $y_n$  such that  $\mathbb{P}(y_n = i) = \lambda_i(t_n)/\lambda^*(t_n)$ ,  $i = 1, \dots, d$ , then  $(N_1^*, \dots, N_d^*)$  has the same distribution as  $(N_1, \dots, N_d)$ .

The following algorithm generates a  $d$ -dimensional multivariate Hawkes process such that  $\lambda_i(t)$  is decreasing between points and  $|\lambda_i(t) - \lambda_i(t^-)| \leq \alpha_i \forall t$ .

### Ogata's Modified Thinning [30]

1.  $n = 1, t_0 = 0.$
2. Generate  $\tau_n \sim \text{Exp}(\lambda_n^*)$  for some  $\lambda_n^* \geq \sum_{i=1}^d (\lambda_i(t_{n-1}) + \alpha_i).$
3. Let  $t_n = t_{n-1} + \tau_n.$
4. Generate  $U_n \sim \text{Unif}(0, 1).$
5. If  $U_n \in (\sum_{i=1}^{k-1} \lambda_i(t_n)/\lambda_n^*, \sum_{i=1}^k \lambda_i(t_n)/\lambda_n^*]$  for some  $k \in \{1, \dots, d\}$  return  $t_n$  and the point is of type  $k$ , else discard  $t_n$  (but keep the value for use in next-generation steps 2 and 3).
6.  $n = n + 1$ , go to step 2.

#### 7.4.1.3 Simulation by branching structure

This method generates points using the branching structure representation of linear marked Hawkes processes. Type- $j$  immigrants arrive according to a nonhomogeneous Poisson process with rate  $\mu_j(t)$ . Next, the type- $j$  parent arriving at  $t_n$  with mark  $y_n$  produces type- $i$  descendants according to a nonhomogeneous Poisson process with rate  $\gamma_{ij}(t - t_n, y_n)$  and the generation is repeated for each descendant until all of them exceed the predefined time  $T$ . Since all the nonhomogeneous Poisson processes are independent, the generations can be done in parallel.

#### *Simulation by the Branching Structure [31]*

1. Generate nonhomogeneous Poisson processes with intensities  $\mu_i(t), i = 1, \dots, d$  on  $[0, T]$ .
2. For each points  $t_n$ , generate  $y_n|t_n$ .
3. Suppose that  $t_n$  is of type  $j$  and generates type- $i$  descendants according to nonhomogeneous Poisson process with intensity  $\gamma_{ij}(t - t_n, y_n)$  on  $[t_n, T], i = 1, \dots, d$ .
4. Repeat steps 2 and 3 for all descendants.

The nonhomogeneous Poisson process with intensity  $\mu(t)$  on  $[0, T]$  can be generated using Lewis' thinning algorithm [32]:

1. Generate  $N \sim \text{Poisson}(\mu^*)$  for some  $\mu^* \geq \max_{t \in (0, T]} \mu(t).$
2. Generate  $U_n \sim \text{Unif}(0, 1), n = 1, \dots, N.$
3.  $T_n = U_{(n)}T, n = 1, \dots, N$  ( $\{U_{(n)}\}$  is the order statistics of  $\{U_n\}$ ).
4. Generate  $V_n \sim \text{Unif}(0, 1), n = 1, \dots, N.$
5. Return  $T_n$  if  $V_n \leq (\mu(T_n)/\mu^*), n = 1, \dots, N$ , otherwise discard  $T_n$ .

## 7.4.2 ESTIMATION

Suppose that we observe a point process on  $(0, T]$  and collect the event times and marks  $\{(t_1, y_1), \dots, (t_N, y_N)\}$ , and now we would like to estimate the functions  $\mu(t)$  and  $\gamma(t, y)$  in the intensity  $\lambda(t)$ , which drives the process  $N(t)$ . We will summarize the various methods appearing in the literature, but so far the focus is on unmarked processes. In the special case where the marks are independent and identically distributed, the mark distribution can be estimated separately from the point process.

If we assume that  $\mu(t)$  and  $\gamma(t)$  have some parametric representations, we can use Maximum Likelihood Estimation (MLE), EM, or Generalized Method of Moments (GMMs) to estimate the parameters. Otherwise, we need to rely on some advanced nonparametric techniques to estimate the whole function curves.

### 7.4.2.1 Maximum likelihood estimation

The log-likelihood of a Hawkes process is given by [29]

$$\log(L(\theta)) = \sum_{i=1}^d \left( - \int_0^T \lambda_i(t; \theta) dt + \int_0^T \log(\lambda_i(t; \theta)) dN_i(t) \right) \quad (7.19)$$

In the case of multivariate linear Hawkes process, it becomes

$$\begin{aligned} \log(L(\theta)) = & - \int_0^T \left( \sum_{i=1}^d \mu_i(t; \theta) \right) dt - \sum_{n=1}^N \int_{t_n}^T \left( \sum_{i=1}^d \gamma_{i,w_n}(t - t_n; \theta) \right) dt \\ & + \sum_{n=1}^N \log \left( \mu_{w_n}(t_n; \theta) + \sum_{t_m < t_n} \gamma_{w_n, w_m}(t_n - t_m; \theta) \right) \end{aligned} \quad (7.20)$$

The parameters  $\theta$  can be estimated by maximizing the log-likelihood. However, the numerical optimization is problematic as the log-likelihood function is usually quite flat (see [33, Figure 2,3]) and may have a lot of local maxima (see, [33, Figure 4]).

### 7.4.2.2 Expectation maximization

For linear Hawkes process, the estimation can also be done via EM [34, 35] as in [33, 36–39]. EM is a variant of MLE where part of the data is missing. In the branching structure representation, the missing data is the parents that produce the descendants. Let  $z_n$  denote the index of the parent

of  $t_n$  and  $w_{z_n}$  represents the type of the parent of  $t_n$ . If  $z_n = m$  and  $w_{z_n} = j$ , that means  $t_n$  is produced by the type  $j$  point  $t_m$ . When  $z_n$  is 0,  $t_n$  is an immigrant. Also, we define  $w_0 = 0$ ,  $\gamma_{i,0}(t) = \mu_i(t)$ , and  $t_0 = 0$  to simplify the expression. Suppose that  $\{t_n, w_n, z_n\}$  are known, since each generation is an independent Poisson process, the complete data log-likelihood is

$$\begin{aligned}\log(L(\theta)) &= \sum_{n=0}^N \sum_{i=1}^d \left\{ - \int_{t_n}^T \gamma_{i,w_n}(t - t_n; \theta) dt \right. \\ &\quad \left. + \sum_{t_m > t_n} \log(\gamma_{i,w_n}(t_m - t_n; \theta)) \mathbb{1}(z_m = n) \mathbb{1}(w_m = i) \right\} \end{aligned}\quad (7.21)$$

$$\begin{aligned}Q(\theta | \theta^{(k)}) &= \mathbb{E}^{\theta^{(k)}}(\log(L(\theta)) | \{(t_k, w_k)\}) \\ &= \sum_{n=0}^N \sum_{i=1}^d \left\{ - \int_{t_n}^T \gamma_{i,w_n}(t - t_n; \theta) dt + \sum_{t_m > t_n} \log(\gamma_{i,w_n}(t_m - t_n; \theta)) \right. \\ &\quad \left. \mathbb{P}^{\theta^{(k)}}(z_m = n | \{(t_k, w_k)\}) \mathbb{1}(w_m = i) \right\} \end{aligned}\quad (7.22)$$

$$\mathbb{P}^{\theta^{(k)}}(z_m = n | \{(t_k, w_k)\}) \mathbb{1}(w_m = i) = \frac{\gamma_{i,w_n}(t_m - t_n; \theta^{(k)}) \mathbb{1}(w_m = i)}{\sum_{l=0}^{m-1} \gamma_{i,w_l}(t_m - t_l; \theta^{(k)})} \quad (7.23)$$

The EM algorithm can be implemented as follows:

1.  $k = 0$  and choose an initial guess  $\theta^{(0)}$ .
2. E-step: compute  $Q(\theta | \theta^{(k)}) = \mathbb{E}^{\theta^{(k)}}(\log(L(\theta)) | \{(t_k, w_k)\})$ .
3. M-step: compute  $\theta^{(k+1)} = \text{argmax}_\theta Q(\theta | \theta^{(k)})$ .
4.  $k = k + 1$ , repeat E-step and M-step until  $\theta^{(k)}$  converges (e.g.,  $||\theta^{(k+1)} - \theta^{(k)}|| < \varepsilon$ ).

In general, the optimization in M-step needs to be solved numerically but when the decay kernel has the exponential form  $\alpha_{ij}\beta_{ij} \exp(-\beta_{ij}t)$ , Olson and Carley [39] suggest a closed form approximate iteration.

$$\mu_i^{(k+1)} = \frac{\sum_{m=1}^N \mathbb{P}^{\theta^{(k)}}(z_m = 0 | \{(t_k, w_k)\}) \mathbb{1}(w_m = i)}{\sum_{n=1}^N \sum_{m=n+1}^N \mathbb{P}^{\theta^{(k)}}(z_m = n | \{(t_k, w_k)\}) \mathbb{1}(w_m = i, w_n = j)} \quad (7.24)$$

$$\alpha_{ij}^{(k+1)} = \frac{\sum_{n=1}^N \sum_{m=n+1}^N \sum_{l=1}^N \mathbb{P}^{\theta^{(k)}}(z_m = n | \{(t_k, w_k)\}) \mathbb{1}(w_m = i, w_n = j) \mathbb{1}(w_l = j)}{\sum_{n=1}^N \sum_{m=n+1}^N (t_m - t_n) \mathbb{P}^{\theta^{(k)}}(z_m = n | \{(t_k, w_k)\}) \mathbb{1}(w_m = i, w_n = j)} \quad (7.25)$$

$$\beta_{ij}^{(k+1)} = \frac{\sum_{n=1}^N \sum_{m=n+1}^N \sum_{l=1}^N \mathbb{P}^{\theta^{(k)}}(z_m = n | \{(t_k, w_k)\}) \mathbb{1}(w_m = i, w_n = j) \mathbb{1}(w_l = j)}{\sum_{n=1}^N \sum_{m=n+1}^N (t_m - t_n) \mathbb{P}^{\theta^{(k)}}(z_m = n | \{(t_k, w_k)\}) \mathbb{1}(w_m = i, w_n = j)} \quad (7.26)$$

$$\begin{aligned} & \mathbb{P}^{\theta^{(k)}}(z_m = n | \{(t_k, w_k)\}) \mathbb{1}(w_m = i, w_n = j) \\ &= \frac{\alpha_{ij}^{(k)} \beta_{ij}^{(k)} \exp(-\beta_{ij}^{(k)}(t_m - t_n)) \mathbb{1}(w_m = i, w_n = j)}{\mu_i^{(k)} + \sum_{l=1}^{m-1} \alpha_{i,w_l}^{(k)} \beta_{i,w_l}^{(k)} \exp(-\beta_{i,w_l}^{(k)}(t_m - t_l))} \end{aligned} \quad (7.27)$$

$$\begin{aligned} & \mathbb{P}^{\theta^{(k)}}(z_m = 0 | \{(t_k, w_k)\}) \mathbb{1}(w_m = i) \\ &= \frac{\mu_i^{(k)} \mathbb{1}(w_m = i)}{\mu_i^{(k)} + \sum_{l=1}^{m-1} \alpha_{i,w_l}^{(k)} \beta_{i,w_l}^{(k)} \exp(-\beta_{i,w_l}^{(k)}(t_m - t_l))} \end{aligned} \quad (7.28)$$

In addition, the summation  $\sum_{l=1}^{m-1} \alpha_{i,w_l}^{(k)} \beta_{i,w_l}^{(k)} \exp(-\beta_{i,w_l}^{(k)}(t_m - t_l))$  can be truncated after  $\exp(-\beta_{i,w_l}^{(k)}(t_m - t_l))$  has decayed to a small value. The speed of EM is reported to be 10–100 times faster than MLE and, more importantly, MLE does not converge within 500 iterations in practically all test cases while EM does [39].

#### 7.4.2.3 Generalized method of moments

Another method for statistical estimation apart from MLE is the *Generalized Method of Moments*<sup>14</sup> [40]. The idea is to find the parameters that minimize the difference between theoretically moments (see Appendix 7.A) in terms of the unknown parameters and the empirical moments computed directly from the data. If we have more moments than the number of parameters, the method involves solving a weighted least squares problem.

Da Fonseca and Zaatour [41] obtain the analytic moment expressions by restricting the process to be univariate with exponential kernel and making use of the Markov property in this special case. The authors claim that this method is extremely fast but no speed comparison result is provided.

#### 7.4.2.4 Nonparametric estimation

Without assuming any parametric form for  $\mu(t)$  nor  $\gamma(t)$ , some nonparametric methods are developed recently to estimate the whole base rate and decay kernel functions. Similar to parametric estimation, penalized MLE or GMM is used to find the function with desirable characteristics

---

<sup>14</sup>Although GMM is consistent under some mild regularity conditions, unlike MLE, it is not asymptotic efficient among the class of consistent estimators.

(e.g., smooth functions, sparse coefficients). Nonetheless, the nonparametric method, which involves finding the unknown functions in infinite-dimensional spaces, requires extensive computational effort and the underlying statistical construction is usually much more involved than the parametric counterpart.

To the best of our knowledge, the first attempt in nonparameteric estimation of Hawkes process is by Gusto and Schbath [42] in 2005. The authors express the kernel function of the multivariate Hawkes process, using B-splines [43] with equally spaced knots. The log-likelihood function involving the basis coefficients is then maximized numerically and the optimal order for the B-splines basis, as well as number of knots, is determined using AIC criteria [44].

Instead of B-splines, Reynaud-Bouret and Schbath [45] find the function within the space of piecewise constant functions that minimizes the empirical  $L^2$ -norm between the true and estimated kernel functions. The method is later extended to multivariate cases [46] with a Lasso-type penalty [47] in the minimization objective.

Instead of EM, Zhou et al. [48] use Minorize-Maximization (MM) algorithm [49], in which EM is a special case. In the E-step of MM algorithm,  $Q(\bullet|\theta^{(k)})$  is any lower bound of the objective function  $\log(L(\bullet))$  such that  $Q(\theta^{(k)}|\theta^{(k)}) = \log(L(\theta^{(k)}))$ . It is then iteratively maximized in the M-steps until convergence. In [48], the kernel functions are expressed using a finite number of basis functions that are estimated nonparametrically in M-step by solving the Euler–Lagrange equation.

Another approach is to use moment matching to find the kernel function as in [50–52]. Bacry and Muzy [52] derive the conditional moment density  $\mathbb{E}(dN_i(t)|dN_j(0) = 1, dy)$  of multivariate marked Hawkes process as the solution of Wiener–Hopf equation [53] involving  $\mu_i, \gamma_{ij}(t), g_{ij}(y)$  for the case that the mark impact kernel is piecewise constant. The conditional moment density can be estimated by any kernel density estimation technique and the Wiener–Hopf equation can be solved numerically via the Nyström method [54].

## 7.4.3 HYPOTHESIS TESTING

### 7.4.3.1 Random time change

The classical method to test the goodness of fit of a point process model on  $\mathbb{R}_+$  is Ogata's residual analysis [14]. Ogata calls  $\{\tilde{t}_n = \int_0^{t_n} \hat{\lambda}(s)ds\}$  the

residual process<sup>15</sup> and according to the random time change theorem (see Appendix 7.A), the residual process should be close to a standard Poisson process if the estimated intensity  $\hat{\lambda}(t)$  is close to the true intensity  $\lambda(t)$ . The hypothesis that  $\{\tilde{t}_n\}$  is a standard Poisson process can be tested by the following methods:

1. QQ Plot [56] of  $\{\tilde{\tau}_n = \tilde{t}_n - \tilde{t}_{n-1}\}$  vs  $\text{Exp}(1)$ .
2. Kolmogorov–Smirnov Test [57–59] to test  $\tilde{\tau}_n \sim \text{Exp}(1)$ .
3. Ljung–Box Test [60] to test the lack of serial correlation of  $\{\tilde{\tau}_n\}$ .

#### 7.4.3.2 Approximate thinning

Another method to test goodness of fit is by thinning, which does not require integration of the intensity function. It is useful if the intensity function is estimated nonparametrically. However, the thinned residual process is only approximately a Poisson process.

By Ogata's modified thinning [30], we know that if there exists  $b > 0$  such that  $b \leq \lambda(t) \forall t$  and we keep point  $t_n$  with probability  $b/\lambda(t_n)$ , the thinned point process is a homogeneous Poisson process with rate  $b$ . However, the infimum  $b$  of the intensity function is often close to 0, making the number of points in the thinned process very small and the test to have little power. A remedy is to use approximate thinning [61] as follows: choose an integer  $k \ll N$  and select one point from  $\{t_1, \dots, t_N\}$  with probability of selecting  $t_n$  proportional to  $\lambda(t_n)^{-1}$ . Repeat the selection (without replacement) until  $k$  points are selected. The resulting  $k$  points will be approximately a homogeneous Poisson process.

## 7.5 Applications of Hawkes processes

After the groundwork of basic theory and statistical inference for Hawkes processes, we now unleash their power to model HF data. First, the readers are reminded how diverse the notion of stock trading frequency can be. According to the Trade And Quote database, between 9:30 a.m. and 4:00 p.m. on May 2, 2014, there were 11 million quote changes (limit + cancellation + market orders) and 0.3 million trades (market orders) for

---

<sup>15</sup>The terminology is not standard. Baddeley et al. [55] refer  $\{N(t_n) - \int_0^{t_n} \hat{\lambda}(s)ds\}$  as residual to extend the concept to higher dimension.

SPDR S&P 500 ETF (SPY). In other words, on average, there are 460 quote changes and 13 trades per second. If we take a snapshot every 5 min as in [4, 5], we will use only 0.03% of trade data and 0.0007% of quote data. In comparison, Pathfinder Bancorp has only 306 quote changes and 11 trades on the the same day, which means that there is a 35-min lag between trades on average and thus the 5-min snapshots will just give a series of repeated information. Regardless of the sampling frequency, we are likely to get some misleading result if we analyze the asynchronous data from a portfolio of liquid and illiquid stocks using models with fixed intervals.

The construction of multivariate point processes shows that each variate can have a completely different arrival intensity  $\lambda_i(t)$ . Nonetheless, the multivariate Hawkes process can still model the dependence structure easily via the  $\gamma_{ij}(t)$ 's, which are estimated by duly considering all the asynchronous data in the highest frequency without any resampling.

Order arrivals and price changes are unarguably two of the most important elements in HF trading. Using Hawkes processes, we can estimate their conditional distributions based on all the historical HF asynchronous data, enabling us to give a more accurate real-time prediction of future event occurrences. In the following subsections, we are going to highlight some of the literature that takes advantage of Hawkes processes to model HF data.

### 7.5.1 MODELING ORDER ARRIVALS

Bowsher [22]<sup>16</sup> is the first to use Hawkes processes to model order arrivals. He uses nonlinear Hawkes processes to allow for inhibitory effect and he considers two rate functions  $\Phi_i(x, t) = \mu_i(t) + \exp(x)$  and  $\Phi_i(x, t) = \max(\mu_i(t) + x, \varepsilon_i)$ ,  $\varepsilon_i > 0$ , where both of them guarantee that the stochastic intensity will be strictly positive at all times. For the deterministic base rate  $\mu_i(t)$ , he exploits a piecewise linear function with knots at 9:30, 10:00, 11:00, ..., 16:00 while the decay kernel is the exponential function without marks. In addition, an extra term is included to represent the spillover effects from the previous trading day.

Bowsher uses MLE to estimate the parameters for the bivariate point process of trade and quote of General Motor, trading on NYSE between

---

<sup>16</sup>Although Bowsher's paper was published in 2007, the first draft appeared in 2002.

July 5, 2000, and August 29, 2000. The model is found to be decent according to the goodness-of-fit test using random time change.

Instead of modeling arrivals of all trades and quotes, Large [62] uses Hawkes processes to model only the arrivals of aggressive orders, which are market orders depleting the queue and limit orders falling inside the bid-ask spread, in order to study the *resiliency* of the LOB. An LOB is called resilient if it reverts to its generic shape promptly after large trades. The idea is that when a large trade causes the bid-ask spread to widen, the arrival intensity of aggressive limit orders in a resilient LOB will surge so that the gap will be filled very quickly. In other words, the cross-excitation effect  $\gamma_{ij}(t)$  from aggressive market orders to aggressive limit orders should be reasonably large in a resilient LOB.

In addition to market orders and limit orders, Large also includes the cancellations of limit orders as well as limit orders falling outside the best quotes. Therefore, he builds a 10-variate linear marked Hawkes process with exponential decay and mark impact kernel to fit the HF data of Barclays (BARC), trading on LSE between January 2 and 31, 2002. The result shows that the widening of bid-ask spread indeed pumps up the intensities of aggressive limit orders, causing the gap to be filled very quickly and hence making the LOB resilient.

More examples of applications of Hawkes processes to order arrivals include the following papers: Muni Toke and Pomponio [63] use a similar approach as Large [62] to model trades-through, namely market orders that deplete the best queues and consume at least one share in the second best. Muni Toke [64] designs a more realistic market simulator, using Hawkes processes with exponential kernel for order arrivals. Hewlett [65] models the arrival of market orders with Hawkes processes for single period market making. Finally, Alfonsi and Blanc [66], Jaisson [67] tackle the problem of optimal execution with market orders following multivariate Hawkes processes.

## 7.5.2 MODELING PRICE JUMPS

### 7.5.2.1 Single asset

Traditionally, the events of price jumps are modeled by Poisson processes, which suffer from the drawbacks mentioned in the Introduction section. Again, Hawkes processes can be applied to model price jumps, which often delineate clustering, self- and cross-excitation behaviors.

Bacry et al. [68] use Hawkes processes to model the price jumps, resulting in a model that can reproduce the microstructure noise [69], Epps effect [70], and jump clustering, while maintaining the coarse scale limit of Brownian diffusion. In their model, the trade price  $X(t)$  has the dynamics

$$X(t) = N_1(t) - N_2(t) \quad (7.29)$$

where  $N(t) = (N_1(t), N_2(t))$  is a bivariate linear Hawkes process with exponential decay kernel.  $N_1(t), N_2(t)$  represents the total number of upward and downward jumps, respectively. The authors make additional assumptions that the Hawkes process  $N$  has only cross-excitation and coefficients are symmetric to simplify computation.

$$\begin{aligned} \lambda_1(t) &= \mu + \int_{(0,t)} \gamma(t-s)dN_2(s), \quad \lambda_2(t) = \mu \\ &\quad + \int_{(0,t)} \gamma(t-s)dN_1(s), \quad \gamma(t) = \alpha e^{-\beta t} \end{aligned} \quad (7.30)$$

According to the model, when  $X$  jumps up (down),  $\lambda_2$  (resp.  $\lambda_1$ ) increases, causing the probability of jumping down (resp. up) to increase. Such a cross-linkage generates the effect of microstructure noise where the trade price is bouncing between best bid and best ask.

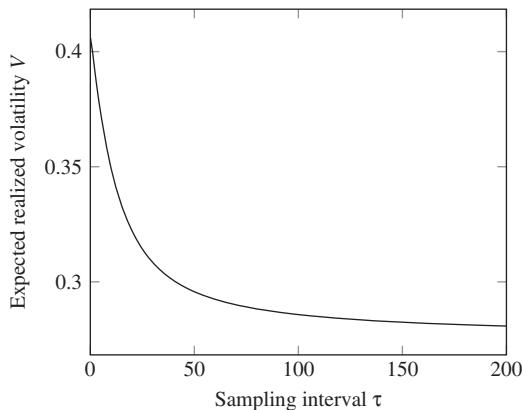
Because of the bid-ask bounce, it is well-known that the realized variance (annualized) increases when the sampling frequency increases [71].

$$V(\tau) = \mathbb{E} \left( \frac{1}{T} \sum_{n=0}^{T/\tau} (X((n+1)\tau) - X(n\tau))^2 \right) \quad (7.31)$$

$$= \frac{2\mu}{1-\alpha/\beta} \left( \frac{1}{(1+\alpha/\beta)^2} + \left( 1 - \frac{1}{(1+\alpha/\beta)^2} \right) \frac{1-e^{-(\alpha+\beta)\tau}}{(\alpha+\beta)\tau} \right) \quad (7.32)$$

Such an effect can be easily demonstrated by computing the expected realized variance (7.32) of the jump model (7.29) and the result with  $\mu = 0.16, \alpha = 0.024, \beta = 0.11$  as shown in Figure 7.1. The authors apply the model to Euro-Bund futures and find a very good fit between the observed and theoretical realized variances under this highly simplified model.

Let  $Y(t) = X(nt)$ , then  $Y(t)$  is a coarse scale version of  $X(t)$ . For example, if  $t$  in  $X$  is in microsecond and  $n = 60,000$ , then  $t$  in  $Y$  will be in minute. When we look at the trade price in a low-frequency setting, Bacry



**FIGURE 7.1** Volatility signature plot of hawkes jump model.

et al. [24] show that the macroscopic Hawkes jump model goes back to the classical model of Brownian motion due to the functional central limit theorem for linear Hawkes process (7.7). Assuming that  $\int_0^\infty \gamma(t)dt < 1$ , then

$$\frac{X(n\bullet)}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{weak}} \sigma W(\bullet), \quad \sigma^2 = \frac{2\mu}{(1 - \int_0^\infty \gamma(t)dt)(1 + \int_0^\infty \gamma(t)dt)^2} \quad (7.33)$$

It is interesting to see how the macroscopic variance  $\sigma^2$  is related to the microscopic base rate  $\mu$  and cross-excitation  $\gamma(t)$ . As  $\int_0^\infty \gamma(t)dt$  approaches 1, the variance goes to  $\infty$ .

Jaisson and Rosenbaum [25] extend the model of Bacry et al. [68] to the case of nearly unstable Hawkes process, where  $\int_0^\infty \gamma(t)dt \simeq 1$ , by constructing a sequence of kernel functions whose integrals converge to 1 at the speed of  $n^{-1}$ . They show that the properly scaled price process converges to Brownian diffusion with Heston stochastic volatility [72]. The full result is stated below.

$$X^{(n)}(t) = N_1^{(n)}(t) - N_2^{(n)}(t) \quad (7.34)$$

$$\lambda_1^{(n)}(t) = \mu + \int_0^t \gamma_1^{(n)}(t-s)dN_1^{(n)}(s) + \int_0^t \gamma_2^{(n)}(t-s)dN_2^{(n)}(s) \quad (7.35)$$

$$\lambda_2^{(n)}(t) = \mu + \int_0^t \gamma_2^{(n)}(t-s)dN_1^{(n)}(s) + \int_0^t \gamma_1^{(n)}(t-s)dN_2^{(n)}(s) \quad (7.36)$$

$$\left( \int_0^\infty \gamma_1^{(n)}(t)dt + \int_0^\infty \gamma_2^{(n)}(t)dt \right) < 1, \quad \gamma_i^{(n)}(t) = \alpha^{(n)}\gamma_i(t) \quad (7.37)$$

$$\gamma_i : \mathbb{R}_+ \longrightarrow \mathbb{R}_+, \\ \int_0^\infty (\gamma_1(t) + \gamma_2(t))dt = 1, \quad \int_0^\infty t(\gamma_1(t) + \gamma_2(t))dt = m < \infty \quad (7.38)$$

$$\int_0^\infty |\gamma'_i(t)|dt < \infty, \quad \sup_{t \in [0, \infty)} |\gamma'_i(t)| < \infty \quad (7.39)$$

$$\alpha^{(n)} \in [0, 1), \quad \lim_{n \rightarrow \infty} \alpha^{(n)} = 1, \quad \lim_{n \rightarrow \infty} n(1 - \alpha^{(n)}) = c > 0 \quad (7.40)$$

$$\psi^{(n)}(t) = \sum_{k=1}^{\infty} \left( \gamma_1^{(n)} + \gamma_2^{(n)} \right)^{\otimes k}(t), \quad \rho^{(n)}(t) = \frac{n\psi^{(n)}(nt)}{\int_0^\infty \psi^{(n)}(t)dt}, \\ |\rho^{(n)}(t)| \leq M \quad \forall n \quad \forall t \quad (7.41)$$

Under the conditions of (7.34–7.41),

$$\frac{X^{(n)}(n\bullet)}{n} \xrightarrow[n \rightarrow \infty]{\text{weak}} Y(\bullet) \quad (7.42)$$

$$dY_t = \frac{\sqrt{V_t}}{1 - \int_0^\infty |\gamma_1(t) - \gamma_2(t)|dt} dW_t^1, \quad Y_0 = 0 \quad (7.43)$$

$$dV_t = \frac{c}{m} \left( \frac{2\mu}{c} - V_t \right) dt + \frac{\sqrt{V_t}}{m} dW_t^2, \quad V_0 = 0 \quad (7.44)$$

Conditions (7.34–7.36) are just a bivariate Hawkes model (with both self- and cross-excitation) but now we have a different  $\gamma_i^{(n)}(t)$  for each  $n$  that is used to scale the time. The rest are the regularity conditions similar to the univariate nearly unstable Hawkes process (7.17) and the most important one is (7.40), which states that  $\alpha^{(n)}$  converges to 1 at the speed of  $n^{-1}$ . However, the interesting result is that instead of converging to an integrated CIR, the price dynamics formed by the difference between two Hawkes processes converges to a stochastic volatility model.

### 7.5.2.2 Two assets

To model the Epps effect, Bacry et al. [68] consider the two-asset case with prices  $(X_1(t), X_2(t))$  given by

$$X_1(t) = N_1(t) - N_2(t), \quad X_2(t) = N_3(t) - N_4(t) \quad (7.45)$$

$$\lambda_i(t) = \mu_i + \sum_{j=1}^4 \int_{(0,t)} \alpha_{ij} \exp(-\beta(t-s)) dN_j(s), \quad i = 1, \dots, 4 \quad (7.46)$$

$(N_1(t), \dots, N_4(t))$  is a 4-variate Hawkes process with exponential kernel, where  $\beta_{ij} = \beta$ . The coupling of excitation effects is constrained to have the form

$$\alpha = \begin{pmatrix} 0 & \alpha_{12} & \alpha_{13} & 0 \\ \alpha_{12} & 0 & 0 & \alpha_{13} \\ \alpha_{31} & 0 & 0 & \alpha_{34} \\ 0 & \alpha_{31} & \alpha_{34} & 0 \end{pmatrix} \quad (7.47)$$

In this case, there is a closed form representation for the realized correlation, which vanishes when the sampling interval goes to 0 (Epps effect).

If we assume  $\mu_1 = \mu_2, \mu_3 = \mu_4, \alpha_{12} = \alpha_{34} = 0, (\int_0^\infty \gamma_{13}(t)dt)$   $(\int_0^\infty \gamma_{31}(t)dt) < 1$ , then the macroscopic bivariate asset prices converges to correlated Brownian diffusion [24]

$$\frac{1}{\sqrt{n}} \begin{pmatrix} X_1(n\bullet) \\ X_2(n\bullet) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{\text{weak}} \frac{\sqrt{2} \left( \begin{array}{l} \sqrt{\nu_1} W_1(\bullet) + \sqrt{\nu_2} \int_0^\infty \alpha_{13}(t)dt W_2(\bullet) \\ \sqrt{\nu_1} \int_0^\infty \alpha_{31}(t)dt W_1(\bullet) + \sqrt{\nu_2} W_2(\bullet) \end{array} \right)}{(1 - (\int_0^\infty \gamma_{13}(t)dt)(\int_0^\infty \gamma_{31}(t)dt))^{3/2}} \quad (7.48)$$

$$\nu_1 = \mu_1 + \left( \int_0^\infty \gamma_{13}(t)dt \right) \mu_3, \nu_2 = \mu_3 + \left( \int_0^\infty \gamma_{31}(t)dt \right) \mu_1$$

$(W_1, W_2)$  is standard 2-dimensional Brownian motion    (7.49)

This convergence result gives us an explicit formula to estimate the macroscopic correlation from the asynchronous HF data.

As a final remark, under this jump representation, the observed trade price is not some hidden continuous fair value process plus some microstructure noise as in [73]. It is the result of the trading interactions between buyers ( $N_1, N_3$ ) and sellers ( $N_2, N_4$ ) on a fixed price grid. There is no such thing as HF volatility or correlation since prices are not diffusions but pure jump processes in the HF scale. Volatility and correlation are only meaningful when we look at the coarse scale diffusion approximation, but those low-frequency representation parameters can be computed directly from the HF jump model characteristics.

### 7.5.3 MODELING JUMP-DIFFUSION

Duffie et al. [74, 75] propose the affine jump-diffusion  $X(t)$ , which has the following structure.<sup>17</sup>

$$dX(t) = (k_0(t) + k_1(t)X(t))dt + (h_0(t) + h_1(t)X(t))dW(t) + \zeta dN(t) \quad (7.50)$$

$$\lambda(t) = a_0(t) + a_1(t)X(t) \quad (7.51)$$

The jump intensity  $\lambda(t)$  of  $N(t)$  is an affine function of  $X(t)$ , which depends on the Brownian motion  $W(t)$  and the jump process  $N(t)$ , with jump size  $\zeta$  drawn from a fixed distribution. When  $k_0 = \beta\theta$ ,  $k_1 = -\beta$ ,  $h_0 = h_1 = 0$ ,  $a_0 = 0$ ,  $a_1 = 1$ ,  $\zeta = \alpha$ , we can see that  $\lambda(t) = X(t)$  and  $d\lambda(t) = \beta(\theta - \lambda(t))dt + \alpha dN(t)$ . Hence in this case,  $N(t)$  is the Hawkes process with exponential kernel.

Zhu [76] derives some convergence results when  $\zeta$  is a constant and the diffusion part is a CIR process.

$$dX(t) = \beta(\mu - X(t))dt + \sigma\sqrt{X(t)}dW(t) + \alpha dN(t) \quad (7.52)$$

$$\lambda(t) = a_0 + a_1X(t) \quad (7.53)$$

Aït-Sahalia et al. [77] model the contagion of financial crisis with the Hawkes jump-diffusion where the price dynamic  $X_i(t)$  is given by

$$dX_i(t) = \mu_i dt + \sqrt{V_i(t)}dW_i^X(t) + Z_i(t)dN_i(t) \quad (7.54)$$

$$dV_i(t) = \kappa_i(\theta_i - V_i(t))dt + \eta_i\sqrt{V_i(t)}dW_i^V(t) \quad (7.55)$$

The diffusion part is the Heston stochastic volatility model and the jump part is a multivariate Hawkes process modeling the clustering and propagation of jumps among multiple assets.  $Z_i(t)$  corresponds to the jump size and direction.

### 7.5.4 MEASURING ENDOGENEITY (REFLEXIVITY)

In terms of the Hawkes branching structure representation of events arrivals, Filimonov et al. [78, 79] portray immigrants as exogenous news whereas the descendants are endogenous incidents. In the context of price movements in the stock market, immigrants are the price discovery due to

---

<sup>17</sup>We show only the one-dimensional case for simplicity.

orders from informed traders, who react to external information, whereas the descendants are the destabilizing ripples created by noise traders, who engage in herding [80], momentum trading [81], parasite trading [82], etc.

Under the univariate linear Hawkes model with exponential decay kernel and constant base rate, the expected number of direct descendants per individual (branching coefficient) is given by

$$n = \int_0^{\infty} \gamma(s)ds = \int_0^{\infty} \alpha e^{-\beta s} ds = \alpha/\beta \quad (7.56)$$

For a given immigrant, the expected number of descendants in all generations is  $n + n^2 + n^3 + \dots = n/(1 - n)$  if  $n < 1$ , so the ratio of descendants (nonimmigrants) versus total population is

$$\frac{\text{Descendants}}{\text{Descendants} + \text{immigrant}} = \frac{n/(1 - n)}{n/(1 - n) + 1} = n \quad (7.57)$$

Therefore, the branching coefficient  $n$  characterizes the amount of endogenous feedback activities while the base rate  $\mu$  measures the arrival rate of exogenous information.

Using E-mini S&P futures as proxy, Filimonov and Sornette [78] find that the level of endogeneity (reflexivity<sup>18</sup>)  $n$  in the US market has gone from 0.3 in 1998 to 0.7 in 2007. Moreover, in the flash crash of May 6, 2010,  $n$  reached a peak of 0.95.

Nonetheless, using the power law decay kernel, Hardiman et al. [84] challenge the result of Filimonov and Sornette [78] by reporting that the branching ratio  $n$  has always been close to 1 since 1998 and that the market could be a critical Hawkes process [23], but Filimonov and Sornette [85] refute that the power law kernel is sensitive to outliers in addition to other counter arguments. Later, Hardiman and Bouchaud [86] devised a nonparametric estimation of the branching ratio in terms of moments, but the result depends heavily on the window size used in the empirical moment computation.

---

<sup>18</sup>Filimonov and Sornette [78] borrow this term from Soros [83].

## Appendix 7.A: Point Processes

### 7.A.1 DEFINITION

Let  $X$  (state space) be a locally compact Hausdorff second countable topological space,<sup>19</sup>  $\mathcal{B}_X$  be the Borel sets on  $X$  and  $\mathcal{B}$  be the collection of bounded (relatively compact) Borel sets on  $X$ . A Borel measure  $\mu$  on  $(X, \mathcal{B}_X)$  is called locally finite if  $\mu(B) < \infty \forall B \in \mathcal{B}$ . Let  $\mathfrak{N}(X)^{20}$  be the set of (positive) locally finite Borel counting (integer-valued) measure on  $(X, \mathcal{B}_X)$  and  $\mathcal{N}(X)^{21}$  be the  $\sigma$ -algebra of  $\mathfrak{N}(X)$  generated by the set of evaluation functionals  $\{\Phi_B : \mathfrak{N}(X) \rightarrow \mathbb{N} \mid B \in \mathcal{B}\}$ , where  $\Phi_B(\mu) = \mu(B)$  and  $\mathbb{N} = \{0, 1, 2, \dots\}$ .

A point process  $N$  on  $X$  is defined as a measurable mapping from a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  to  $(\mathfrak{N}(X), \mathcal{N}(X))$ ; thus, a point process is formally a measure-valued random element. However, for any point process  $N$ , there exist random variables  $b_i \in \mathbb{Z}_+ = \{1, 2, \dots\}$ ,  $x_i \in X$ ,  $n \in \overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{\infty\}$  such that  $N(\bullet) = \sum_{i=1}^n b_i \delta_{x_i}(\bullet)$ , where  $\delta_x$  is the Dirac measure ( $\delta_x(A) = \mathbb{1}(x \in A)$ ) [8, p. 20]. If we think of  $b_i$  as the number of points at  $x_i$ , we can see that the point process  $N$  is indeed the random counting measure showing the total number of points in any given region and this matches our intuition that a point process is a random set of points  $\{x_i\}$  on  $X$ .

The point process  $N$  is called simple if  $\mathbb{P}(N(\{x\}) > 1) = 0 \forall x \in X^{22}$ ; that is, each location has at most one point. In this case,  $N(\bullet) = \sum_{i=1}^n \delta_{x_i}(\bullet)$ .

Suppose that  $X$  is also a topological vector space (e.g.,  $\mathbb{R}^m$ ), the shift operator  $S_t : \mathcal{B}_X \rightarrow \mathcal{B}_X$  is defined as  $S_t(A) = A + t = \{(s + t) \in X \mid s \in A\}$ . A point process  $N$  is called stationary if the shifted process  $N \circ S_t^{23}$  has the same distribution as  $N \forall t \in X$ .

<sup>19</sup>Some textbooks use complete separable metric space, but locally compact Hausdorff second countable space has a complete separable metrization and all the results here do not depend on any particular choice of metric [8, p. 11]. In most cases,  $X = \mathbb{R}^m$ .

<sup>20</sup>On locally compact Hausdorff second countable space, all locally finite Borel measures are Radon measures.

<sup>21</sup> $\mathcal{N}(X)$  is the same as the Borel  $\sigma$ -algebra generated by the vague topology of  $\mathfrak{N}(X)$  [8, p. 32].

<sup>22</sup> $X$  is Hausdorff, so all singletons are closed and thus measurable.

<sup>23</sup> $N \circ S_t : \Omega \rightarrow (\mathfrak{N}(X), \mathcal{N}(X))$ ,  $((N \circ S_t)(\omega))(A) = (N(\omega))(S_t(A))$ ; that is,  $N$  is shifted  $t$  unit to the left when  $X = \mathbb{R}$ .

### 7.A.2 MOMENTS

Let  $k \in \mathbb{Z}_+$ , the  $k^{\text{th}}$  moment measure<sup>24</sup>  $M^k : \mathcal{B}_X^{\otimes k} \rightarrow [0, \infty]$  of a point process  $N$  is defined as

$$M^k(A_1, \dots, A_k) = \mathbb{E}(N(A_1) \dots N(A_k)) = \mathbb{E}\left(\sum_{x_1} \dots \sum_{x_k} \delta_{(x_1, \dots, x_k)}(A_1 \times \dots \times A_k)\right) \quad (7.58)$$

The first moment measure is also called mean (intensity) measure and denoted as  $M(\bullet)$ . The covariance measure is defined as

$$C^2(A_1, A_2) = \text{Cov}(N(A_1), N(A_2)) = M^2(A_1, A_2) - M(A_1)M(A_2) \quad (7.59)$$

The second and higher moment measures have concentration along diagonals, so we also have the  $k^{\text{th}}$  factorial moment measure.

$$M^{(k)}(A_1, \dots, A_k) = \mathbb{E}\left(\sum_{x_1 \neq \dots \neq x_k} \dots \sum_{x_k} \delta_{(x_1, \dots, x_k)}(A_1 \times \dots \times A_k)\right) \quad (7.60)$$

The name factorial comes from the fact that  $M^{(k)}(A, \dots, A) = \mathbb{E}(N(A)(N(A) - 1) \dots (N(A) - k + 1))$ . Obviously,  $M(A) = M^{(1)}(A)$  and for  $k = 2$ , we have  $M^2(A_1, A_2) = M^{(2)}(A_1, A_2) + M(A_1 \cap A_2)$ .

If  $X = \mathbb{R}^m$  and  $N$  is stationary, it can be shown that  $M(A) = \lambda|A|$ , where  $\lambda = M((0, 1]^m)$  and  $|\bullet|$  is Lebesgue measure. That implies the mean measure  $M$  of a stationary point process is absolutely continuous with respect to Lebesgue measure with constant density  $M((0, 1]^m)$ . If the covariance factorial moment measure  $C^{(2)}$  is also absolutely continuous, we denote its density function as  $c^{(2)}(x, y)$ . Since  $N$  is stationary,  $c^{(2)}(x, y) = c^{(2)}(y - x)$  and  $c^{(2)}(\bullet)$  is called reduced covariance density. The covariance measure  $C^2$  is usually not absolutely continuous but for simple point process  $N$  on  $\mathbb{R}_+$ , the quantity below is still called (reduced) covariance density and is useful in estimation:

$$\begin{aligned} c^2(dx) &= \mathbb{E}(N(x + dx)N(x))/dx^2 - \lambda^2 = \lambda\delta(dx) + c^{(2)}(dx) \\ &\left( \int_{-\infty}^{\infty} \delta(x)dx = 1 \right) \end{aligned} \quad (7.61)$$

---

<sup>24</sup>The notations of moment, covariance, factorial moment, and reduced moment vary between authors.

### 7.A.3 MARKED POINT PROCESSES

When an event happens, it may carry an additional information (mark). For instance, each order arrival is associated with an order quantity (volume) and each earthquake is reported with a magnitude. A point process with marks is called marked point process.

Let  $Y$  (mark space) be a locally compact Hausdorff second countable space,  $(Y, \mathcal{B}_Y)$  be a measurable space and  $\nu$  (mark distribution) be a probability measure on  $(Y, \mathcal{B}_Y)$ . A marked point process (MPP)  $N$  is a measurable mapping  $N : \Omega \longrightarrow (\mathfrak{N}(X \times Y), \mathcal{N}(X \times Y))$  such that the ground measure  $N_g(\bullet) = N(\bullet \times Y)$  is a point process (i.e., locally finite).<sup>25</sup> Hence a marked point process is nothing but a point process on a product space, but usually we treat the location  $x$  and mark  $y$  differently and we have a few more definitions.

$N$  is called a multivariate point process if  $Y = \{1, \dots, d\}$ . In this case,  $N_i(\bullet) = N(\bullet \times \{i\})$  is called the marginal process of type  $i$  points. An MPP  $N$  is called simple if  $N_g$  is simple.<sup>26</sup> The marks of an MPP are called *unpredictable* if  $y_n$  is independent of  $\{(x_i, y_i)\}_{i < n}$  and they are called *independent* if  $y_n$  is independent of  $\{(x_i, y_i)\}_{i \neq n}$ .<sup>27</sup>

### 7.A.4 STOCHASTIC INTENSITY

In this section,  $X = \mathbb{R}_+$ <sup>28</sup> and  $N_t = N((0, t])$ . Let  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$  be a filtered complete probability space. A stochastic process  $Z : \mathbb{R}_+ \times \Omega \longrightarrow \mathbb{R}$  is called  $\mathcal{F}$ -predictable if it is measurable with respect to the predictable  $\sigma$ -algebra  $\mathcal{P} = \sigma(\{(s, t] \times A | 0 \leq s < t, A \in \mathcal{F}_s\})$ . If  $Z_t$  is adapted and left-continuous, then  $Z_t$  is predictable [10, p. 9]. In practice, all the predictable processes we use are in this category. Also, if  $Z_t$  is predictable, then  $Z_t \in \mathcal{F}_t$ ; in other words, the value of the predictable process  $Z_t$  at time  $t$  is “known” just before time  $t$ .

We assume that the filtration  $\{\mathcal{F}_t\}$  satisfies the usual condition (complete and right-continuous) and  $\{N_t\}$  is adapted and simple. A stochastic

<sup>25</sup>From this definition, Poisson random measure  $N$  on  $\mathbb{R}^2$  is not MPP on  $\mathbb{R} \times \mathbb{R}$  as  $N(A \times \mathbb{R}) = \infty$ .

<sup>26</sup>Any point process can be treated as simple MPP, with the mark being the number of points at  $x_i$ .

<sup>27</sup>Notice that if marks are independent, future location  $x_{n+1}$  cannot depend on previous mark  $y_n$ .

<sup>28</sup>The stochastic intensity of point process on  $\mathbb{R}_+$  is extended to higher dimension in [87].

process  $A : \mathbb{R}_+ \times \Omega \longrightarrow \mathbb{R}_+$  is called a  $\mathcal{F}$ -compensator of a point process  $N$  if  $A_t$  is increasing, right-continuous,  $\mathcal{F}$ -predictable,  $A_0 = 0$  a.s. and  $(N_t - A_t)$  is a  $\mathcal{F}$ -local martingale. If  $A_t = \int_0^t \lambda_s ds$  a.s.,  $\lambda_t$  is nonnegative and  $\mathcal{F}$ -predictable, then  $\lambda_t$  is called the stochastic or conditional  $\mathcal{F}$ -intensity of  $N$ .<sup>29</sup><sup>30</sup><sup>31</sup> A defining properties of  $\lambda_t$  is that

$$\mathbb{E} \left( \int_s^t \lambda_u du \mid \mathcal{F}_s \right) = \mathbb{E}(N_t - N_s \mid \mathcal{F}_s) \quad \text{a.s.} \quad \forall s < t \quad (7.62)$$

When  $s \rightarrow t$ , this becomes  $\lambda_t dt = \mathbb{E}(N(dt) \mid \mathcal{F}_{t^-})$ . We can see that the stochastic intensity  $\lambda_t$  is the instantaneous rate of arrival conditioned on all information just before time  $t$ . For a multivariate point process,  $\lambda_i(t)$  is the intensity of the marginal process  $N_i(t)$ .

From the definition, we notice that intensity exists if and only if the compensator is absolutely continuous. In fact, the compensator of a point process can be expressed in terms of the conditional interarrival time  $(t_n - t_{n-1}) \mid \mathcal{F}_{t_{n-1}}$  if the conditional distribution has support over  $\mathbb{R}_+$ . Under this condition, the intensity exists if and only if the conditional interarrival time is absolutely continuous. In this case, the intensity is given by [9, p. 70]

$$\lambda_t = h_n(t - t_{n-1}) \text{ if } t \in (t_{n-1}, t_n] \quad (7.63)$$

$$h_n(t) = \frac{g_n(t)}{1 - G_n(t^-)}, \quad (t_n - t_{n-1}) \mid \mathcal{F}_{t_{n-1}} \sim G_n \quad (7.64)$$

Once we know the intensity, we know the conditional distributions of all interarrival times and hence the complete distribution of the point process [11, p. 233].

$A : \mathbb{R}_+ \times \mathcal{B}_Y \times \Omega \longrightarrow \mathbb{R}_+$  is called a compensator of the MPP  $N$  if  $A(\bullet, B)$  is a compensator of  $N(\bullet \times B)$   $\forall B \in \mathcal{B}_Y$  and  $A(t, \bullet)$  is a measure on  $(Y, \mathcal{B}_Y)$   $\forall t \in \mathbb{R}_+$ . If  $A(t, B) = \int_0^t \int_B \lambda(s)v(s, dy)ds$  a.s. where  $\lambda(t)$  is nonnegative and predictable, then  $\lambda(t)$  is called the stochastic intensity of the MPP  $N$  and  $v(t_n, dy) = \mathbb{P}(y_n \in dy \mid \mathcal{F}_{t_n^-})$  is called the conditional mark distribution.

<sup>29</sup>Stochastic intensity is unique up to modification [10, p. 31].

<sup>30</sup>Notice that stochastic intensity depends on the underlying filtration, so some text use the notation  $\lambda(t|\mathcal{F}_t)$  but we will simply use  $\lambda(t)$  and call it stochastic intensity or intensity when there is no confusion about the filtration.

<sup>31</sup>if  $A_t$  is absolutely continuous with respect to Lebesgue measure and  $\lambda_t$  is the Radon-Nikodym derivative (may not be predictable) then  $\mathbb{E}(\lambda_t \mid \mathcal{F}_{t^-})$  is a version of the stochastic intensity. Some authors require only the intensity to be adapted, but using the conditional expectation, one can always find a predictable version of intensity provided that the intensity has finite first moment.

### 7.A.5 RANDOM TIME CHANGE

If the filtration is usual, a point process  $N$  on  $\mathbb{R}_+$  is simple and adapted, its intensity  $\lambda(t)$  exists, and  $\int_0^\infty \lambda(s)ds = \infty$  a.s., then  $\{\tilde{t}_n = \int_0^{t_n} \lambda(s)ds\}$  is a standard Poisson process (rate = 1). The above theorem is called random time change theorem [88, 89] and is extremely useful in testing the goodness of fit of a stochastic intensity model.

## Appendix 7.B: A Brief History of Hawkes processes

Hawkes processes are proposed by Hawkes [13] in 1971 to model contagious processes like epidemics, neuron firing, and particle emission, where the occurrences of events trigger further events. Although the intensity of Cox processes [90], introduced in 1955, are stochastic, they are determined before the events are unfolded.<sup>32</sup> In order to portray the excitation behavior in contagious processes, Hawkes extends the model in such a way that the intensity is a predictable stochastic process with an intuitive autoregressive form, which allows it to adapt to events that happen over time.

After Hawkes' seminal paper, there are a number of theoretical developments including the branching structure representation by Hawkes and Oakes [18] in 1974, Markov property for intensity with exponential decay kernel by Oakes [16] in 1975, MLE for Hawkes processes by Ozaki [29] in 1979, Ogata's modified thinning simulation algorithm [30] in 1981, nonlinear Hawkes processes by Brmaud and Massoulié [20] in 1996, non-parametric estimation by Gusto and Schbath [42] in 2005, EM for Hawkes processes by Veen and Schoenberg [33] in 2008, and functional central limit theorem for Hawkes processes by Bacry et al. [24], Jaisson and Rosenbaum [25], Zhu [26] in 2013–2015.

Although the first application of Hawkes processes to earthquake occurrences appeared in 1982 [15], it was not until 1988 [14] that Hawkes processes received much attention. Since then, the versatility of Hawkes model is leveraged in seismology [14, 91], finance (risk and credit default modeling) [92, 93], social networks [94, 95], neuroscience [96, 97], etc. (see [98, 99] for more applications).

The use of Hawkes processes in HF financial data modeling starts with Bowsher [22] in 2007<sup>33</sup> and then Large [62] in the same year. Both

<sup>32</sup>For processes on  $\mathbb{R}_+$ , it means that intensity of Cox process is  $\mathcal{F}_0$  measurable.

<sup>33</sup>Although Bowsher's paper was published in 2007, the first draft appeared in 2002.

authors exploit Hawkes processes so as to describe the interactions among different types of order arrivals. Later, Bacry et al. [68] and Aït-Sahalia et al. [77] employ Hawkes processes to reproduce jump clustering in pure jump and jump diffusion representation of stock prices in 2013 and 2015, respectively. On the other hand, an interesting idea from Filimonov and Sornette [78] appears in 2012, which utilizes the branching coefficient of linear Hawkes model to measure the level of endogenous activities in the US stock market, although the debate about the validity of the result is still ongoing [84–86].

## References

1. J.-P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. *Handbook of Financial Markets: Dynamics and Evolution*, 1:57, 2009.
2. R. Cont. Statistical modeling of high-frequency financial data. *Signal Processing Magazine, IEEE*, 28(5):16–25, 2011.
3. A. Chakraborti, I. M. Toke, M. Patriarca, and F. Abergel. Econophysics review: I. empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.
4. T. Andersen, T. Bollerslev, F. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001.
5. O. E. Barndorff-Nielsen. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002.
6. M. Garman. Market microstructure. *Journal of Financial Economics*, 3(3):257–275, 1976.
7. T. Ho and H. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73, 1981.
8. O. Kallenberg. *Random measures*. Akademie-Verlag, 1983.
9. A. Karr. *Point processes and their statistical inference*. CRC press, 1991.
10. P. Brémaud. *Point processes and queues*. Springer, 1981.
11. D. Daley and D. Vere-Jones. *An introduction to the theory of point processes: Volume I: Elementary theory and methods*. Springer, 2003.

12. D. Daley and D. Vere-Jones. *An introduction to the theory of point processes: Volume II: General theory and structure*. Springer, 2007.
13. A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
14. Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.
15. Y. Ogata and H. Akaike. On linear intensity models for mixed doubly stochastic poisson and self-exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 102–107, 1982.
16. D. Oakes. The Markovian self-exciting process. *Journal of Applied Probability*, 69–77, 1975.
17. J. Da Fonseca and R. Zaatour. Clustering and mean reversion in a Hawkes microstructure model. *Journal of Futures Markets*, 2014.
18. A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 493–503, 1974.
19. J. Møller and J. G. Rasmussen. Perfect simulation of Hawkes processes. *Advances in Applied Probability*, 629–646, 2005.
20. P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 1563–1588, 1996.
21. P. Brémaud, G. Nappo, and G. Torrisi. Rate of convergence to equilibrium of marked Hawkes processes. *Journal of Applied Probability*, 39(1):123–136, 2002.
22. C. G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.
23. P. Brémaud and L. Massoulié. Hawkes branching point processes without ancestors. *Journal of Applied Probability*, 38(1):122–135, 2001.
24. E. Bacry, S. Delattre, M. Hoffmann, and J. Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and Their Applications*, 2013.
25. T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, 25(2):600–631, 2015.

26. L. Zhu. Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability*, 50(3):760–771, 2013.
27. P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2009.
28. J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*. Springer-Verlag, Berlin, 1987.
29. T. Ozaki. Maximum likelihood estimation of Hawkes’ self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.
30. Y. Ogata. On Lewis’ simulation method for point processes. *Information Theory, IEEE Transactions on*, 27(1):23–31, 1981.
31. J. Zhuang, Y. Ogata, and D. Vere-Jones. Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*, 109(B5), 2004.
32. P. A. Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.
33. A. Veen and F. P. Schoenberg. Estimation of space–time branching process models in seismology using an EM–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624, 2008.
34. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38, 1977.
35. G. McLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, Inc., 2007.
36. D. Marsan and O. Lengline. Extending earthquakes’ reach through cascading. *Science*, 319(5866):1076–1079, 2008.
37. P. Halpin and P. Boeck. Modelling dyadic interaction with Hawkes processes. *Psychometrika*, 78(4):793–814, 2013.
38. P. F. Halpin. *A scalable EM algorithm for Hawkes processes*. New developments in quantitative psychology. Springer, 2013.
39. J. F. Olson and K. M. Carley. Exact and approximate EM estimation of mutually exciting Hawkes processes. *Statistical Inference for Stochastic Processes*, 16(1):63–80, 2013.
40. L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 1029–1054, 1982.

41. J. Da Fonseca and R. Zaatour. Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 2013.
42. G. Gusto and S. Schbath. Fado: A statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Statistical Applications in Genetics and Molecular Biology*, 4(1):1119, 2005.
43. C. De Boor. *A practical guide to splines*. Springer-Verlag, New York, 1978.
44. H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory*, pp. 267–281. Akademinai Kiado, 1973.
45. P. Reynaud-Bouret and S. Schbath. Adaptive estimation for Hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822, 2010.
46. N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 2015. ISSN 1350-7265.
47. R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288, 1996.
48. K. Zhou, H. Zha, and L. Song. Learning triggering kernels for multi-dimensional Hawkes processes. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 1301–1309, 2013.
49. D. R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37, 2004.
50. E. Bacry, K. Dayri, and J. F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B*, 85(5), 2012.
51. E. Bacry and J.-F. Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.
52. E. Bacry and J.-F. Muzy. Second order statistics characterization of Hawkes processes and non-parametric estimation. *arXiv preprint arXiv:1401.0903*, 2014.
53. B. Noble. *Methods based on the Wiener-Hopf technique*. Pergamon Press, New York, 1958.

54. G. Chandlen and I. G. Graham. The convergence of Nyström methods for Wiener–Hopf equations. *Numerische Mathematik*, 52(3):345–364, 1987.
55. A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):617–666, 2005.
56. M. Wilk and R. Gnanadesikan. Probability plotting methods for the analysis of data. *Biometrika*, 55(1):1–17, 1968.
57. A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4(1):83–91, 1933.
58. N. V. Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin of Mathematics University of Moscow*, 2(2), 1939.
59. N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 279–281, 1948.
60. G. M. Ljung and G. E. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.
61. F. P. Schoenberg. Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98(464), 2003.
62. J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.
63. I. Muni Toke and F. Pomponio. Modelling trades-through in a limited order book using Hawkes processes. *Economics: The Open-Access, Open-Assessment E-Journal*, 6(2012-22), 2012.
64. I. Muni Toke. “Market Making” in an order book model and its impact on the spread. In *Econophysics of order-driven markets*, pages 49–64. Springer, 2011.
65. P. Hewlett. Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*, pages 6–8, 2006.
66. A. Alfonsi and P. Blanc. Dynamic optimal execution in a mixed-market-impact Hawkes price model. *arXiv preprint arXiv:1404.0648*, 2014.

67. T. Jaisson. Market impact as anticipation of the order flow imbalance. *Quantitative Finance*, 15(7):1123–1135, 2015.
68. E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.
69. Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18(2):351–416, 2005.
70. T. W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a):291–298, 1979.
71. T. Andersen, T. Bollerslev, F. Diebold, and P. Labys. Great realizations. *Risk Magazine*, 2000.
72. S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.
73. F. M. Bandi and J. R. Russell. Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3):655–692, 2006.
74. D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406, 1996.
75. D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.
76. L. Zhu. Limit theorems for a Cox–Ingersoll–Ross process with Hawkes jumps. *Journal of Applied Probability*, 51(3):699–712, 2014.
77. Y. Aït-Sahalia, J. Cacho-Díaz, and R. J. A. Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015.
78. V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.
79. V. Filimonov, D. Bicchetti, N. Maystre, and D. Sornette. Quantification of the high level of endogeneity and of structural regime shifts in commodity markets. *Journal of International Money and Finance*, 2013.
80. R. J. Shiller. *Irrational exuberance*. Princeton University Press, 2005.

81. J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance*, 379–395, 1990.
82. L. Harris. Order exposure and parasitic traders. *Unpublished manuscript*, 1997.
83. G. Soros. *The alchemy of finance*. John Wiley Sons, 2003.
84. S. J. Hardiman, N. Bercot, and J.-P. Bouchaud. Critical reflexivity in financial markets: A Hawkes process analysis. *The European Physical Journal B*, 86(10):1–9, 2013.
85. V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance*, 15(8):1293–1314, 2015.
86. S. J. Hardiman and J.-P. Bouchaud. Branching-ratio approximation for the self-exciting Hawkes process. *Physical Review E*, 90(6):062807, 2014.
87. F. Papangelou. The conditional intensity of general point processes and an application to line processes. *Probability Theory and Related Fields*, 28(3):207–226, 1974.
88. P. A. Meyer. Démonstration simplifiée d'un théorème de knight. In *Séminaire de probabilités v université de strasbourg*, pages 191–195. Springer, 1971.
89. F. Papangelou. Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society*, 165:483–506, 1972.
90. D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):129–164, 1955.
91. Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.
92. V. Chavez-Demoulin, A. Davison, and A. Mcneil. Estimating value-at-risk: A point process approach. *Quantitative Finance*, 5(2):227–234, 2005.
93. E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

94. R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.
95. C. Blundell, J. Beck, and K. A. Heller. Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems*, pages 2609–2617, 2012.
96. M. Krumin, I. Reutsky, and S. Shoham. Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Frontiers in Computational Neuroscience*, 4, 2010.
97. V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. How structure determines correlations in neuronal networks. *PLoS Computational Biology*, 7(5):e1002059, 2011.
98. T. Liniger. *Multivariate Hawkes processes*. PhD thesis, Swiss Federal Institute of Technology, 2009.
99. L. Zhu. *Nonlinear Hawkes processes*. PhD thesis, New York University, 2013.

## Chapter Eight

# Multifractal Random Walk Driven by a Hermite Process

Alexis Fauth<sup>1,2</sup> and Ciprian A. Tudor<sup>3,4,\*</sup>

<sup>1</sup>SAMM, Université de Paris I Panthéon-Sorbonne, 90, rue de Tolbiac, 75634, Paris, France

<sup>2</sup>Invivoo, 13 rue de l'Abreuvoir, 92400 Courbevoie, France

<sup>3</sup>Laboratoire Paul Painlevé, Université de Lille 1, F-59655 Villeneuve d'Ascq, France

<sup>4</sup>Department of Mathematics, Academy of Economical Studies, Bucharest, Romania

### 8.1 Introduction

The so-called random multifractal processes (or simply multifractals) have been used to model natural and man-made phenomena in a variety of fields, such as hydrodynamics, genetics, Internet traffic analysis, or stock prices. Our work concerns the applications of multifractals to this last field, mathematical finance. Starting with the seminal work of Mandelbrot about cotton price [22], several studies of financial stock prices times series have allowed to exhibit some particularities of their fluctuations. Without making a comprehensive list, we can mention the appearance in

---

\*Supported by the CNCS grant PN-II-ID-PCCE-2011-2-0015. Associate member of the team Samm, Université de Panthéon-Sorbonne Paris 1.

empirical data of the following properties: non-Gaussian distributions due to now well-known fat tails of financial returns, the so-called volatility clustering that means that the volatility fluctuations are of intermittent and of correlated nature, scaling invariance, long-run correlation in volatility, leverage effect, and so on (see, e.g., [12, 18, 34] for an extensive review). Thus, constructing theoretical models for financial returns that include all the properties listed before appears as a very interesting challenge.

The multifractal models are introduced to capture the aforementioned stylized facts but multifractality started to emerge as an additional stylized fact since the nineties. The multifractality can be understood as an invariance scaling property. At any scales, we observe the same properties. The concept of multifractality represents an extension of the notion of self-similarity. A process  $(X_t)_{t \geq 0}$  is multifractal if it satisfies the following scaling property:

$$\mathbb{E} |X_{t+l} - X_t|^q = K_q l^{\xi(q)} \quad (8.1)$$

at any scale  $l$  and for any  $q \geq 0$ . The exponent  $\xi(q)$  is usually called the multifractal spectrum of the process  $X$ .

When the exponent  $\xi(q)$  is linear in  $q$ , that is,  $\xi(q) = qH$ , the process is referred to as a monofractal process. The self-similar processes (e.g., fractional Brownian motions) are a particular case of monofractal processes. When  $\xi(q)$  is a nonlinear function of  $q$ , it is referred to as a multifractal process, or sometimes, as a process displaying multiscaling or intermittency.

In high-frequency data, one observes a nontrivial multifractal scaling in the sense that the higher moments of the return process satisfy the property (8.1) where the index  $\xi(q)$  is not a constant and it is not a linear function of  $q$ . The multifractal random walk (MRW) constructed in our chapter has this multiscaling character.

The MRW (see, e.g., [8], and see [1]) is defined as

$$Z(t) = \int_0^t Q(u) dY(u) \quad (8.2)$$

where  $Q$  is a suitable fractal noise and  $Y$  is a self-similar process with stationary increments, independent of  $Q$ . In his initial form [6],  $Y$  is chosen as a standard Brownian motion. More recent approaches ([1, 15]) used as an integrator in (8.2) the fractional Brownian motion (fBm in the sequel) and this choice is more appropriate to obtain long-memory properties for the financial log-returns.

In addition to multifractality, this model is able to capture several properties observed on the empirical data, clustering, autocorrelation in log-returns, and squared log-returns. This process is also able to reproduce the leverage effect when the fractal noise  $Q$  and the integrator  $Y$  in (8.2) are dependent processes, which is not the considered case here. Such an example can be found in [15].

The purpose of this work is to propose a generalization of the model (8.2) that allows more flexibility for the log-return. Actually the integrator  $Y$  in (8.2), who is usually chosen to be a standard Brownian motion, will be replaced by a larger class of long-memory processes, called the Hermite processes that includes the fractional Brownian motion and Rosenblatt process.

The Hermite process of order  $k \geq 1$  is a self-similar process with stationary increments and exhibits long memory. When  $k = 1$ , it reduces to the fractional Brownian motion but is not a Gaussian process for  $k \geq 2$ . The Hermite process of order 2 is called the Rosenblatt process.

We will show that this new model, with the fBm replaced by a Hermite process (this will be called a Hermite multifractal random walk and abbreviated HMRW) of order  $k$  in (8.2) keeps the multifractal character. Indeed, we prove in Section 8.3 that equation (8.1) is satisfied by the HMRW. This is because the integrator in (8.2) is self-similar and it has stationary increments. Moreover, the HMRW is also able to capture other empirical stylized facts, such as the decay of the autocorrelation of the return process. This property is induced by the long-range dependence of the Hermite process. All these facts, together with the non-Gaussian nature of the object (8.2), make the HMRW a natural candidate to model high-frequency financial data.

One of the main challenges that appear when working with such a process as an integrator in (8.2) is how to construct the stochastic integral with respect to it. Since the Hermite process is not a semimartingale, and it is not Gaussian, no classical stochastic integral can be defined with respect to it. Therefore, we will restrict to the case when  $Y$  (which will be the Hermite process) and the fractal noise  $Q$  in (8.2) are independent. In this case, the stochastic integral in (8.2) can be viewed as a Wiener integral, which is easier to define and handle. The model still keeps very well many properties of the empirical data, as the initial MRW, including long-memory in log-returns.

We have organized our chapter as follows. In Section 8.2, we recall the definition and the main properties of the Hermite process and we explain

how the Wiener integrals with respect to it are constructed. In Section 8.3, we define and analyze the HMRW, while in Section 8.4, we describe how our object can be simulated and how they fit to real-world data.

## 8.2 Preliminaries

In this paragraph, we first describe the integrator in the integral in (8.2) and then we explain how the Wiener integral with respect to a Hermite process is constructed. More details can be found in [26]. In the last part of this section, we present the definition and the basic properties of the fractal process  $Q$  that appears in the expression (8.2).

### 8.2.1 FRACTIONAL BROWNIAN MOTION AND HERMITE PROCESSES

Let us first introduce the integrator in the integral in (8.2). The fractional Brownian motion  $(B_t^H)_{t \geq 0}$  with Hurst parameter  $H \in (0, 1)$  is a centered Gaussian process starting from 0 with covariance function

$$R^H(t, s) := \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}), \quad s, t \geq 0.$$

$B_t^H$  is a  $H$ -self-similar process with stationary increments. Actually, it is the only Gaussian process  $H$ -self-similar with stationary increments.

The fractional Brownian process  $(B_t^H)_{t \in [0, 1]}$  with Hurst parameter  $H \in (0, 1)$  can be written as a Wiener integral with respect to the Wiener process as

$$B_t^H = \int_0^t K^H(t, s)dW_s, \quad t \geq 0$$

where  $(W_t, t \geq 0)$  is a standard Wiener process, the kernel  $K^H(t, s)$  has the expression (when  $H > \frac{1}{2}$ )  $c_H s^{1/2-H} \int_s^t (u - s)^{H-3/2} u^{H-1/2} du$  for  $t > s$  (and it vanishes if  $s \geq t$ ),  $c_H$  is an explicit positive constant. The above integral is a Wiener integral with respect to the standard Wiener process  $W$ . For  $t > s$ , the kernel's derivative is  $\frac{\partial K^H}{\partial t}(t, s) = c_H (\frac{s}{t})^{1/2-H} (t - s)^{H-3/2}$ . Fortunately, we will not need to use these expressions explicitly, since they will be involved below only in integrals whose expressions are known.

We will denote by  $(Z_H^k(t))_{t \geq 0}$  the  $q$ th Hermite process with self-similarity parameter  $H \in (1/2, 1)$ . Here  $q \geq 1$  is an integer. The Hermite

process can be defined as a multiple integral with respect to the standard Wiener process  $W$ ; details concerning the definition of multiple stochastic integrals can be found in Chapter 1 in [25]. These multiple integrals can be viewed as iterated Itô integrals. Concretely, we have the following definition.

**Definition 8.1.** Let  $(B(t))_{t \in \mathbb{R}}$  be a Wiener process. The Hermite process  $(Z_H^k(t))_{t \geq 0}$  of order  $k$  and with self-similarity index  $H \in (\frac{1}{2}, 1)$  is defined as

$$Z_H^k(t) = c(H, k) \int_{\mathbb{R}^k} \int_0^t \left( \prod_{i=1}^k (s - y_i)_+^{-\left(\frac{1}{2} + \frac{1-H}{k}\right)} \right) ds dB(y_1) \dots dB(y_k), \quad (8.3)$$

where  $x_+ = \max(x, 0)$ . The above integral is a multiple integral of order  $k$  with respect to the Brownian motion  $B$  and the constant  $c(H, k)$  is a normalizing constant that ensures that  $\mathbb{E}(Z_H^k(t))^2 = 1$ .

**Remark 8.1.** Throughout, a random variable that has the same law as  $Z_H^k(1)$  will be called as Hermite random variable.

The most studied Hermite process is, of course, *the fractional Brownian motion* (which is obtained in (8.3) for  $k = 1$ ) due to its large range of applications. The process obtained in (8.3) for  $k = 2$  is known as *the Rosenblatt process*. It was introduced by Rosenblatt in [29] and it has been named in this way by M. Taqqu in [30].

First, we mention that  $Z_H^k$  is a centered process since it is defined by a multiple stochastic integral. Of fundamental importance is the fact that the covariance of  $Z_H^k$  is identical to that of fBm, namely, for every  $s, t \geq 0$

$$R(t, s) = \mathbb{E}[Z_H^k(t)Z_H^k(s)] = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}). \quad (8.4)$$

The basic properties of the Hermite process are listed below:

- the Hermite process  $Z_H^k$  is  $H$ -self-similar and it has stationary increments.
- the mean square of the increment is given by, for  $s, t \geq 0$

$$\mathbb{E} \left[ \left| Z_H^k(t) - Z_H^k(s) \right|^2 \right] = |t - s|^{2H}; \quad (8.5)$$

as a consequence, it follows will little extra effort from Kolmogorov's continuity criterion that  $Z_H^k$  has Hölder-continuous paths of any exponent  $\delta < H$ . From the self-similarity and the stationary of increments of the Hermite process, we also have, for every  $p > 2$ , that

$$\mathbb{E} \left[ \left| Z_H^k(t) - Z_H^k(s) \right|^p \right] = |t - s|^{pH}; \quad (8.6)$$

- it exhibits long-range dependence in the sense that  $\sum_{n \geq 1} \mathbb{E} [Z_H^k(1)(Z_H^k(n+1) - Z_H^k(n))] = \infty$ . In fact, the summand in this series is of order  $n^{2H-2}$ . This property is identical to that of fBm since the processes share the same covariance structure, and the property is well-known for fBm with  $H > 1/2$ .

### 8.2.2 WIENER INTEGRALS WITH RESPECT TO THE HERMITE PROCESS

In this section, we recall the definition of the Wiener integrals with respect to the Hermite process. This construction has been done in [26]. Consider a Hermite process given by (8.3).

Let us denote by  $\mathcal{E}$  the class of elementary functions on  $\mathbb{R}$  of the form

$$f(u) = \sum_{l=1}^n a_l 1_{(t_l, t_{l+1}]}(u), \quad t_l < t_{l+1}, \quad a_l \in \mathbb{R}, \quad l = 1, \dots, n. \quad (8.7)$$

For  $f \in \mathcal{E}$  as above, it is natural to define its Wiener integral with respect to the Hermite process  $Z_H^k$  by

$$\int_{\mathbb{R}} f(u) dZ_H^k(u) = \sum_{l=1}^n a_l (Z_H^k(t_{l+1}) - Z_H^k(t_l)). \quad (8.8)$$

In order to extend the definition (8.8) to a larger class of integrands, let us make first some observations. By formula (8.3), we can write

$$Z_H^k(t) = \int_{\mathbb{R}^k} I(1_{[0,t]})(y_1, \dots, y_k) dB(y_1) \dots dB(y_k),$$

where by  $I$  we denote the mapping on the set of functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  to the set of functions  $f : \mathbb{R}^k \rightarrow \mathbb{R}$

$$I(f)(y_1, \dots, y_k) = c(H, k) \int_{\mathbb{R}} f(u) \prod_{i=1}^k (u - y_i)_+^{-\left(\frac{1}{2} + \frac{1-H}{k}\right)} du.$$

Note that for  $k = 1$  this operator can be expressed in terms of fractional integrals and derivatives, (see [28]). Thus, the definition (8.8) can also be written in the form, because of the obvious linearity of  $I$

$$\begin{aligned} \int_{\mathbb{R}} f(u) dZ_H^k(u) &= \sum_{l=1}^n a_l (Z_H^k(t_{l+1}) - Z_H^k(t_l)) \\ &= \sum_{l=1}^n a_l \int_{\mathbb{R}^k} I(1_{(t_l, t_{l+1})})(y_1, \dots, y_k) dB(y_1) \dots dB(y_k) \\ &= \int_{\mathbb{R}^k} I(f)(y_1, \dots, y_k) dB(y_1) \dots dB(y_k). \end{aligned} \quad (8.9)$$

We introduce now the following space:

$$\mathcal{H} = \left\{ f : \mathbb{R} \rightarrow \mathbb{R} \mid \int_{\mathbb{R}^k} (I(f)(y_1, \dots, y_k))^2 dy_1 \dots dy_k < \infty \right\}$$

endowed with the norm

$$\|f\|_{\mathcal{H}}^2 = \int_{\mathbb{R}^k} (I(f)(y_1, \dots, y_k))^2 dy_1 \dots dy_k.$$

It holds that

$$\begin{aligned} \|f\|_{\mathcal{H}}^2 &= c(H, k)^2 \int_{\mathbb{R}^k} \left( \int_{\mathbb{R}} \int_{\mathbb{R}} f(u) f(v) \prod_{i=1}^k (u - y_i)_+^{-\left(\frac{1}{2} + \frac{1-H}{k}\right)} \right. \\ &\quad \times (v - y_i)_+^{-\left(\frac{1}{2} + \frac{1-H}{k}\right)} dv du \Bigg) dy_1, \dots, dy_k \\ &= c(H, k)^2 \int_{\mathbb{R}} \int_{\mathbb{R}} f(u) f(v) \left( \int_{\mathbb{R}} (u - y)_+^{-\left(\frac{1}{2} + \frac{1-H}{k}\right)} (v - y)_+^{-\left(\frac{1}{2} + \frac{1-H}{k}\right)} dy \right)^k \\ &\quad \times dv du \\ &= H(2H - 1) \int_{\mathbb{R}} \int_{\mathbb{R}} f(u) f(v) |u - v|^{2H-2} dv du. \end{aligned}$$

Hence we have

$$\mathcal{H} = \left\{ f : \mathbb{R} \rightarrow \mathbb{R} \mid \int_{\mathbb{R}} \int_{\mathbb{R}} f(u) f(v) |u - v|^{2H-2} dv du < \infty \right\}$$

and

$$\|f\|_{\mathcal{H}}^2 = H(2H - 1) \int_{\mathbb{R}} \int_{\mathbb{R}} f(u) f(v) |u - v|^{2H-2} dv du.$$

Let us observe that the mapping

$$f \mapsto \int_{\mathbb{R}} f(u) dZ_H^k(u) \quad (8.10)$$

provides an isometry from  $\mathcal{E}$  to  $L^2(\Omega)$ . Indeed, for  $f$  of the form (8.7), it holds that

$$\begin{aligned} \mathbb{E}[I(f)^2] &= \sum_{i,j=0}^{n-1} a_i a_j \mathbb{E}[Z_H^k(t_{i+1}) - Z_H^k(t_i)] (Z_H^k(t_{j+1}) - Z_H^k(t_j)) \\ &= \sum_{i,j=0}^{n-1} a_i a_j (R(t_{i+1}, t_{j+1}) - R(t_{i+1}, t_j) - R(t_i, t_{j+1}) + R(t_i, t_j)) \\ &= \sum_{i,j=0}^{n-1} a_i a_j H(2H-1) \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} |u-v|^{2H-2} dv du \\ &= \sum_{i,j=0}^{n-1} a_i a_j \langle 1_{(t_i, t_{i+1})}, 1_{(t_j, t_{j+1})} \rangle_{\mathcal{H}} = \|f\|_{\mathcal{H}}^2, \end{aligned}$$

where  $R$  represents the covariance function (8.4). On the other hand, it has been proved in [28] that the set of elementary functions  $\mathcal{E}$  is dense in  $\mathcal{H}$ . As a consequence, the mapping (8.10) can be extended to an isometry from  $\mathcal{H}$  to  $L^2(\Omega)$  and relation (8.9) still holds. This extension will be called the Wiener integral with respect to the Hermite process.

The space  $\mathcal{H}$  coincides with the canonical Hilbert space associated with the fBm. Therefore, the followings facts hold (see [28] or [25]):

- The elements of  $\mathcal{H}$  may be not functions but distributions; it is therefore more practical to work with subspaces of  $\mathcal{H}$  that are sets of functions. Such a subspace is

$$|\mathcal{H}| = \left\{ f : \mathbb{R} \rightarrow \mathbb{R} \mid \int_{\mathbb{R}} \int_{\mathbb{R}} |f(u)| |f(v)| |u-v|^{2H-2} dv du < \infty \right\}.$$

Then  $|\mathcal{H}|$  is a strict subspace of  $\mathcal{H}$  and we actually have the inclusions

$$L^2(\mathbb{R}) \cap L^1(\mathbb{R}) \subset L^{\frac{1}{H}}(\mathbb{R}) \subset |\mathcal{H}| \subset \mathcal{H}. \quad (8.11)$$

- The space  $|\mathcal{H}|$  is not complete with respect to the norm  $\|\cdot\|_{\mathcal{H}}$ , but it is a Banach space with respect to the norm

$$\|f\|_{|\mathcal{H}|}^2 = \int_{\mathbb{R}} \int_{\mathbb{R}} |f(u)| |f(v)| |u-v|^{2H-2} dv du.$$

### 8.2.3 INFINITELY DIVISIBLE CASCADING NOISE

Let  $M$  denote an infinitely divisible, independently scattered random measure on the set  $\mathbb{R} \times \mathbb{R}_+$  with generating infinitely divisible distribution  $G$  satisfying

$$\int_{\mathbb{R}} e^{qx} G(dx) = e^{-\rho(q)}$$

for some function  $\rho$  and for every  $q \in \mathbb{R}$ . We assume that  $M$  has control measure  $m$  on  $\mathbb{R} \times \mathbb{R}_+$  meaning that for every Borel set  $A \subset \mathbb{R} \times \mathbb{R}_+$  it holds

$$\mathbb{E} e^{qM(A)} = e^{-\rho(q)m(A)} \text{ for every } q \in \mathbb{R}. \quad (8.12)$$

The fact that  $M$  is independently scattered means that the random variables

$$M(A_1), M(A_2), \dots, M(A_n)$$

are independent whenever the Borel sets  $A_1, \dots, A_n \in \mathbb{R} \times \mathbb{R}_+$  are disjoint. We define the *Infinitely Divisible Cascading noise* (IRC) by

$$Q_r(t) = \frac{e^{M(C_r(t))}}{\mathbb{E} e^{M(C_r(t))}} \quad (8.13)$$

for every  $r > 0$  and  $t \in \mathbb{R}$ . Here  $C_r(t)$  is the cone in  $\mathbb{R} \times \mathbb{R}_+$  defined by

$$C_r(t) = \left\{ (t', r'), r \leq r' \leq 1, t - \frac{r'}{2} \leq t' \leq t + \frac{r'}{2} \right\}. \quad (8.14)$$

We will use the following facts throughout our chapter. We refer to [11] or [1] for their proofs. First, let us note the scaling property of the moments of the IRC

$$\mathbb{E} Q_r(t)^q = e^{-\varphi(q)m_r(0)}$$

and the expression of its covariance: for every  $r > 0$  and  $t, s \in \mathbb{R}$

$$\mathbb{E} Q_r(t) Q_r(s) = e^{-\varphi(2)m_r(|t-s|)}, \quad (8.15)$$

where we denoted by for  $u \geq 0, r > 0$ .

$$m_r(u) = m(C_r(0) \cap C_r(u))$$

and by

$$\varphi(q) = \rho(q) - q\rho(1). \quad (8.16)$$

In this work, we consider

$$m(dt, dr) = dt \frac{cdr}{r^2} \text{ if } 0 < r \leq 1 \quad (8.17)$$

and  $m$  vanishes if  $r \geq 1$ . Here  $c$  is a strictly positive constant. The case when the measure  $m$  is given by (8.17) is called in [4] (see also [1]) the exact invariant scaling case. This implies

$$\begin{aligned} m_r(u) &= 0 \text{ if } u > 1, m_r(u) = -c \ln u + c(u-1) \text{ if } r < u \leq 1 \\ \text{and } m_r(u) &= -c \ln r + cu \left(1 - \frac{1}{r}\right) \text{ if } 0 \leq u \leq r. \end{aligned}$$

The scaling of the moment of  $Q$  can be extended to the following scaling property in distribution: for  $t \in (0, 1)$

$$(Q_{rt}(tu))_{u \in \mathbb{R}} =^{(d)} e^{\Omega_t} (Q_r(u))_{u \in \mathbb{R}} \quad (8.18)$$

where “ $=^{(d)}$ ” means equivalence of finite dimensional distributions. Here  $\Omega_t$  denotes a random variable independent by  $Q$ , which satisfies, if the measure  $m$  is given by (8.17),

$$\mathbb{E} e^{q\Omega_t} = t^{q\varphi(q)}. \quad (8.19)$$

**Remark 8.2.** As noticed in [1], we have  $\varphi(2) < 0$ .

In [1] (see also [20]) the MRW based on fractional Brownian motion is defined as limit when  $r \rightarrow 0$  (in some sense) of the family of stochastic integrals  $(Z_r^H(t))_{t \geq 0}$  defined by

$$Z_r^H(t) = \int_0^t Q_r(u) dB^H(u), \quad t \in [0, T], \quad (8.20)$$

where  $(B_t^H)_{t \in [0, T]}$  is a fractional Brownian motion with Hurst parameter  $H \in (0, 1)$ . The fractional Brownian motion  $(B_t^H)_{t \in [0, T]}$  with Hurst parameter  $H \in (0, 1)$  is a centered Gaussian process starting from 0 with covariance function

$$R^H(t, s) := \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}), \quad s, t \in [0, T]. \quad (8.21)$$

In [1], it is assumed that  $M$  and  $B^H$  are independent. We will follow this assumption in our work. We mention that a situation when  $M$  and  $B^H$  are dependent has been treated in [15]. Therefore, the stochastic integral with respect to  $B^H$  in (8.20) behaves mainly as a Wiener integral since, because of the independence, the integrand  $Q_r(u)$  can be viewed as deterministic function for the integrator  $B^H$ .

Another important fact in the development of this theory is that the IRC  $Q$  is a martingale with respect to the argument  $r$ . Let us recall the following result (see [11]):

**Lemma 8.1.** *For every  $u > 0$  the stochastic process  $(Q_r(u))_{r>0}$  is a martingale with respect to its own filtration. As a consequence, for every  $u, v, r, r' > 0$  with  $r < r'$  it holds*

$$\mathbb{E}Q_r(u)Q_{r'}(v) = \mathbb{E}Q_r(u)Q_r(v). \quad (8.22)$$

The property (8.22) plays an important role in the construction of the MRW process in [1] or [11].

## 8.3 Multifractal random walk driven by a Hermite process

### 8.3.1 DEFINITION AND EXISTENCE

Let us fix  $H \in (\frac{1}{2}, 1)$ , and define for every  $r > 0$

$$X_r^H(t) = \int_0^t Q_r(u)dZ_H^k(u) \quad (8.23)$$

where  $Z_H^k$  denotes a Hermite process of order  $k$  and with self-similarity order  $H$  (see Section 8.2). This is the approximating Hermite Multifractal Random Walk in the sense that its limit, when  $r \rightarrow 0$ , will be defined as the Hermite MRW.

For the sake of simplicity, let us denote  $Z_k^H$  by  $Z$  in the sequel. We will assume that the multifractal measure  $M$  and the Hermite process  $Z$  are independent and therefore the integral in (8.23) can be understood as a Wiener integral with respect to the Hermite process as described in Section 2.2.

We can easily prove that  $X_r^H(t)$  is well-defined for every  $r, t$ . Indeed, the stochastic integral in (8.23) can be viewed as a Wiener integral with respect to  $Z$  and it follows easily from the definition of the Wiener integral with respect to the Hermite process that  $X_r^H(t)$  has the same square mean independently on  $k \geq 1$  and we can directly use the results obtained in [1] in the case  $k = 1$ . In fact, with  $\alpha_H = H(2H - 1)$ ,

$$\begin{aligned}\mathbb{E} |X_r^H(t)|^2 &= \alpha_H \int_0^t \int_0^t dudv \mathbb{E}(Q_r(u)Q_r(v))|u-v|^{2H-2} \\ &= \alpha_H \int_0^t \int_0^t dudv |u-v|^{2H-2} e^{-\rho(2)m_r(|u-v|)} \\ &\leq C_{r,t}\end{aligned}$$

where we used the fact that  $\mathbb{E}Q_r(u)Q_r(v) = e^{-\rho(2)m_r(|u-v|)} \leq c_r$  with  $c_r$  a strictly positive constant.

We can prove the following result.

**Theorem 8.1.** *Let  $\phi$  be given by (8.16) and assume*

$$c\varphi(p) + Hp > 1. \quad (8.24)$$

*For every  $t \geq 0$  and for every  $p \geq 2$ , the sequence  $(X_r^H(t))_{r>0}$  converges in  $L^p(\Omega)$  to a random variable  $X^H(t)$ .*

**Proof:** The  $L^2$  convergence is immediate. Let  $r, r' > 0$  and consider the difference  $X_r^H(t) - X_{r'}^H(t)$ . Its square mean can be written as follows:

$$\begin{aligned}\mathbb{E} |X_r^H(t) - X_{r'}^H(t)|^2 &= \mathbb{E} \left( \int_0^t (Q_r(u) - Q_{r'}(u))dZ(u) \right)^2 \\ &= \mathbb{E} \left( \int_0^t (Q_r(u) - Q_{r'}(u))dB^H(u) \right)^2 \\ &= \mathbb{E} \left( \int_0^t Q_r(u)dB^H(u) \right)^2 - \mathbb{E} \left( \int_0^t Q_{r'}(u)dB^H(u) \right)^2 \\ &\rightarrow_{r,r' \rightarrow 0} 0.\end{aligned}$$

Throughout, we denote by  $B^H$  a fBm with  $H > \frac{1}{2}$  (which could be  $Z_H^1$ ) and the last convergence to 0 is a consequence of the proof of Proposition 3.1 in [1] under assumption (8.24).

If  $p \geq 2$ , we need a different argument based on the hypercontractivity property of multiple stochastic integrals (see, e.g., [21]). For every  $p > 2$ , since  $X_r^H(t)$  is conditionally a multiple integral,

$$\begin{aligned}&\mathbb{E} |X_r^H(t) - X_{r'}^H(t)|^p \\ &\leq c_{p,k,H} \mathbb{E} \left[ \left( \int_0^t \int_0^t (Q_r(u) - Q_{r'}(u))(Q_r(v) - Q_{r'}(v))|u-v|^{2H-2} dudv \right)^2 \right]^{\frac{p}{2}} \\ &\rightarrow_{r,r' \rightarrow 0} 0.\end{aligned}$$

**Definition 8.2.** *The process  $(X^H(t))_{t \geq 0}$  from Theorem 8.1 will be called the Hermite Multifractal Random Walk (Hermite MRW or HMRW).*

**Remark 8.3.** *The limit as  $r \rightarrow 0$  is the above construction in order to reproduce the multifractal character described by the scaling property (8.1).*

### 8.3.2 PROPERTIES OF THE HERMITE MULTIFRACTAL RANDOM WALK

The first observation is that  $X^H$  is a process with stationary increments. Indeed, for any  $h, r > 0$ , since  $Q_r$  is stationary and  $Z$  has stationary increments,

$$\begin{aligned} X_r^H(t+h) - X_r^H(h) &= \int_h^{t+h} Q_r(u)dZ(u) \\ &= \int_0^h Q_r(u+h)dZ(u+h) \\ &=^{(d)} \int_0^h Q_r(u)dZ(u) \end{aligned}$$

where  $=^{(d)}$  stands for the equivalence of finite dimensional distributions. Taking the limit as  $r \rightarrow 0$ , we obtain the stationarity of the increments of the process  $X^H$ .

Other properties can be obtained by using the results in the fractional MRW case and from the fact that the covariance and the scaling properties of  $X^H$  are independent of the order  $k \geq 1$ . Let us list these properties in the following proposition:

**Proposition 8.1.** *Let  $X^H$  be the Hermite MRW from Definition 8.2.*

a. Denote for every  $l \geq 1$ ,

$$I_l := X^H((l+1)\Delta) - X^H(l\Delta).$$

Then as  $l \rightarrow \infty$ ,

$$\mathbb{E} I_l I_0 \sim \alpha_H \Delta^{2H} l^{2H-2}. \quad (8.25)$$

b. The following scaling property holds

$$(X^H(at))_{t \in [0,1]} =^{(d)} a^H (e^{\Omega_a} X^H(t))_{t \in [0,1]} \quad (8.26)$$

where  $\Omega_a$  denotes a random variable independent by  $X^H$  that satisfies (8.19).

b. Under (8.24), it holds that

$$\mathbb{E} |X^H(t)|^p < \infty$$

for every fixed  $t$  and for every  $p \geq 2$ . Moreover, for every  $t > s$ , and  $q \geq 2$

$$\mathbb{E} |X^H(t) - X^H(s)|^q = |t - s|^{q\varphi(q)+Hq} \mathbb{E} |X^H(1)|^q. \quad (8.27)$$

**Proof:** As we mentioned, the proof follows the arguments in the case of the MRW driven by the fractional Brownian motion, due to the fact that the covariance structure of the Hermite process is the same for all  $k \geq 1$ . For the sake of completeness, let us give some details. In order to compute  $\mathbb{E} I_l I_0 = \mathbb{E}(X^H((l+1)\Delta) - X^H(l\Delta))X^H(\Delta)$ , we notice that we can formally write  $X^H(t) = \int_0^t Q_0(u)dZ(u)$ , where  $Q_0$  satisfies (8.15) with  $r = 0$ . Therefore, with  $\alpha_H = H(2H - 1)$ , we have

$$\begin{aligned} \mathbb{E} I_l I_0 &= \alpha_H \int_{l\Delta}^{(l+1)\Delta} \mathbb{E} Q_0(u)Q_0(v)|u - v|^{2H-2}dudv \\ &= \alpha_H \Delta^{2H} \int_l^{l+1} \int_0^1 |u - v|^{2H-2}dudv \\ &= \frac{\alpha_H \Delta^{2H}}{2H(2H - 1)} (|l + 1|^{2H} + |l - 1|^{2H} - 2|l|^{2H}) \end{aligned}$$

and point a. follows by noting that  $(|l + 1|^{2H} + |l - 1|^{2H} - 2|l|^{2H})$  converges as  $H(2H - 1)l^{2H-2}$  when  $l$  goes to infinity. Let us justify (8.27). This follows from the stationarity of the increments of  $X^H$  and point b. above. Indeed,

$$\begin{aligned} \mathbb{E} |X^H(t) - X^H(s)|^q &= \mathbb{E} |X^H(t-s)|^q \\ &= |t - s|^{q\varphi(q)+Hq} \mathbb{E} |X^H(1)|^q \end{aligned}$$

using (8.19).

## 8.4 Financial applications

Let us now describe our numerical results. Although the theoretical results in Section 8.3 have been obtained for a general Hermite process of order

$k \geq 1$ , we will restrict our numerical study to the cases  $k = 1$  and  $k = 2$ . As mentioned before, the case  $k = 1$  corresponds to the fractional Brownian motion while for  $k = 2$  the driving process is a Rosenblatt process. While fBm is a Gaussian process, the Rosenblatt process is not Gaussian and its numerical analysis is more complex and have been developed quite recently. Our restriction to the cases  $k = 1, 2$  comes from the difficulty to simulate a Hermite process a order  $k \geq 3$ ; therefore, the results in Section 8.3 remain at this point only theoretical for  $k \geq 3$ . The simulation of a general Hermite process is still an open problem.

Our purpose is to make here a comparative numerical case between the Gaussian case (meaning that the integrator in (8.2) is the fractional Brownian motion) and the non-Gaussian case (when the integrator in (8.2) is a Rosenblatt process). We will put in light the advantages and the weak points of each model and we hope in this way to give to practitioners the possibility to choose their own model consistent with the data taken into account.

#### 8.4.1 SIMULATION OF THE HMRW

Before discussing how the fractional or Rosenblatt MRW compares with the high-frequency financial data, we will explain the main ingredients to simulate these processes. While the MRW based on fractional Brownian motion has been already simulated (see [1,3] or [15] for related results), the numerical study of the MRW based on the Rosenblatt process is completely new. Our R code is freely available and documented upon request.

The simulation of the object (8.2) contains three steps:

- The simulation of the noise  $Y$ , which is a fBm or a Rosenblatt process,
- the simulation of the integrand  $Q$ , which is an IDC (Infinite Divisible cascade), and
- the simulation of the stochastic integral (8.2).

A first step is to simulate the noise, that is, the fBm or the Rosenblatt process. Since the methodology to simulate the trajectory of a fractional Brownian motion is pretty well-known nowadays (actually, in our work, we adopt the approach in [35] to simulate a stationary Gaussian sequence), we will focus on the case of the Rosenblatt process.

The numerical analysis of the Rosenblatt process is recent. One possible approach is to use the approximation of the Rosenblatt process by two-dimensional perturbed random walks proven in [31]. This is based on the fact that the sequence

$$Z_H^n(t) = \sum_{i,j=1; i \neq j}^{[nt]} n^2 \int_{\frac{i-1}{n}}^{\frac{i}{n}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} F\left(\frac{[nt]}{n}, u, v\right) dv du \frac{\xi_i}{\sqrt{n}} \frac{\xi_j}{\sqrt{n}}, \quad (8.28)$$

where the kernel  $F$  is given by

$$F(t, y_1, y_2) = d(H) \mathbb{1}_{[0,t]}(y_1) \mathbb{1}_{[0,t]}(y_2) \int_{y_1 \vee y_2} \frac{\partial K^{H'}}{\partial u}(u, y_1) \frac{\partial K^{H'}}{\partial u}(u, y_2) du.$$

( $K$  is the usual kernel of the fBm that allows the representation  $B_t^H = \int_0^t K^H(t, s) dW_s$  and  $H' = \frac{H+1}{2}$ ), convergence weakly to the Rosenblatt process in the Skorohod topology (see [31]).

Even if this simulation approach seems quite natural, the cost is expensive due to five sums that appear in order to simulate  $Z^n$  (two sums, two integrals, and  $F$  itself are an integral).

The second approximation (introduced in the paper [33]) is related to the so-called noncentral limit theorem involving long-range dependent random variables. Let  $(\xi_k, k > 0)$  be a stationary Gaussian sequence with  $\mathbb{E}\xi_k = 0$ ,  $\mathbb{E}\xi_k^2 = 1$  and a correlation structure  $\mathbb{E}\xi_0\xi_k = k^{-(1-H)}$  with  $H \in (1/2, 1)$ , it can be shown (see [30] or [13]) that the process defined by

$$\tilde{Z}_H^n(t) = \frac{1}{d_n} \sum_{k=1}^{[nt]} (\xi_k^2 - 1) \quad (8.29)$$

converges in law to the Rosenblatt process  $Z(t)$  as  $n$  tends to infinity. Here  $d_n$  is the normalization constant chosen asymptotically proportional to  $\mathbb{E}(\sum_{i=1}^n \xi_k^2 - 1)^2$ . More precisely, we have  $d_n = n^H \left(\frac{2}{H(2H+1)}\right)^{1/2}$ .

This numerical method requires only to simulate correlated Gaussian random variables and therefore we can employ the algorithm already utilized to simulate the fBm (we recall that we employ the standard circulant matrix idea from [35]). Let us note that using (8.29), we obtain a convergence in law to the Rosenblatt process. In order to have a better approximation, we employ an idea from [2]. We can rewrite the sum (8.29) as

$$\tilde{Z}_H^n(t) = \frac{1}{2^{2jk}} \sum_{k=1}^{[2^j t]} (Y_k^2 - \mathbb{E}Y_k^2), \quad (8.30)$$

where  $Y_k$  is understood as a FARIMA(0,k,0) object. Above  $(\varepsilon_t)_t$  a Gaussian white noise and then the process  $(X_t)_{t \leq 0}$  given by

$$X_t - kX_t + \frac{k(k-1)}{2!}X_{t-2} - \dots = \varepsilon_t$$

is also a FARIMA (0,k,0) object. One can show that we have the same covariance structure as the fBm. This approximation, introduced in [2] and using a wavelet-based method, leads us to an almost sure convergence of (8.30) to the Rosenblatt process (instead of the convergence in law). Therefore, we have chosen to use an algorithm based on the sequence (8.30). Since the algorithm is a little bit more complicated, involving wavelets and multiresolution analysis (MRA) (see [27] for a nice review), we prefer to omit the details (but, as mentioned, our code is freely available).

Once the trajectory of the fBm or the Rosenblatt process numerically obtained, we have to simulate in addition the infinite divisible cascade noise (IDC in our short notation, denoted by  $Q$  in our chapter) in order to obtain the MRW. We briefly explain how we proceed. Basically, when the random measure  $M$  has normal distribution,  $Q_r$  (which is given by formula (8.13)) follows a lognormal distribution. A crucial step in the simulation is to find the covariance structure of the process  $\log Q_r(t)$ ,  $t \geq 0$ . From relation (8.12), we have

$$\rho(q) = -\log \mathbb{E} e^{qX}$$

where  $X$  is a  $\mathcal{N}(\mu, \sigma_G^2)$  random variable. Hence,

$$\varphi(q) = \rho(q) - q\rho(1) = -\log \left( \frac{\mathbb{E} e^{qX}}{(\mathbb{E} e^X)^q} \right).$$

Simple calculations show that

$$\varphi(q) = \frac{\sigma_G^2}{2}q(1-q). \quad (8.31)$$

Second, we have chosen the control measure to be  $dm(t, r) = \frac{c}{r^2}drdt$  for  $0 < r \leq 1$  [see (8.17)]. Other choices are possible but this particular case leads us to an exact invariance scaling, that is,

$$\mathbb{E}[|X(t) - X(t-\tau)|^q] = K_q t^{\zeta_q}, \quad q = 1, 2, \dots,$$

while other choice induces asymptotic invariance scaling

$$\mathbb{E}[|X(t) - X(t-\tau)|^q] \sim K_q t^{\zeta_q}, \quad q = 1, 2, \dots,$$

for some constant  $K_q$ . Then, for  $r < |\tau| \leq 1$ ,

$$m(C_r(t) \cap C_r(t + \tau)) = -c \log |\tau|. \quad (8.32)$$

Finally, using (8.31), (8.32), and (8.15), we have for any  $r \leq \tau \leq 1$

$$\text{Cov}(\log(Q_r(t)), \log(Q_r(t + \tau))) = -c\sigma_G^2 \log(|\tau|). \quad (8.33)$$

Since the IDC is normalized, the expectation of the process is, for every  $r, t$ ,

$$\mathbb{E} \log(Q_r(t)) = 0. \quad (8.34)$$

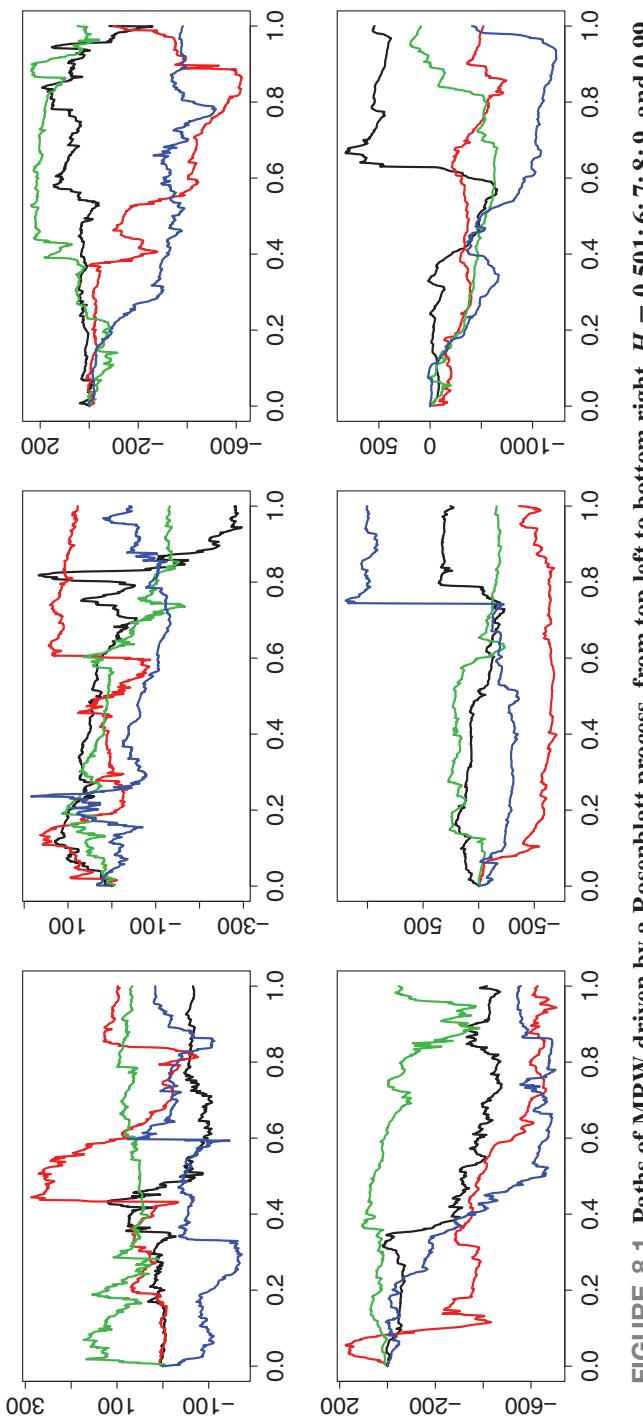
Now, the trajectory of  $Q_r$  can be easily simulated (as the exponential of a Gaussian process with covariance (8.33)). Of course, there are several parameters to take into account ( $r$ ,  $\sigma_G^2$ , and  $c$ ) and we will explain below how these parameters are chosen or estimated.

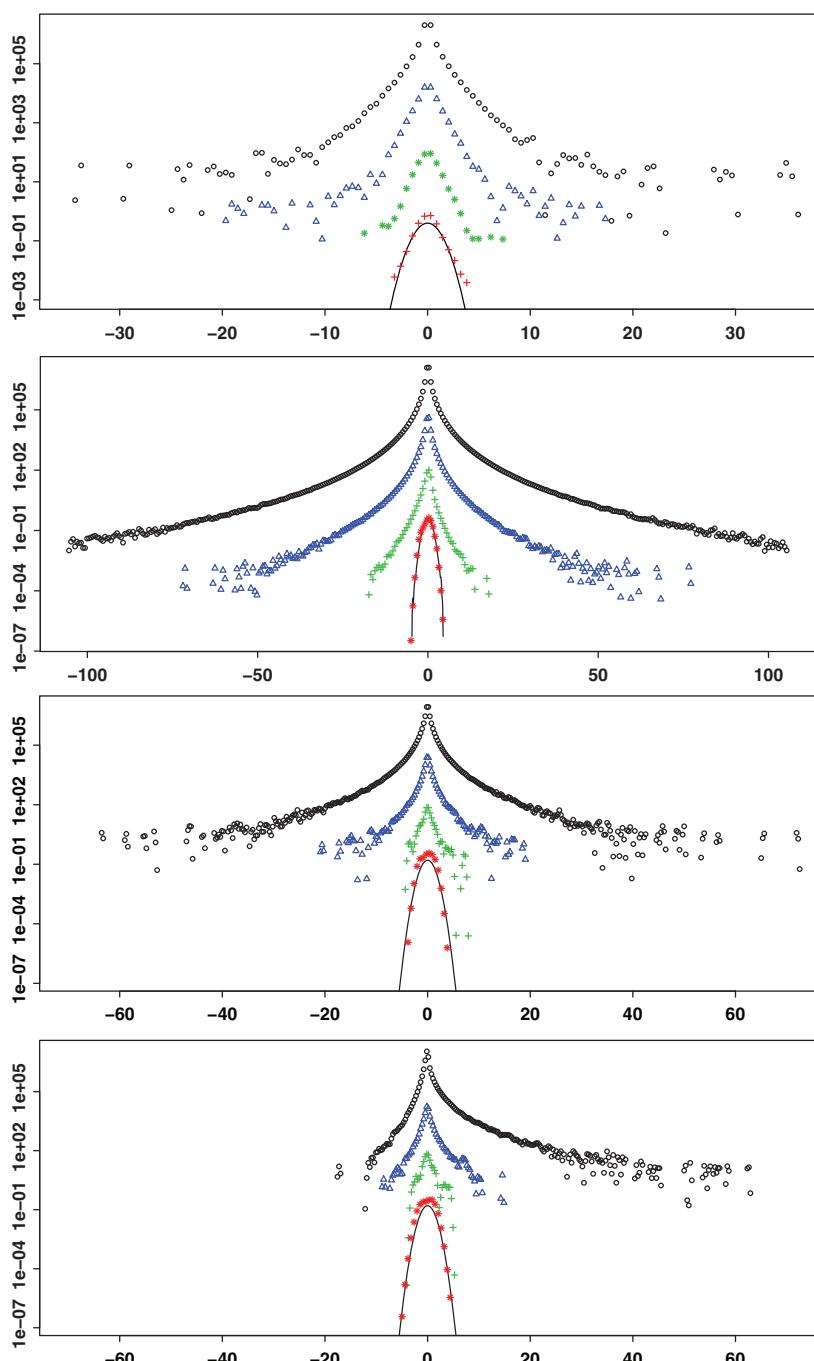
The last step is to simulate the MRW that is defined as the integral of  $Q$  with respect to the noise. This integral, which behaves as a Wiener integral, can be naturally approximated by standard Riemann sums. That is, the MRW driven by a Hermite process  $Z^H$  of self-similarity parameter  $H \in (1/2, 1)$  will be approximated by standard Riemann sums given by

$$X_r^H(t) = \sum_{s=1}^{[rt]} Q_r(s)(Z^H(s) - Z^H(s - r)). \quad (8.35)$$

We give below some realizations of MRW driven by a Rosenblatt process on Figure 8.1 for different value of  $H$ ,  $\sigma_G = 0.15$  and  $c = 1$ . Empirically, the difference between the Rosenblatt MRW and fBm-based MRW is not obvious and we will make a deeper numerical analysis in order to distinguish their properties.

If  $X$  is an MRW driven by a standard Brownian motion, the probability distribution of the increments  $X(t) - X(t - \tau)$  changes according to the considered lag  $\tau$  (this is because the MRW satisfies the so-called Castaing equation, see, e.g., [10]). We present in Figure 8.2 the estimation of the probability density in the fBm and Rosenblatt cases and a common example of real-world data (the financial index S&P 500). As we can see, in both cases, we keep this property (meaning the law of the increment changed with the lag) and thus in principle, the extended MRW (fBm or Rosenblatt) can be used to model high-frequency financial data. The MRW process can adapt to any frequency and, in particular, to high-frequency sampling. For the fBm we plot in Figure 8.2, the probability distribution for  $H = 0.5$ ,





**FIGURE 8.2** From top left to bottom right: Empirical density of S&P 500, with (from top to bottom)  $\tau = 15\text{ s}$ ,  $30\text{ min}$ ,  $4\text{ h}$ , and  $1\text{ day}$ ; empirical density of the MRW driving by a fBm with the following values of  $\tau$  from top to bottom:  $\tau = 10^{-4}, 10^{-2}, 1, 10^2$ , and  $H = 0.5$  (and  $\sigma^2 = 0.02$ ), fBm with  $H = 0.56$  and with a Rosenblatt process with  $H = 0.52$ . In any cases, the black line corresponds to the Gaussian prediction. We used the logarithmic scale.

which is a standard Brownian motion and for  $H = 0.56$  (this value is chosen because in the next section, we study in detail the particular case of the stock price for Netflix for which we found after estimation  $H = 0.56$ ). For the Rosenblatt case, also for the Netflix index, we found after estimation  $H = 0.52$  and we plot the empirical density for this value. As we can see, for this case and more generally, when  $H$  is close to 0.5, we have a positive skew. Actually, this is more or less expected since equation (8.30) indicated that the Rosenblatt process is somehow a shifted chi-square distribution, then, since the IDC and the noise are independent, the density of the HMRW is simply the product of the two densities. Note that this is not obvious but for  $H \rightarrow 1$ , the Rosenblatt process has a nonshifted density function (see [33]), which will be better for financial comparison, but the counterpart is to increase long memory in log-returns.

### 8.4.2 FINANCIAL STATISTICS

The main statistical properties of the MRW driven by the standard Brownian motion are preserved when the driving process is replaced by a Hermite process (and in particular by a Rosenblatt process or by a fractional Brownian motion). Recall that the Hermite process is defined for  $H > \frac{1}{2}$ . In this case, its increments (which corresponds to the financial returns) are positively correlated, meaning that an upward jump is followed by another upward jump. This situation is most frequently observed in the high-frequency financial data and with liquid badly arbitrated stock. When  $H$  becomes closer and closer to  $1/2$ , the correlation of the increments of the our process is close to 0 and this can be observed in our empirical study.

In the previous section, we described the algorithm to simulate the MRW driven by the fBm and by the Rosenblatt process. In this numerical analysis, several parameters appear: these are the self-similarity index  $H$ , the parameters  $c, \sigma_G$ , and  $r$  that are involving in the definition of  $Q$  (see Section 8.2).

First notice that the parameter  $c$  is always linked to parameter  $\sigma_G$ , see (8.33). The parameter  $c$  appears only in the covariance structure of  $Q$  and it is associated to the volatility  $\sigma_G$ . Without loss of generality, we can assume  $c = 1$  and then,  $c\sigma_G^2 = \sigma_G^2$ . On the other hand,  $r$  is chosen to be close to 0, we precisely consider  $r = 10^{-8}$ . This reduces the number of parameters to estimate, which are now  $H$  and  $\sigma_G$  defined above. We will explain how these parameters are estimated in practice using the theoretical results proven in our chapter.

So, the set of parameters is reduced to  $\{\sigma_G, H\}$ . We will not complete an estimation theory and asymptotic properties in this general situation (this is beyond the purpose of our work), but we can easily find an empirical approximation.

The model has to satisfy the major stylized facts, that is, the autocorrelation of log-return  $R_1$  and of the squared log-return  $R_2$  and the multifractal character with spectrum  $\zeta(q)$ ; see (8.27). These quantities are, respectively, defined by

$$\begin{aligned} R_1(k) &= \text{Cor}(\delta_\tau X(t), \delta_\tau X(t+k)), \quad k \geq 0 \\ R_2(k) &= \text{Cor}(\delta_\tau X^2(t), \delta_\tau X^2(t+k)), \quad k \geq 0 \\ \zeta(q) &= \frac{\log \frac{1}{n} \sum_{i=1}^n |\delta_\tau X(t)|^q}{\log \tau}, \end{aligned}$$

where  $\text{Cor}$  denotes correlation operator and  $\delta_\tau X(t) = X(t + \tau) - X(t)$ . We used a least square estimation procedure; that is, we find the values minimizing the mean squared error

$$\begin{aligned} (\tilde{\sigma}_G, \tilde{H}) = \arg \max_{\sigma_G, H} \{ & ||R_1(k) - \hat{R}_1(k)||^2 + ||R_2(k) - \hat{R}_2(k)||^2 \\ & + ||\zeta(q) - \hat{\zeta}(q)||^2 \} \end{aligned} \quad (8.36)$$

where  $\hat{R}_1$ ,  $\hat{R}_2$ , and  $\hat{\zeta}(q)$  are the empirical values. We have minimized this sum, using a genetic algorithm.

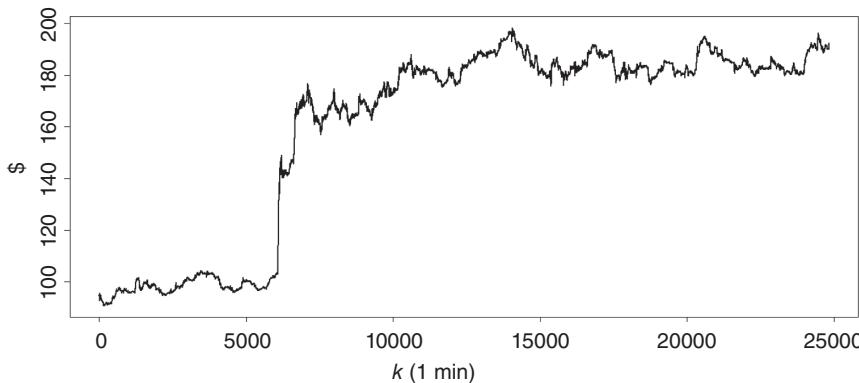
Note an alternative estimation procedure. Note that the autocorrelation function of the MRW driven by a Hermite process is given by (8.25) and this formula depends only on  $H$ . Therefore, we can simply obtain from the empirical autocorrelation the index  $H$ . Concerning  $\sigma_G$ , recall that we chose a log-normal distribution for the IDC with  $\varphi(q) = \frac{\sigma_G^2}{2} q(1 - q)$ . From equation (8.27), we have

$$\zeta(q) = \frac{\sigma_G^2}{2} q^2 (1 - q) + Hq$$

and this allows to estimate the volatility.

We compared the two approaches. Using the second procedure, the computing cost is strictly inferior, but we think that the result are poorest. Therefore, the figures presented in the chapter are based on the least squares estimation approach.

We will apply our model to the Nasdaq 100 index from January 2 to March 28, 2013, at a 1-min frequency, meaning about 23,000 observations



**FIGURE 8.3** NFLX from January 2 to March 28, 2013, at 1-min frequency, 24,832 observations.

in the series. One of the most interesting cases is the asset Netflix, Inc. (NFLX), that we plot in Figure 8.3.

After parameter estimation for the HMRW driven by a Rosenblatt process using (8.36), we found  $H = 0.52$  and  $\sigma_G^2 = 0.02$  while for the HMRW driven by a fBm  $H = 0.56$  our estimation gives and  $\sigma_G^2 = 0.2$ .

In Figure 8.4, we give the autocorrelation functions of the returns and of the squared returns and below their corresponding values obtained via HMRW.

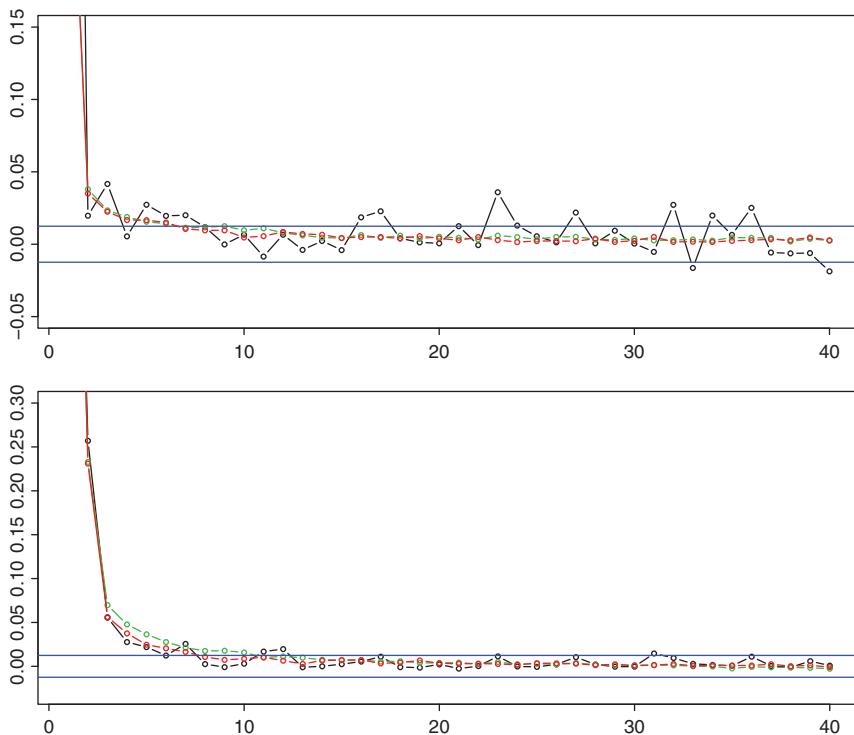
Finally, in Figure 8.5, we trace the multifractal spectrum for  $q = 0, \dots, 5$ . As we can notice, the multifractal nature of the spectrum can be well observed and the spectrum of HMRW well describes the spectrum of the asset.

Since the Rosenblatt process is not defined for  $H = 1/2$ , our theory cannot be applied. Nevertheless, taking the limit  $H$  close to  $1/2$ , for example,  $H = 0.501$ , we find significant empirical results. See Figure 8.6 for the case of the Electronics Art Inc. (EA) asset price with  $H = 0.501$ .

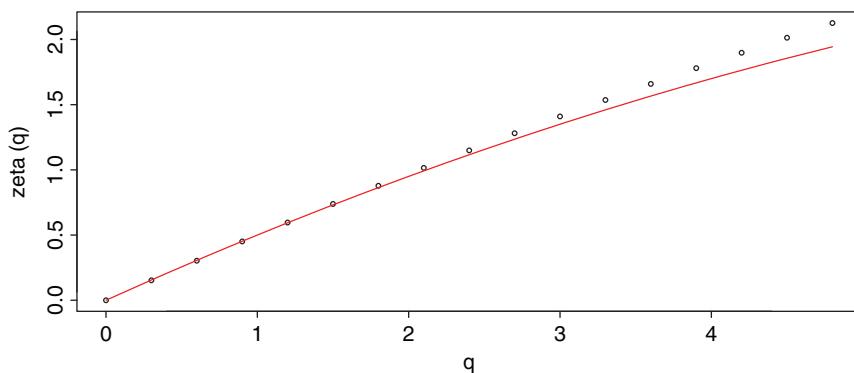
In Figure 8.7, we compare the autocorrelations functions and the spectra of the EA index and those obtained using HMRW.

## 8.5 Concluding remarks

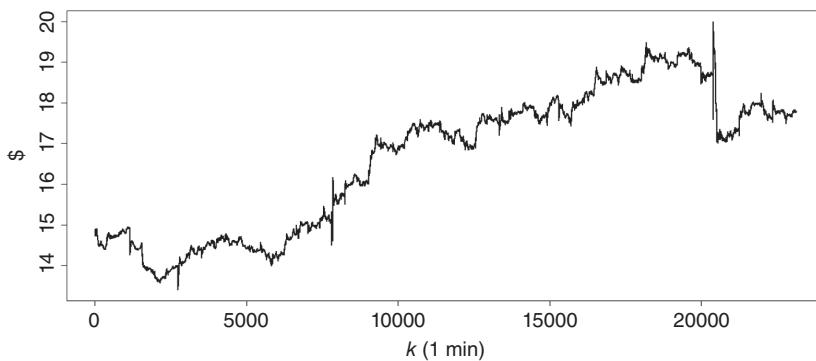
The multifractal models capture the main stylized facts observed in empirical data. The multifractality itself, as defined by the property (8.1), became a stylized fact for the financial models. Our primary purpose here was



**FIGURE 8.4** Autocorrelation function of the returns (left) and of the squared returns (right) of NFLX. In red, the average of the autocorrelations of 10 paths of the HMRW driven by an fBm ( $H = 0.56$ ), with squared increments (right) and with standard increments (left). In green for the HRMW driven by a Rosenblatt process ( $H = 0.52$ ).



**FIGURE 8.5** Multifractal spectrum of NFLX (black dots) and of HMRW driven by an mbf in red with  $H = 0.56$  and  $\sigma_G = 0.19$ .  $q$  varies from 0. to 5.

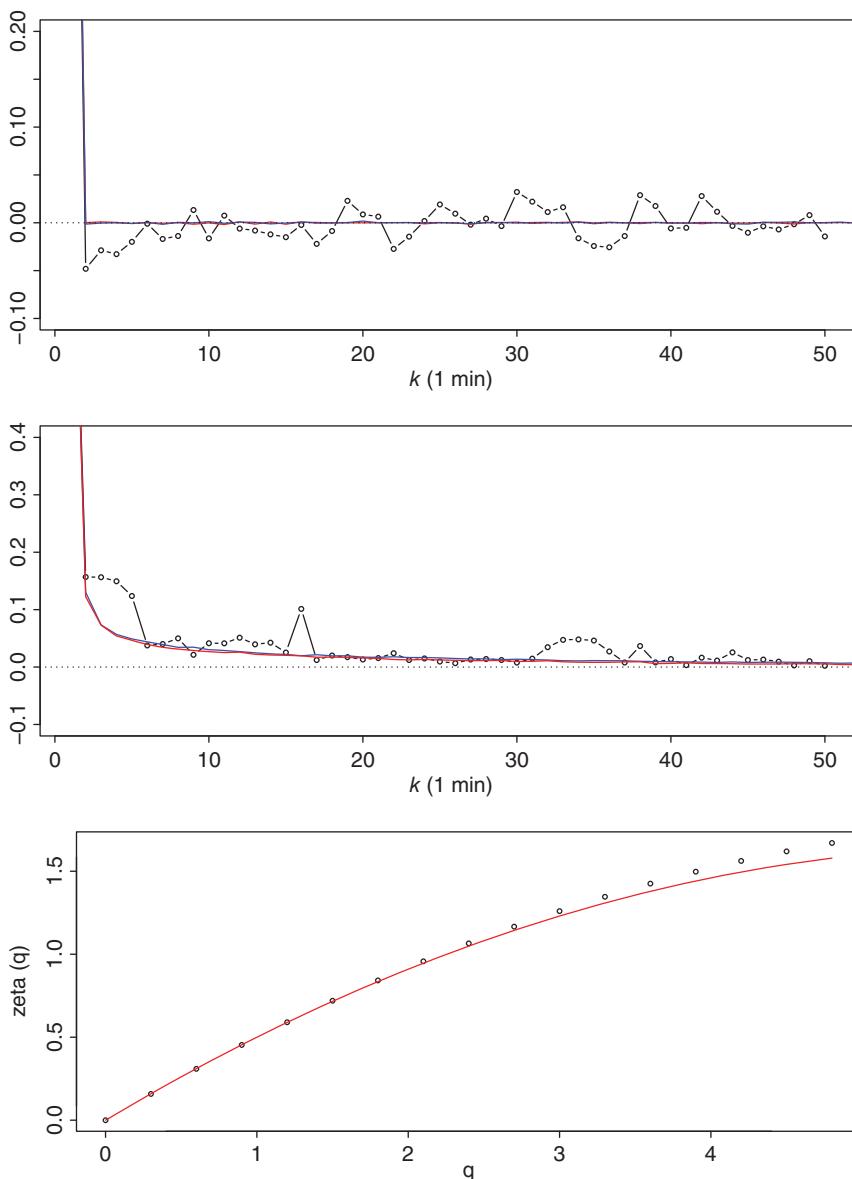


**FIGURE 8.6 EA from January 2 to March 28, 2013, at 1 min, 23,139 observations.**

to enlarge the class of multifractal models and in this way to enlarge the modelling tool kit. We constructed models of the form (8.2), which are integrals of an Infinite Divisible Cascade  $Q$  with respect to a Hermite process of order  $k$ . The most interesting cases are  $k = 1$  (the integrator in (8.2) is a fractional Brownian motion) and  $k = 2$  (the integrator in (8.2) is the Rosenblatt process). We show in Section 8.3 that these integrals are well-defined and they present the multifractal structure; see Proposition 8.1 in Section 8.3.

We explain in Section 8.4 how the object (8.2) can be simulated, for  $k = 1$  and  $k = 2$ . While the fBm case  $k = 1$  was already studied before, the algorithm to simulate the MRW based on the Rosenblatt noise is new. As explained in the first part of Section 8.2, in the case  $k = 2$ , the algorithm is more complex, because of the non-Gaussian structure of the Rosenblatt process. Also, in Section 8.4, we compare these models with the real-world data in the high-frequency context. The main conclusion of our numerical analysis is the following: both models (obtained for  $k = 1$  and  $k = 2$ ) capture some of the stylized facts, such as autocorrelation of the returns or of the squared returns or the multifractal spectrum. On the other hand, in the Rosenblatt model with  $H$  close to one half, the probability distribution presents a skewness, which is not the case for financial data. This skewness is less visible when  $H$  is close to 1.

Taking into account the above remarks and the complexity of the algorithm, the choice of the fBm instead of a Rosenblatt process appears to be much more appropriate in general. Nevertheless, the Rosenblatt process with a self-similarity index could be used in situations when the non-Gaussian character of the data appears to be more pronounced.



**FIGURE 8.7** Down: autocorrelation of the returns (left) and of the squared returns (right) from EA. In red, the average of 10 paths of the HMRW with squared increments (right) and standard increments (left). Up: Multifractal spectrum of EA (black dots) and of HMRW (red) with  $H = 0.501$  and  $\sigma_G = 0.19$ .  $q$  varies from 0 to 5.

## References

1. P. Abry, P. Chainais, L. Coutin, and V. Pipiras (2009): *Multifractal random walks as fractional Wiener integrals*. IEEE Transactions on Information Theory, 55(8), 3825–3846.
2. P. Abry and V. Pipiras (2006): *Wavelet-based synthesis of the Rosenblatt process*. Signal Processing, 86, 2326–2339.
3. E. Bacry, L. Duvernet, and J.F. Muzy (2012): *Continuous-time skewed multifractal processes as a model for financial returns*. Journal of Applied Probability, 49, 482–502.
4. E. Bacry and J.-F. Muzy (2002): *Multifractal stationary random measures and multifractal walks with log-infinitely divisible scaling laws*. Physical Review E, 66, 056121.
5. E. Bacry, A. Kozhemyak, and J.-F. Muzy (2008): *Continuous cascade models for asset returns*. Journal of Economic Dynamics and Control, 32(1), 156–199.
6. E. Bacry, J.F. Muzy, and J. Delour (2001): *Multifractal random walks*. Physical Review E, 64, 026103–026106.
7. J.M. Bardet, G. Lang, G. Oppenheim, A. Philippe, and M. Taqqu (2003): *Generators of long-range dependent processes: A survey*. Long-range Dependence: Theory and Applications, Birkhauser.
8. J. Barral and B. Mandelbrot (2002): *Multifractal products of cylindrical pulses*. Probability Theory and Related Fields, 124, 409–430.
9. L. Calvet and A. Fisher (2008): *Multifractal Volatility: Theory, Forecasting and Pricing*. Cambridge University Press.
10. B. Castaing (1996): *The temperature of turbulent flows*. Journal de Physique II France, 6, 105–114.
11. P. Chainais, R. Riedi, and P. Abry (2005): *On non-scale invariant infinitely divisible cascades*. IEEE Transactions on Information Theory, 51(3), 1063–1083.
12. A. Chakraborti, I. Muni Toke, M. Patriarca, and F. Abergel (2011): *Econophysics review: I. Empirical Facts*. Quantitative Finance, 11(7), 991–1012.
13. R.L. Dobrushin and P. Major (1979): *Non-Central Limit Theorems for Non-Linear Functionals of Gaussian Fields*. Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete, 50(N1), 27–52.

14. R.F. Engle (1982): *Autoregressive conditional heteroskedasticity with estimates of the variance of the U.K. inflation*. *Econometrica*, 987–1108.
15. A. Fauth and C.A. Tudor (2012): *Multifractal random walks with fractional Brownian motion via Malliavin calculus*. *IEEE Transactions on Information Theory*, 60 (2014), 3, 1963–1975.
16. A. Fisher, L. Calvet, and B. Mandelbrot (1997): *Multifractality of Deutschemark/US Dollar Exchange Rates*. Cowles Foundation Discussion Paper No. 1166.
17. S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge (1996): *Turbulent Cascades in Foreign Exchange Markets*. *Nature*, 381, 767–770.
18. P. Gopikrishnan, V. Plerou, L.A. Amaral, M. Meyer, and H.E Stanley (1999): *Scaling of the Distribution of Fluctuations of Financial Market Indices*. *Physical Review E*, 60, 5305–5316.
19. F. Lillo and J. D. Farmer (2004): *The long memory of the efficient market*. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 1558–3708.
20. C. Ludena (2008):  *$L^p$  variations for multifractal fractional random walks*. *Annals of Applied Probability*, 18(3), 1138–1163.
21. Major, P. (2005). *Tail behavior of multiple random integrals and  $U$ -statistics*. *Probability Surveys*, 2, 448–505.
22. B. Mandelbrot (1963): The variation of certain speculative prices. *Journal of Business*, 36, 394.
23. J.F. Muzy, J. Delour, and E. Bacry. (2000): *Modeling fluctuations of financial time series: from cascade process to stochastic volatility model*. *European Physical Journal B*, 17, 537–548.
24. D. Nelson (1991): *Conditional heteroskedasticity in asset returns: A new approach*. *Econometrica*, 59, 347–370.
25. D. Nualart (2006): *The Malliavin Calculus and Related Topics*. Second Edition, Springer.
26. M. Maejima and C.A. Tudor (2007): *Wiener integrals and a non-central limit theorem for Hermite processes*. *Stochastic Analysis and Applications*, 25(5), 1043–1056.
27. S. Mallat (2008): *A Wavelet Tour of Signal Processing*, Academic Press, 3rd edition, 2008.

28. V. Pipiras and M. Taqqu (2001): *Integration questions related to the fractional Brownian motion.* Probability Theory and Related Fields, 118(2), 251–281.
29. M. Rosenblatt (1961): *Independence and dependence.* Proceedings of 4th Berkeley Symposium on Mathematics and Statistics, II, 431–443.
30. M. Taqqu (1975): *Weak convergence to the fractional Brownian motion and to the Rosenblatt process.* Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete, 31, 287–302.
31. S. Torres and C.A. Tudor (2009): *Donsker type theorem for the Rosenblatt process and a binary market model.* Stochastic Analysis and Applications, 27(3), 555–573.
32. C.A. Tudor (2006): *Analysis of the Rosenblatt process.* ESAIM Probability and Statistics, 12, 230–257. MR2374640
33. M.S. Veillette and M.S. Taqqu (2013): *Properties and numerical evaluation of the Rosenblatt distribution.* Bernoulli, 19(3), 982–1005.
34. J. Voit (2005): *The Statistical Mechanics of Financial Markets.* Springer, 3rd edition, 378.
35. A. Wood and G. Chan (1994): *Simulation of stationary Gaussian processes in  $[0, 1]^d$ .* Journal of Computational and Graphical Statistics, 3(4), 409–432.

## Chapter Nine

# Interpolating Techniques and Nonparametric Regression Methods Applied to Geophysical and Financial Data Analysis

K. Basu<sup>1</sup> and Maria C. Mariani<sup>2</sup>

<sup>1</sup>*Department of Mathematics, Occidental College, Los Angeles, CA, USA*

<sup>2</sup>*Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX, USA*

### 9.1 Introduction

In the past few years, researchers have a renewed interest in analyzing “critical value phenomena,” and in ([12, 15]), the authors applied three powerful modeling techniques to predict and estimate parameters for these major-events (major earthquakes or financial market crashes) efficiently. Powerful modeling techniques for earthquake data can be found in [7, 22]. In particular, the role of stress transfer leading to earthquake and the earthquake hazards after the mainshock is discussed in detail in [20, 21]. There are several models available in literature that deal with both space and time aspects of the earthquake data such as [19], although our approach to deal with geophysical and high-frequency finance data is based

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

on a completely different approach. Instead of estimating the major earthquake date (i.e., deal with the time series data), we implemented a spatial analysis where we have fixed the time (in this case, the year) and we have collected the earthquake data from different locations of a particular geographical region. On the basis of these data trends, we applied two different versions of the local weighted regression method to estimate the magnitude of the earthquake in the same year at a location whose data were not used in the calculations. The method has proved to be efficient and useful. We also applied a similar smoothing technique for a time series data arising in the financial market for five different financial institutions. In this case, we used four different versions of weighted local regression curves that fitted very well the financial data.

We first introduce nonparametric regressions techniques, specially the local regression method [11]. We focus our work in two different applications: geophysical data and financial high-frequency data. Spatial analysis is performed to different data sets corresponding to different locations by varying the latitude and longitude, to estimate the magnitude of the earthquake at any given time. The application of these models was proved to be efficient by looking at the results. A modified local regression method that is suitable for applications to one-dimensional input data is introduced, and four different modified versions are applied to the high-frequency financial data leading to the Bear Stearns crash, which occurred in mid-March 2008. One of our main goals is to find out any possible future avenue where we can use these deterministic modeling techniques along with traditional stochastic models.

In the second part of our work, we present five powerful interpolation methods for two-dimensional input data. Numerical estimations are implemented with these interpolation techniques and, in addition, we present the “goodness of fit” parameters such as sum of squares of error (SSE) and  $R$ -square value. Looking at the parameters, it seems that some of the techniques produced more promising results than others. This is probably due to the fact that these techniques are very much dependent on the local structure of the data. Although the overall application has opened a new perspective to analyze the data.

The results of our numerical methods applied to both earthquake data and financial data indicate that the weighted local regression approach and the interpolation methods are very powerful techniques to extract information from large-scale data sets. The simplicity yet robustness of these methods makes them easy to apply and the estimations are very

accurate. In this chapter, we applied these deterministic modeling approach to look for any alternative modeling endeavor dealing with similar data sets. We believe that these parameter approximation methods with the help of all the well-established stochastic processes analysis can produce better models in the future. In the conclusions, we discuss the advantages and disadvantages of using these modeling techniques and indicate future work directions. We recall that in this work, it is not our intention to perform a statistical data analysis of the given data, but rather to present an innovative and mathematical model approach to handle these kinds of data set.

## 9.2 Nonparametric regression models

The traditional nonlinear regression model fits the following model

$$y_i = f(\beta, \mathbf{x}_i') + \varepsilon_i$$

where  $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$  is a vector of parameters to be estimated and  $\mathbf{x}_i' = (x_1, x_2, \dots, x_k)$  is a vector of predictors for the  $i$ th of  $n$  observations; here we assume that the corresponding errors  $\varepsilon_i$  for this model are independently and normally distributed with mean 0 and constant variance  $\sigma^2$ . The function  $f(\cdot)$ , relating the average value of the response  $y$  to the predictors, is generally specified in advance, whereas it is a linear or nonlinear function.

The *general nonparametric regression model* is written in a similar manner, but the function  $f$  is left unspecified,

$$\begin{aligned} y_i &= f(\mathbf{x}_i') + \varepsilon_i \\ &= f(x_{i1}, x_{i2}, \dots, x_{ik}) + \varepsilon_i \end{aligned}$$

Moreover, the object of the nonparametric regression is not to estimate the parameters, but rather estimate the function  $f(\cdot)$  itself. Most of the nonparametric regression techniques implicitly assume that the function  $f(\cdot)$  is a smooth and continuous function, except [16, 17]. In nonlinear regression, it is a standardized assumption that  $\varepsilon_i \sim \text{NID}(0, \sigma^2)$ .

An important special case of the general model is nonparametric simple regression, where there is only one predictor:

$$y_i = f(x_i) + \varepsilon_i$$

The nonparametric simple regression is often termed as “scatterplot smoothing” because an important application is to tracing a smooth curve through a scatterplot of  $y$  against  $x$ .

As it is complicated and difficult to fit a general nonparametric regression model when there are more than one predictor and to display the fitted model when there are more several predictors, more restrictive models have been developed. As an example, we can present *additive regression model*,

$$y_i = \alpha + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_k(x_{ik}) + \varepsilon_i$$

where the partial regression functions  $f_j(\cdot)$  are assumed to be smooth and have to be estimated by using the given data. This model is substantially more restrictive than the general nonparametric model, but less restrictive than the linear regression model, which assumes that all the partial regression functions are linear.

There are some variations of the additive regression model, such as semiparametric models, in which some of the predictors are entered linearly, for example,

$$y_i = \alpha + \beta_1 x_{i1} + f_2(x_{i2}) + \cdots + f_k(x_{ik}) + \varepsilon_i$$

(particularly useful when some of the predictors are factors), and models in which some predictors enter into interactions, which appears as higher-dimensional terms in the model, for example,

$$y_i = \alpha + f_{12}(x_{i1}, x_{i2}) + f_3(x_{i3}) + \cdots + f_k(x_{ik}) + \varepsilon_i$$

All of these models extend to *generalized nonparametric regression* straightforward. The random and link components are as in the generalized linear models, with the linear predictor of the generalized linear model

$$\eta_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

is replaced, for example, by an unspecified smooth function

$$\eta_i = f(x_{i1}, x_{i2}, \dots, x_{ik})$$

in the most general case, or by a sum of smooth partial-regression functions

$$\eta_i = \alpha + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_k(x_{ik})$$

for the generalized additive model.

A detailed discussion of all the aforementioned nonparametric models can be found in [4, 5].

## 9.2.1 LOCAL POLYNOMIAL REGRESSION

There are several approaches to evaluate the nonparametric regression models, such as local polynomial regression and smoothing splines. We will discuss the local regression method that is the main focus of this work.

### 9.2.1.1 Simple regression

Under simple regression, we are looking to fit the model:

$$y_i = f(x_i) + \varepsilon_i$$

Let us concentrate in evaluating the regression function at a particular value of  $x$ , say focal point,  $x^*$ . (Finally, we will fit the model at a representative range of values of  $x$  or simply at the  $n$  number of observations,  $x_i$ .) We proceed to perform a  $p$ th-order weighted least-squares polynomial regression of  $y$  on  $x$ ,

$$y_i = a + b_1(x_i - x^*) + b_2(x_i - x^*)^2 + \cdots + b_p(x_i - x^*)^p + e_i$$

we choose a weight function to give the greatest weight to the observations that are closest to the focal point,

$$W(z) = \begin{cases} (1 - |z|^3)^3 & \text{for } |z| < 1 \\ 0 & \text{for } |z| \geq 1 \end{cases}$$

Let  $z_i = \frac{(x_i - x^*)}{h}$  be the scaled distance between the predictor value for the  $i$ th observations,  $x_i$ , and the focal point. Here  $h$  is the half-width of the window enclosing the observations for the local regression centered at  $x^*$ . The fitted value at  $x^*$ , that is, the estimated height of the regression curve, is simply  $\hat{y}^* = a$  (produced conveniently by having centered the predictor  $x$  at the focal point  $x^*$ ).

It is typical to adjust  $h$  so that each local regression includes a fixed proportion  $s$  of the data; then,  $s$  is called the span of the local-regression smoother. The larger the span, the smoother the fit; conversely, the larger the order of the local regression  $p$ , the more flexible the smooth.

### 9.2.1.2 Multiple regression

The nonparametric multiple regression model is

$$\begin{aligned}y_i &= f(\mathbf{x}_i') + \varepsilon_i \\&= f(x_{i1}, x_{i2}, \dots, x_{ik}) + \varepsilon_i\end{aligned}$$

Simple extension of the local-polynomial approach to multiple regression is conceptually very simple but can run into practical difficulties.

Followings are the ordered steps to execute the technique:

- (i) The first step is to define a multivariate neighborhood around the focal point  $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_k^*)$ . The default approach for *Loess* function is to employ scaled Euclidean distances:

$$D(\mathbf{x}_i, \mathbf{x}^*) = \sqrt{\sum_{j=1}^k (z_{ij} - z_j^*)^2}$$

where the  $z_j$  are the scaled predictors,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

Here  $\bar{x}_j$  is the mean of the  $j$ th predictor and  $s_j$  is its standard deviation.

- (ii) Weights are defined by using the scaled distances:

$$w_i = W \left[ \frac{D(\mathbf{x}_i, \mathbf{x}^*)}{h} \right]$$

where  $W(\cdot)$  is a suitable weight function, such as a tricube, in which  $h$  is the half-width of the window of the neighborhood. As in local regression,  $h$  may be adjusted to define a neighborhood including the nearest neighborhood [ $ns$ ] of the focal point (where the square bracket denotes the box function).

- (iii) Perform a weighted polynomial regression of  $y$  on the  $x$ 's; for example, a local linear fit takes the following form:

$$y_i = a + b_1(x_{i1} - x_1^*) + b_2(x_{i2} - x_2^*) + \dots + b_k(x_{ik} - x_k^*) + e_i$$

The fitted value at  $x^*$  is then simply  $\hat{y}^* = a$ .

- (iv) The procedure is repeated for representative combinations of predictor values to create a regression surface.

### 9.2.2 LOWESS/LOESS METHOD

Lowess and Loess (locally weighted scatterplot smoothing) are two strongly related nonparametric regression methods that include multiple regression models in a  $k$ -nearest-based metamodel. “Loess” is a much generic version of “Lowess”; its name arises in “LOcal regrESSion.” They are constructed on both the linear and nonlinear least square regressions. These methods are more powerful and effective for studies in which the classical regression procedures cannot produce satisfactory results or cannot be efficiently applied without undue labor. Loess incorporates much of the simplicity of the linear least squares regression with some room for nonlinear regression. It works by fitting simple models to localized subsets of the data to construct a function that describes pointwise the deterministic part of the variation of data. The main advantage of this method is that we need no data analyst to determine a global function of any form to fit a model to the entire data set, only to the segment of data.

This method involves a lot of increased computation as it is a computationally intense procedure. But in the modern computational setup, Lowess/Loess has been designed to take the advantage of current computational ability to fullest advantage in order to achieve goals not easily achieved by traditional methods.

A smooth curve through a set of data points obtained with a statistical technique is called a **Loess curve**, particularly when smoothed value is obtained by a weighted quadratic least squares regression over the span of values of the  $y$ -axis scattergram criterion variable. Similarly, the same process is called **Lowess curve** when each smoothed value is given by weighted linear least squares regression over the span, although some literature presents **Lowess** and **Loess** as synonymous. Some key features of the local regression models are presented in the next paragraphs.

Lowess/Loess, originally proposed by Cleveland [1] and further improved by Cleveland and Devlin [2], specifically denoted a method that is also known as locally weighted polynomial regression. At each point in the data set, a low-degree polynomial is fitted to a subset of the data, with explanatory variable values near the point whose response is being estimated. Weighted least square method is implemented to fit the polynomial where more weightage is given to the points near the point whose response is being estimated and less importance to the points further away. The value of the regression function for the point is then evaluated by calculating the local polynomial using the explanatory variable values for

that data point. One needs to compute the regression function values for each of the  $n$  data points in order to complete the Lowess/Loess processes. Many of the details of these methods, such as degree of the polynomial model and weights, are flexible.

The subset of data used for each weighted least square fit in Lowess/Loess is decided by a nearest neighbors algorithm. One can pre-determine the specific input for the process called the “bandwidth” or “smoothing parameter” that determines how much of the data is used to fit each local polynomial according to the need. The smoothing parameter  $\alpha$  is restricted between the value  $\frac{(\lambda+1)}{n}$  and 1, with  $\lambda$  denoting the degree of the local polynomial. The value of  $\alpha$  is the proportion of data used in each fit. The subset of data used in each weighted least squares fit comprises the  $n\alpha$  points (rounded to the next larger integer) whose explanatory variable values are closest to the point at which the response is being evaluated.

The smoothing parameter  $\alpha$  is named because it controls the flexibility of the Lowess/Loess regression function. Large values of  $\alpha$  produce the smoothest functions that wiggle the least in response to fluctuations in the data. The smaller  $\alpha$  is, the closer the regression function will conform to the data, but using a very small value for the smoothing parameter is not desirable since the regression function will eventually start to capture the random error in the data. For the majority of the Lowess/Loess applications,  $\alpha$  values are chosen from the range of 0.25–0.5. First- and second-degree polynomial are used to fit local polynomials to each subset of data. That means, either a locally linear or quadratic function, are most useful; using a zero polynomial turns Lowess/Loess into a weighted moving average. Such a simple model might work well for some situations and may approximate the underlying function well enough. Higher-degree polynomials work great in theory, but the Lowess/Loess methods are based on the idea that any function can be approximated in a small neighborhood by a low-degree polynomial and simple models can be fit data easily. High-degree polynomials would tend to overfit the data in each subset and are numerically unstable, making precise calculation almost impossible.

As mentioned above, Lowess/Loess methods use the traditional tri-cubed weight function. However, any other weight function that satisfies the properties listed in Cleveland [1] could also be taken into consideration. The process of calculating the weight for a specific point in any localized subset of data is done by evaluating the weight function at the distance between the point and the point of estimation, after scaling the distance so that the maximum absolute distance over all possible points in the subset of data is exactly one.

### 9.2.3 NUMERICAL APPLICATIONS

In this section, we present the numerical simulations performed with Lowess/Loess methods applied to geophysical data and high-frequency financial data.

#### 9.2.3.1 Application to geophysics

The geological data was obtained at the U.S. Geological Survey (USGS) from January 1, 1973, to November 9, 2010 ([23]). The data contains information about the date, longitude, latitude, and magnitude of each recorded earthquake in the region.

The location of the major earthquake chosen defines the area studied. The area chosen cannot be too small (lack of data) or too big (noise from unrelated events). The data is obtained using a *square* centered at the coordinates of the major event. The sides of the square were usually chosen as  $\pm 0.1^\circ$ – $0.2^\circ$  in latitude and  $\pm 0.2^\circ$ – $0.4^\circ$  in longitude. A segment  $0.1^\circ$  of latitude at the equator is  $\approx 11.11$  km  $\approx 6.9$  miles in length.

The earthquake magnitude is the recorded data used in the analysis. The policy of the USGS regarding recorded magnitude is the following [23]:

- Magnitude is a dimensionless number between 1 and 12.
- The reported magnitude should be moment magnitude, if available.
- The least complicated, and probably most accurate terminology is to just use the term “*magnitude*” and to use the symbol  $M$ .

In the numerical study, we use data collected from different locations at a given time to estimate the magnitude of the earthquake at a given location, where the real magnitude is known. The magnitude is recorded in the data used and where available moment magnitude is used. For more information, we refer to the specific USGS documentation available at: [http://earthquake.usgs.gov/aboutus/docs/020204mag\\_policy.php](http://earthquake.usgs.gov/aboutus/docs/020204mag_policy.php).

To study the efficiency and accuracy of the Lowess/Loess methods on the geophysical data set, we have applied both processes to the same data and listed the estimated magnitude of the earthquake in 10 different geographical locations. Moreover, the relative errors with respect to its actual given magnitude were computed to compare the strength between linear least square local fit (Lowess) versus quadratic least square local

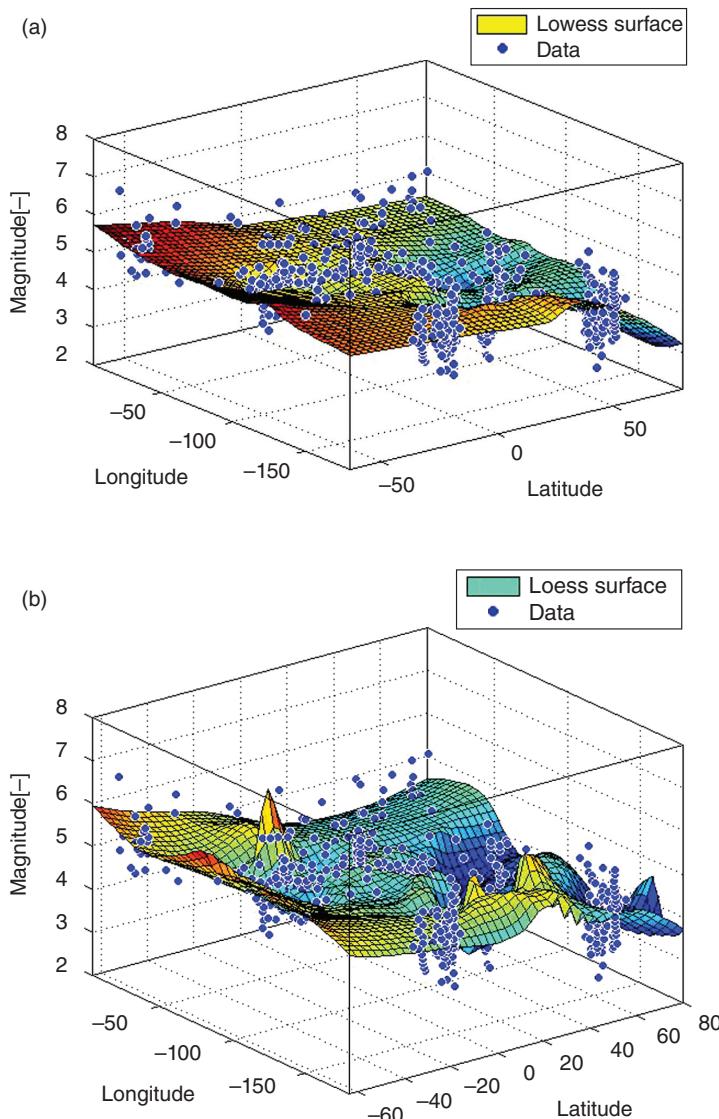
fit (Loess). In the calculation, we used “lowess” and “loess” built in subroutine under curve fitting tool in Matlab. Results are presented for five randomly selected years where the magnitude for the earthquake in different locations is available. In Figures 9.1, 9.2, 9.3, 9.4, and 9.5 below, we display typical results for the earthquake estimation surface simulated by the Lowess/Loess techniques in some areas of California. We used data from 1973, 1979, 1988, 1996, and 2008. Real value data were used to draw the estimation surface.

The data for these figures was measured in the Western Hemisphere (i.e.,  $-180^\circ$  in longitude). Cross-validation was used to get a better estimate of the estimation error. Almost 25% of the randomly selected original data was estimated and only 10 of the random results are presented in the table format. The entire set of earthquakes analyzed (from 1973, 1978, 1988, 1996, and 2008) is presented in Tables 9.1, 9.2, 9.3, 9.4, and 9.5, respectively. Each table displays the date of the earthquake, the actual magnitude, estimated magnitude of the earthquake and finally the relative precision taking into consideration the actual magnitude of the earthquake.

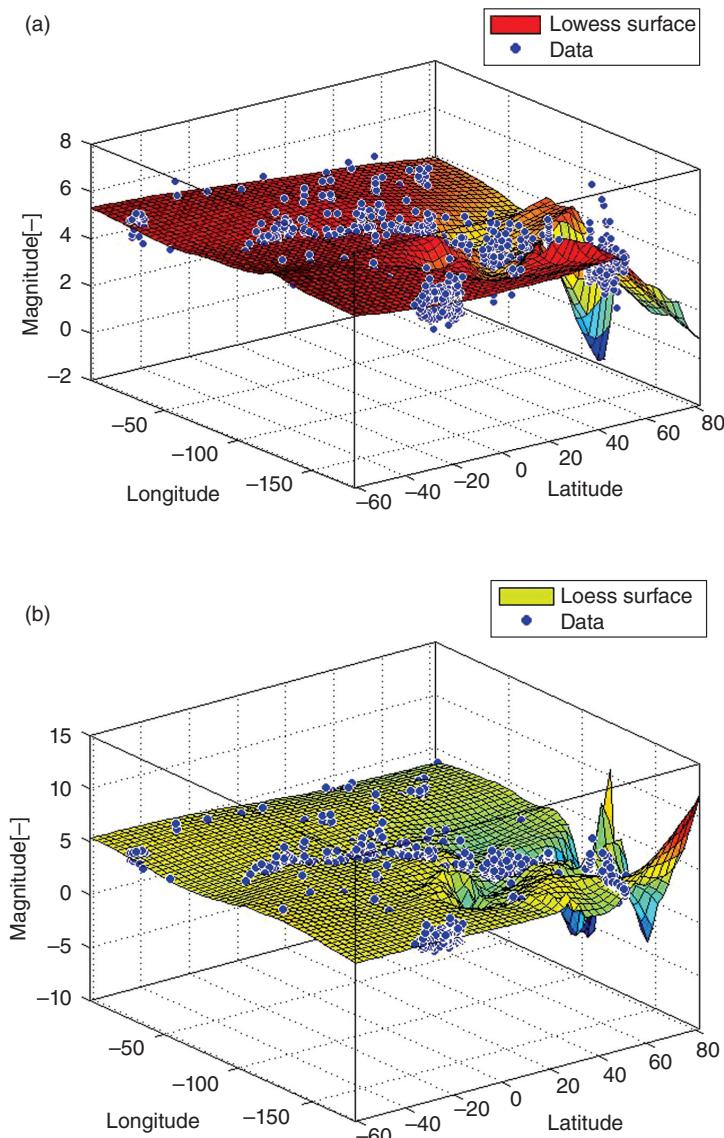
### 9.2.3.2 Application to financial data sampled with high frequency

In this section, we study the high-frequency data corresponding to the collapse of the Bear Stearns in March 2008. The data involves the week (five trading days) March 10–14, 2008, before the merging announcement over the weekend as well as the two following trading days March 17 and 18. On Friday, March 14, 2008, at around 9:14 a.m., JP Morgan Chase and the Federal Reserve Bank of New York announced an emergency loan (of about 29 billion, term undisclosed) to prevent the firm from becoming insolvent. This bailout was declared to prevent the very likely crash of the market as a result of the fall of one of the biggest investment banks at the time. This measure proved to be insufficient to keep the firm alive and 2 days later on Sunday March 16, 2008, Bear Stearns signed a merger agreement with JP Morgan Chase essentially selling the company for \$2 a share (price revised on March 24 to \$10/share). The same stock traded at \$172 in January 2007 and \$93 a share in February 2007. Today, this collapse is viewed as the first sign of the risk management meltdown of investment bank industry in September 2008 and the subsequent global financial crisis and recession.

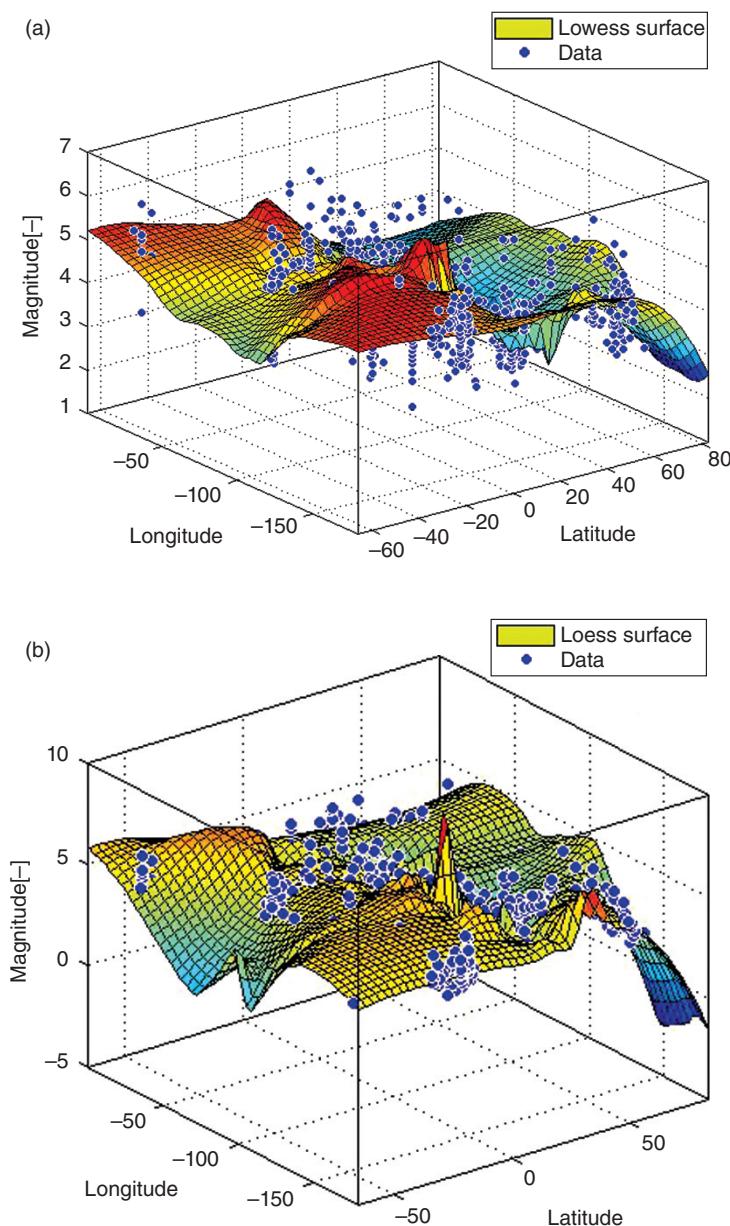
In this part of the chapter, we study how to fit different local regression models (one-dimensional version to deal with time-series data) into



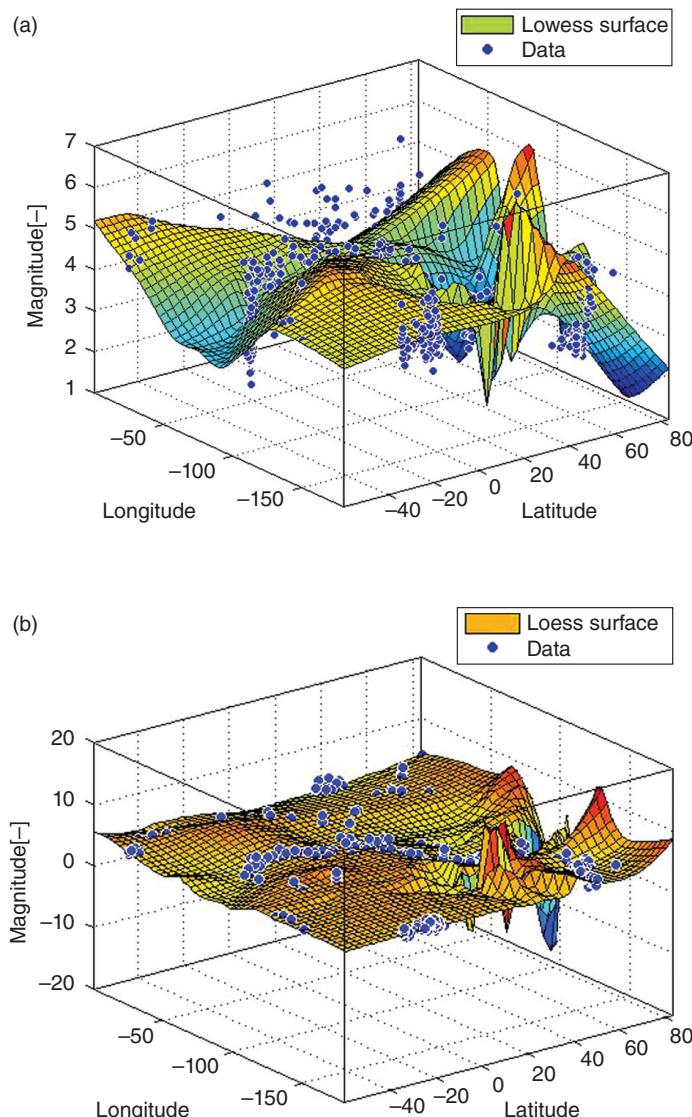
**FIGURE 9.1** Simulation results for the earthquake that happened in the months of January–April in 1973. For both simulations, we have used 653 data points that contain the magnitude of the earthquake collected from different locations. (a) The local regression estimation surface generated by Lowess method and (b) the local regression estimation surface generated by Loess method.



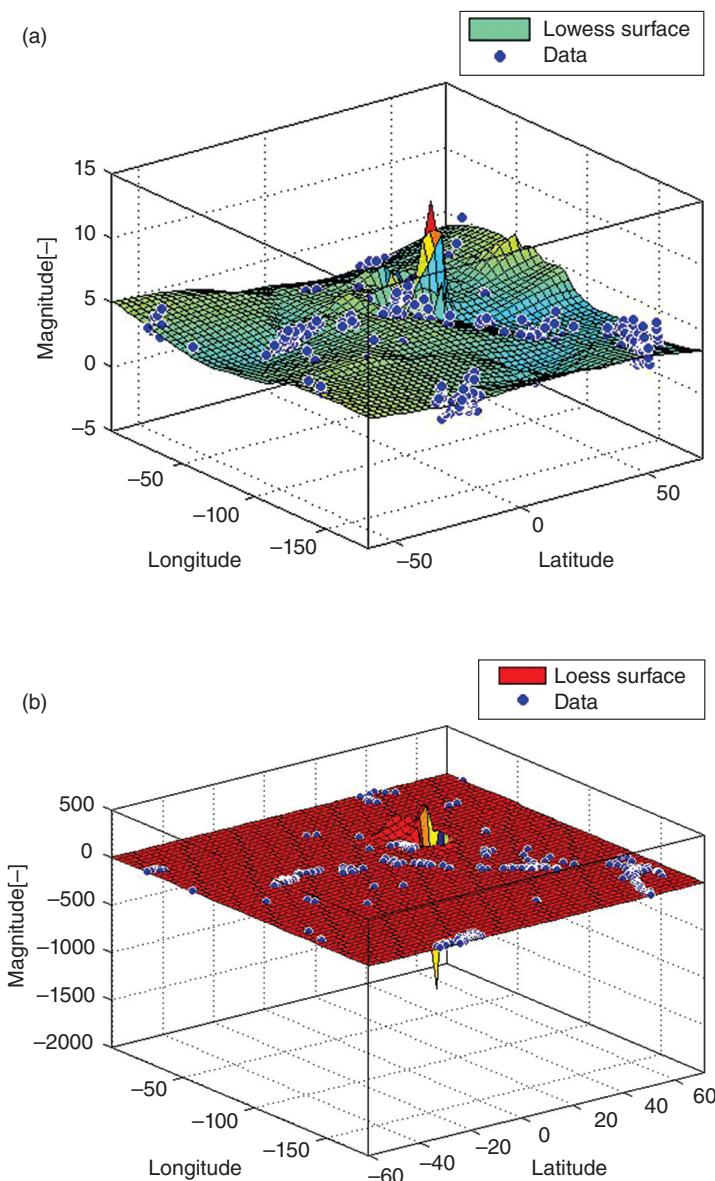
**FIGURE 9.2** Simulation results for the earthquake that happened in the months of January–May in 1979. For both simulations, we have used 1139 data points that contain the magnitude of the earthquake collected from different locations. (a) The local regression estimation surface generated by Lowess method and (b) the local regression estimation surface generated by Loess method.



**FIGURE 9.3** Simulation results for the earthquake that happened in the months of April–June in 1988. For both simulations, we have used 700 data points that contain the magnitude of the earthquake collected from different locations. (a) The local regression estimation surface generated by Lowess method and (b) the local regression estimation surface generated by Loess method.



**FIGURE 9.4** Simulation results for the earthquake that happened in the months of September–October in 1996. For both simulations, we have used 560 data points that contain the magnitude of the earthquake collected from different locations. (a) The local regression estimation surface generated by Lowess method; and (b) the local regression estimation surface generated by Loess method.



**FIGURE 9.5** Simulation results for the earthquake that happened in the months of November–December in 2008. For both simulations, we have used 700 data points that contain the magnitude of the earthquake collected from different locations. (a) The local regression estimation surface generated by Lowess method; and (b) the local regression estimation surface generated by Loess method.

**TABLE 9.1 Results for earthquakes in 1973.**

Latitude (°)	Longitude (°)	Actual magnitude	Estimated magnitude		Relative error (%)	
			Lowess	Loess	Lowess	Loess
56.825	-153.845	4.00	4.024672	4.021482	0.62	0.54
-20.608	-176.426	4.50	4.861483	4.864919	8.03	8.11
62.148	-144.870	3.60	3.653676	3.654828	1.49	1.52
6.408	-79.262	4.60	4.662425	4.724964	1.36	2.72
-26.854	-177.731	4.60	4.872853	4.867820	5.93	5.82
16.744	-86.502	4.80	4.517581	4.498685	5.88	6.28
28.336	-111.934	4.40	4.028162	3.882251	8.45	11.77
33.583	-117.733	3.60	3.824336	3.646300	6.23	1.29
-18.178	-178.158	4.60	4.843991	4.807303	5.30	4.51
37.423	-121.802	3.60	3.685618	3.571850	2.38	0.78

high-frequency data collected for five different financial institutions: JP Morgan Chase, Exxon Mobile Corporation, The Cal Group, International Business Machines Corporation, and MFA Financial, Inc. We work with the stock return value for each of the five financial companies with sample period of time being  $T = 1$  min.

We have applied the “smooth” subroutine from Matlab with different given models as mentioned in the figures. The “smooth” function smooths the given data using a moving average filter, and the results are returned in the new column vector. The default span for the moving average is 5.

**TABLE 9.2 Results for earthquakes in 1979.**

Latitude (°)	Longitude (°)	Actual magnitude	Estimated magnitude		Relative error (%)	
			Lowess	Loess	Lowess	Loess
34.932	-116.683	3.30	3.465456	3.403476	5.01	3.14
-27.829	-66.582	4.90	4.880643	4.884803	0.38	0.31
5.224	-75.801	4.90	4.731234	4.638491	3.44	5.39
-25.829	-70.370	4.80	4.887798	4.894319	1.83	1.96
44.947	-111.869	3.60	3.547208	3.386407	1.47	5.93
-15.140	-173.522	5.10	4.825427	4.865640	5.38	4.60
61.735	-149.881	3.40	3.566559	3.534701	4.88	3.96
60.048	-152.263	3.60	3.803186	3.820450	5.64	6.12
-22.684	-175.375	5.10	4.889236	4.903117	4.13	3.86
-9.961	-74.822	4.90	4.814565	4.877262	1.74	0.46

**TABLE 9.3 Results for earthquakes in 1988.**

Latitude (°)	Longitude (°)	Actual magnitude	Estimated magnitude		Relative error (%)	
			Lowess	Loess	Lowess	Loess
57.002	-142.372	3.90	3.886364	4.132071	0.35	5.95
-21.366	-67.937	4.70	4.712730	5.002906	0.27	6.44
40.445	-124.753	3.30	3.386009	3.303831	2.61	0.12
53.941	-163.268	4.40	4.365138	4.684361	0.79	6.46
51.660	-173.708	4.50	4.602204	4.710451	2.27	4.68
-0.163	-77.793	4.80	4.686259	4.652767	2.37	3.07
-32.382	-69.710	4.50	4.469986	4.168801	0.67	7.36
48.228	-116.387	3.40	3.498185	3.133430	2.89	7.84
57.813	-142.916	4.00	3.882614	4.079561	2.93	1.99
50.966	-177.640	4.70	4.701313	4.603498	0.03	2.05

Under different modeling techniques, the smooth function will produce different smothered data. We have considered the four most popular local regression models applied for modeling the data arising in five different financial institutions, as mentioned above; then we have compared the results provided by the different methods for the same set of data.

In this section, we present the numerical study of the high-frequency data described in Section 9.2.3.1. We have used Matlab as our computational software where different subroutines were applied for the different

**TABLE 9.4 Results for earthquakes in 1996.**

Latitude (°)	Longitude (°)	Actual magnitude	Estimated magnitude		Relative error (%)	
			Lowess	Loess	Lowess	Loess
-18.010	-178.542	4.70	4.563134	4.551168	2.91	3.17
49.011	-128.009	3.80	3.866504	4.163363	1.75	9.56
51.719	-179.526	4.60	4.687235	4.637800	1.89	0.82
-20.619	-178.744	4.60	4.562643	4.554294	0.81	0.99
44.549	-110.527	3.30	3.343416	3.083925	1.32	6.55
60.203	-140.855	3.20	3.178109	3.298102	0.68	3.07
11.114	-86.074	4.30	4.462813	4.651453	3.79	8.17
44.804	-111.276	3.40	3.362970	3.179219	1.09	6.49
-32.744	-179.065	4.60	4.524058	4.580376	1.65	0.42
16.710	-98.731	4.50	4.423162	4.591036	1.71	2.03

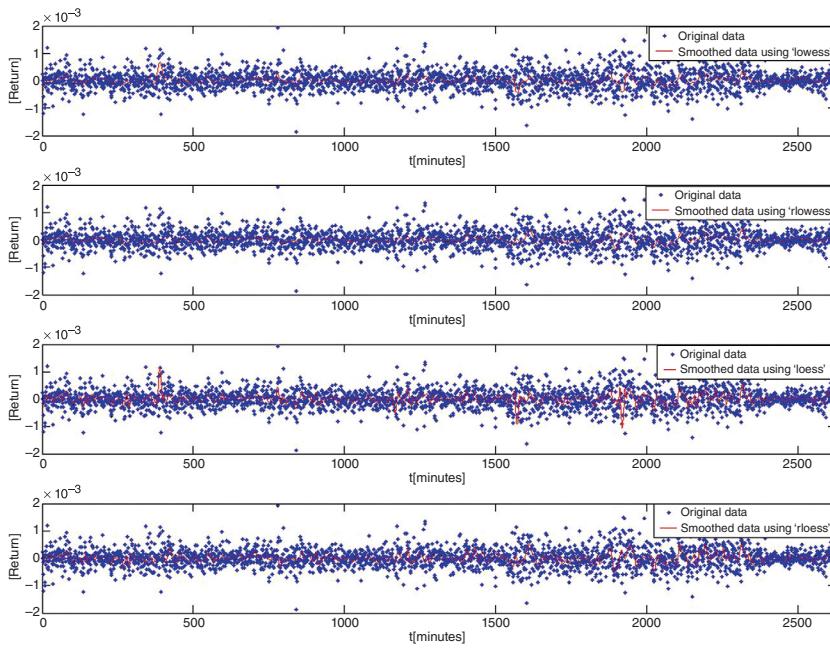
**TABLE 9.5 Results for earthquakes in 2008.**

Latitude (°)	Longitude (°)	Actual magnitude	Estimated magnitude		Relative error (%)	
			Lowess	Loess	Lowess	Loess
40.320	-124.595	3.10	3.199942	3.132174	3.22	1.04
54.054	-162.958	3.60	3.554314	3.377265	1.27	6.19
-32.007	-177.174	4.40	4.533094	4.460330	3.02	1.37
59.323	-152.959	3.30	3.215022	3.077414	2.58	6.75
19.016	-64.781	3.30	3.212336	3.197769	2.66	3.10
61.526	-149.953	3.00	3.089869	3.074387	2.99	1.17
65.889	-140.218	3.10	2.989829	3.074387	3.55	0.83
58.534	-155.296	3.40	3.281178	3.118368	3.49	8.28
-59.452	-23.717	4.50	4.457317	4.726484	0.94	5.03
17.310	-100.770	4.20	4.204719	4.359230	0.11	3.79

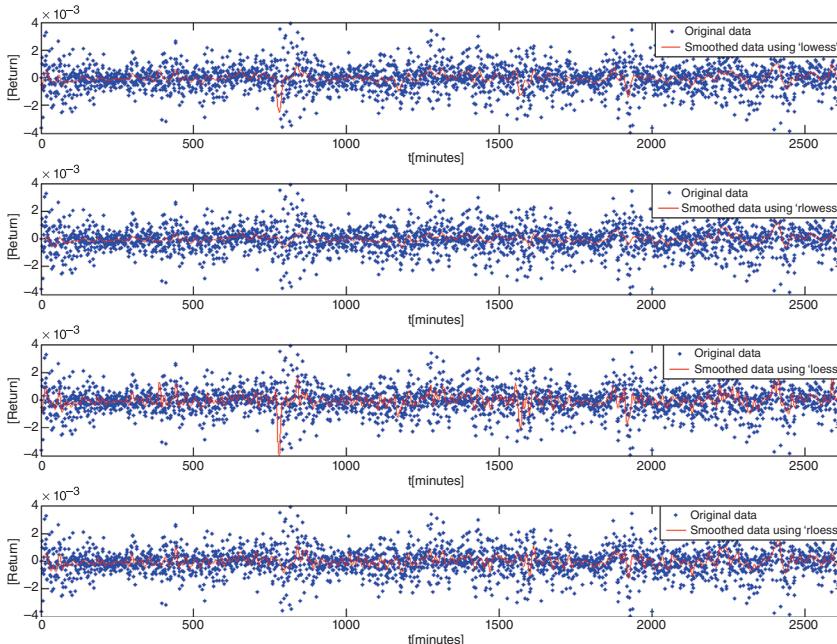
local regression methods. The results are presented for subroutines “lowess,” “rlowess,” “loess,” and “rloess” as different smothering techniques. The results are quite promising even when applied to the random high-frequency data. The data set we used contains highly random data and the feedback obtained from the results is the indication of this method efficiency. Because of the design of robust “r” versions of “lowess” and “loess,” the results seem to be more desirable than the ones obtained when using weighted local regression with first- and second-order polynomial approximations.

The Matlab command  $yy = \text{smooth}(y, \text{"method"})$  smooths the data in  $y$  by using one of the Lowess/Loess methods and the default span. The supported description for each method is listed in the table below.

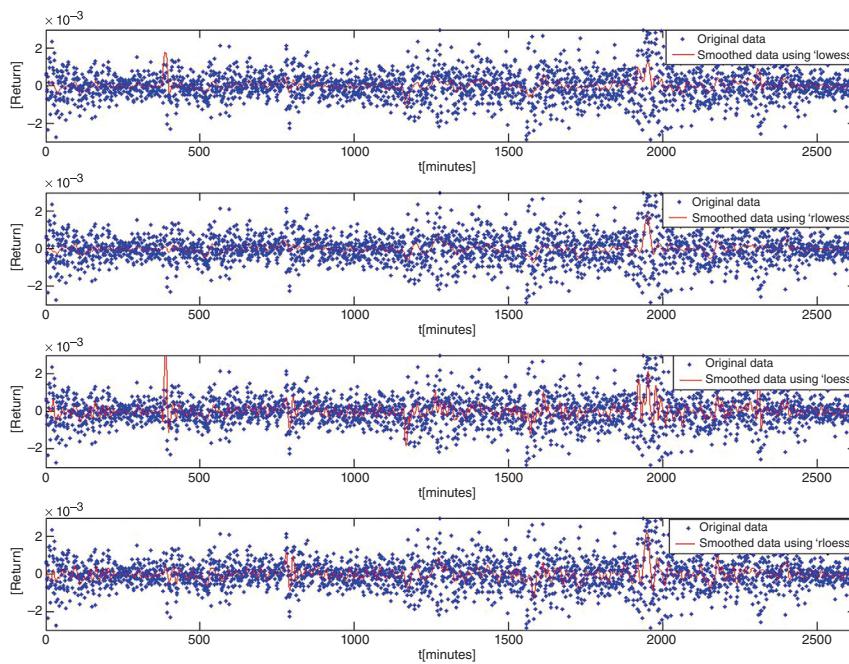
Method	Description
“lowess”	Local regression using weighted linear least squares and a first-degree polynomial model.
“loess”	Local regression using weighted linear least squares and a second-degree polynomial model.
“rlowess”	A robust version of “lowess” that assigns lower weight to outliers in the regression. The method assigns zero weight to data outside six mean absolute deviations.
“rloess”	A robust version of “loess” that assigns lower weight to outliers in the regression. The method assigns zero weight to data outside six mean absolute deviations.



**FIGURE 9.6** Exxon Mobile Corporation—The solid red line represents the best fit curve with four different methodology using same data.



**FIGURE 9.7** The Cal Group (CAL)—The solid red line represents the best fit curve with four different methodology using same data.



**FIGURE 9.8** JP Morgan Chase—The solid red line represents the best fit curve with four different methodology using same data.

### 9.2.3.3 Highlights and discussions

In Table. 9.6, we have listed all the results from the year-wise data analysis. Sometimes it is surprising to observe that Lowess method has produced a better approximation than its quadratic counterpart (Loess). This can happen because of different reasons, in order to address this question accurately, we need to know sufficient background of the data. If the whole

**TABLE 9.6** Highlights of geophysical data analysis.

Year	Max rel. error (%)		Min. rel. error (%)	
	Lowess	Loess	Lowess	Loess
1973	8.45	11.77	0.62	0.54
1979	5.64	6.12	0.38	0.31
1988	2.93	7.84	0.03	0.12
1996	3.79	9.56	0.81	0.42
2008	3.55	8.28	0.11	0.83

system is more toward steady state then data can behave linearly or most of the data collection tools are linear in nature. As we do not have sufficient background for the data set, it is difficult to explain these results.

## 9.3 Interpolation methods

---

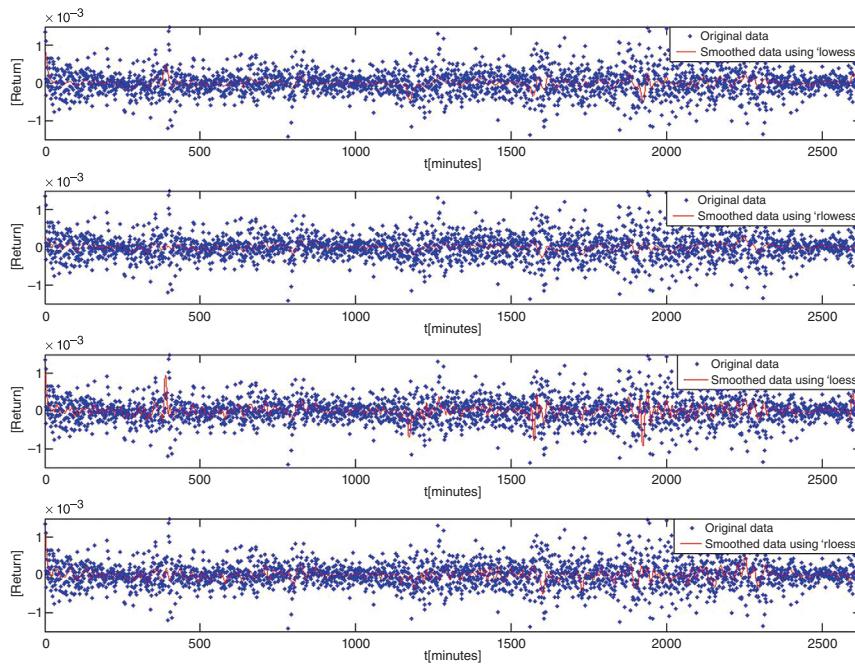
In numerical analysis, interpolation is a process for estimating values that lie within the range of a known discrete set of data points. In engineering and science, one often has a number of data points, obtained by sampling or experimentation, which represents the values of a function for a limited number of values of the independent variable. It is often required to interpolate (i.e., estimate) the value of that function for an intermediate value of the independent variable. This may be achieved by curve fitting or regression analysis.

Another problem that is similar to the interpolation problem is to approximate complicated functions with the means of simple functions. Suppose that we know a formula to evaluate a function but it is too complex to calculate for the given data points. A few known data points from the original function can be used to create an interpolation based on a simpler function. Of course, when a simple function is used to estimate data points from the original, interpolation errors are usually present; however, depending on the problem, domain, and the interpolation method that is used, the gain in simplicity may be of greater value than the resultant loss in accuracy. There is another kind of interpolation in mathematics called “Interpolation of operators.”

In this work, we have implemented five simple interpolation models to our geophysical data set that lists the magnitude of earthquake intensities in different parts of California. Here we will explain the details of our interpolation methods.

### 9.3.1 NEAREST-NEIGHBOR INTERPOLATION

Nearest-neighbor interpolation (also known as proximal interpolation) is a simple method of multivariate interpolation in one or more dimensions. The nearest neighbor algorithm selects the value of the nearest point and does not consider the values of neighboring points at all, yielding a piecewise-constant interpolant. The algorithm is very simple to implement and is commonly used (usually along with mipmapping) in real-time 3D rendering to select color values for a textured surface.



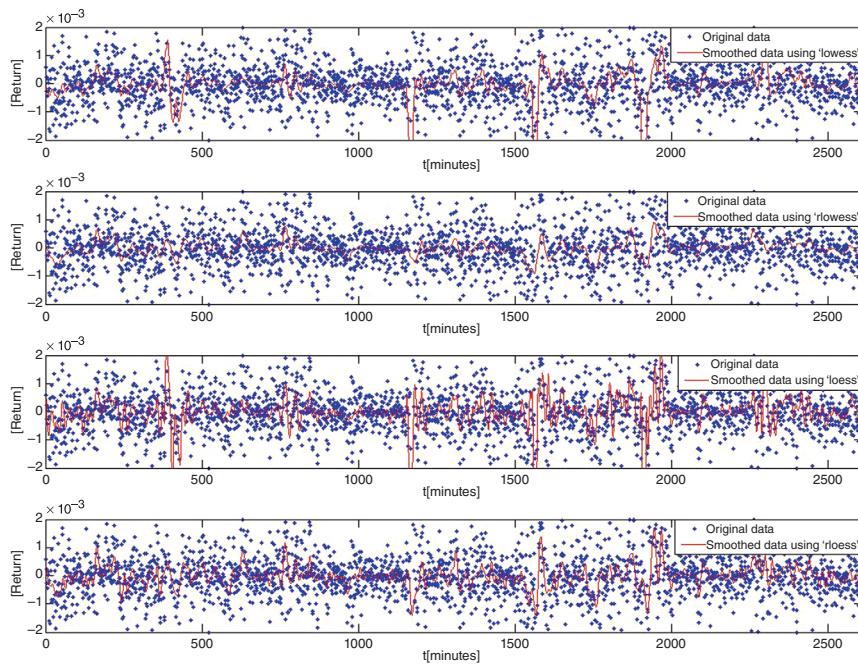
**FIGURE 9.9 International Business Machines**—The solid red line represents the best fit curve with four different methodology using same data.

### 9.3.2 BILINEAR INTERPOLATION

In numerical methods, bilinear interpolation is a special technique that is an extension of regular linear interpolation for interpolation functions of two variables (i.e.,  $x$  and  $y$ ) on a regular two-dimensional grid. The main idea is to perform linear interpolation first in one direction and then again in the other direction. Although each step is linear in the sampled values and in the position, the interpolation as a whole is not linear but rather quadratic in the sample location. Bilinear interpolation is a continuous fast method where one needs to perform only two operations: one is to multiply and the other one to divide. For these methods, bounds are fixed at extremes.

#### 9.3.2.1 Algorithm

Suppose that we want to find the value of the unknown function  $f$  at the point  $P = (x, y)$ . It is assumed that we know the value of  $f$  at the four points  $Q_{11} = (x_1, y_1)$ ,  $Q_{12} = (x_1, y_2)$ ,  $Q_{21} = (x_2, y_1)$ , and  $Q_{22} = (x_2, y_2)$ .



**FIGURE 9.10** MFA Financial, Inc-The solid red line represents the best fit curve with four different methodologies using same data.

We first do linear interpolation in the  $x$ -direction. This yields

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21})$$

where  $R_1 = (x, y_1)$ ,

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22})$$

where  $R_2 = (x, y_2)$ .

We next proceed by interpolating in the  $y$ -direction

$$f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2).$$

This follows the desired estimate of  $f(x, y)$ .

$$\begin{aligned} f(x, y) &\approx \frac{f(Q_{11})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y_2 - y) \\ &+ \frac{f(Q_{21})}{(x_2 - x_1)(y_2 - y_1)} (x - x_1)(y_2 - y) \end{aligned}$$

$$\begin{aligned}
 & + \frac{f(Q_{12})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y - y_1) \\
 & + \frac{f(Q_{22})}{(x_2 - x_1)(y_2 - y_1)} (x - x_1)(y - y_1) \\
 f(x) = & \frac{1}{(x_2 - x_1)(y_2 - y_1)} (f(Q_{11})(x_2 - x)(y_2 - y) + f(Q_{21})(x - x_1)(y_2 - y) \\
 & + f(Q_{12})(x_2 - x)(y - y_1) + f(Q_{22})(x - x_1)(y - y_1))
 \end{aligned}$$

Here we can note that the same result will be achieved if we execute the  $y$ -interpolation first and  $x$ -interpolation second.

### 9.3.2.2 Unit square

If we select the four points where  $f$  is given to be  $(0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$  as the four vertices of the unit square, then the interpolation formula simplifies to

$$f(x, y) \approx f(0, 0)(1 - x)(1 - y) + f(1, 0)x(1 - y) + f(0, 1)(1 - x)y + f(1, 1)xy.$$

### 9.3.2.3 Nonlinear

Contrary to what the name suggests, the bilinear interpolant is not linear; nor is it the product of two linear functions. In other words, the interpolant can be written as

$$b_1 + b_2x + b_3y + b_4xy$$

where

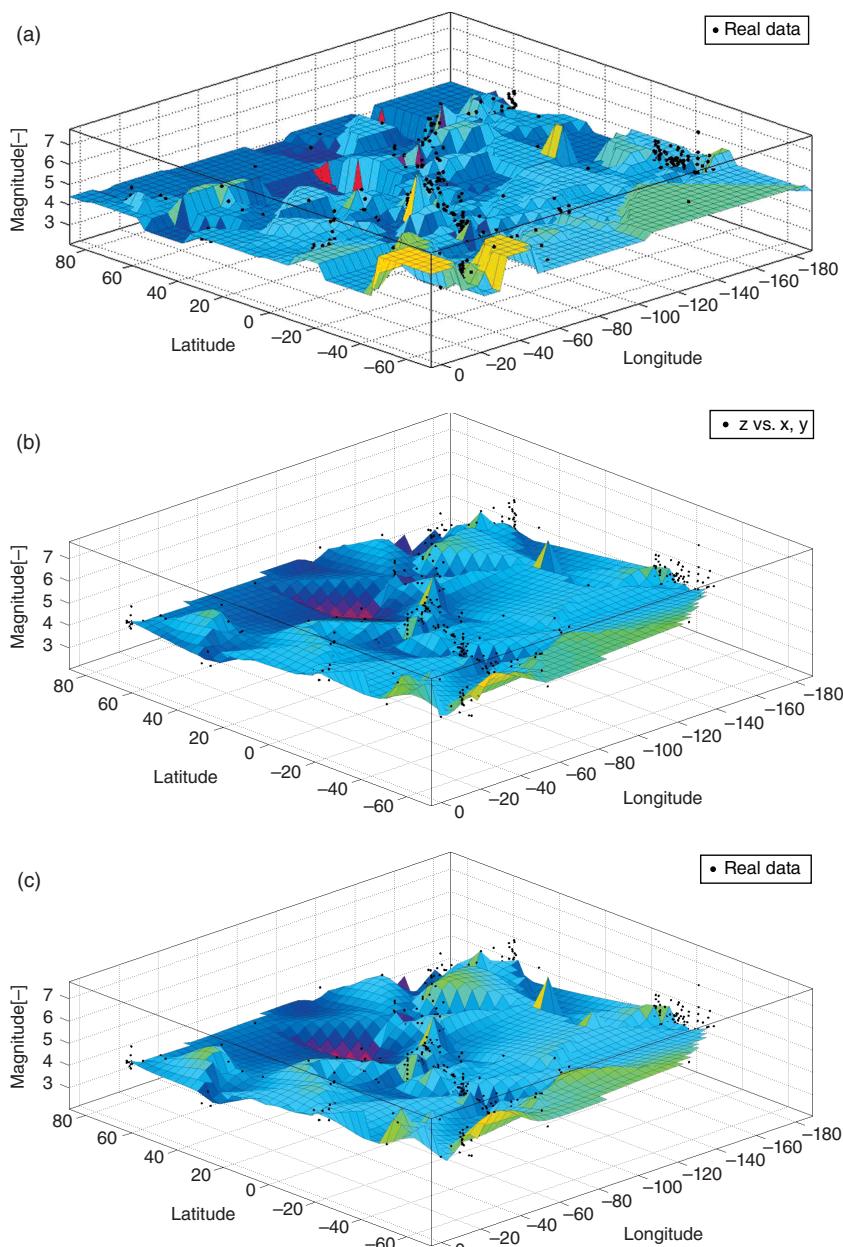
$$b_1 = f(0, 0)$$

$$b_2 = f(1, 0) - f(0, 0)$$

$$b_3 = f(0, 1) - f(0, 0)$$

$$b_4 = f(0, 0) - f(1, 0) - f(0, 1) + f(1, 1)$$

In both cases, the number of constants (four) corresponds to the number of data points where  $f$  is given. The interpolant is linear along lines parallel to either the  $x$  or  $y$  direction, equivalently if  $x$  or  $y$  is set constant. Along any other straight line, the interpolant is quadratic. However, even if the interpolation is not linear in the position ( $x$  and  $y$  variables), it is linear in the amplitude, as it follows from the equations above: all the coefficient  $b_j, j = 1, \dots, 4$ , are proportional to the value of the function  $f$ .



**FIGURE 9.11** Simulation results for the earthquake that happened in the months of January–April in 1973. For all the simulations, we have used 653 data points that contain the magnitude of the earthquake collected from different locations. The Interpolant estimation surface is generated by the following: (a) nearest neighborhood method; (b) bilinear method; (c) bicubic method; (d) biharmonic method; and (e) TPS.

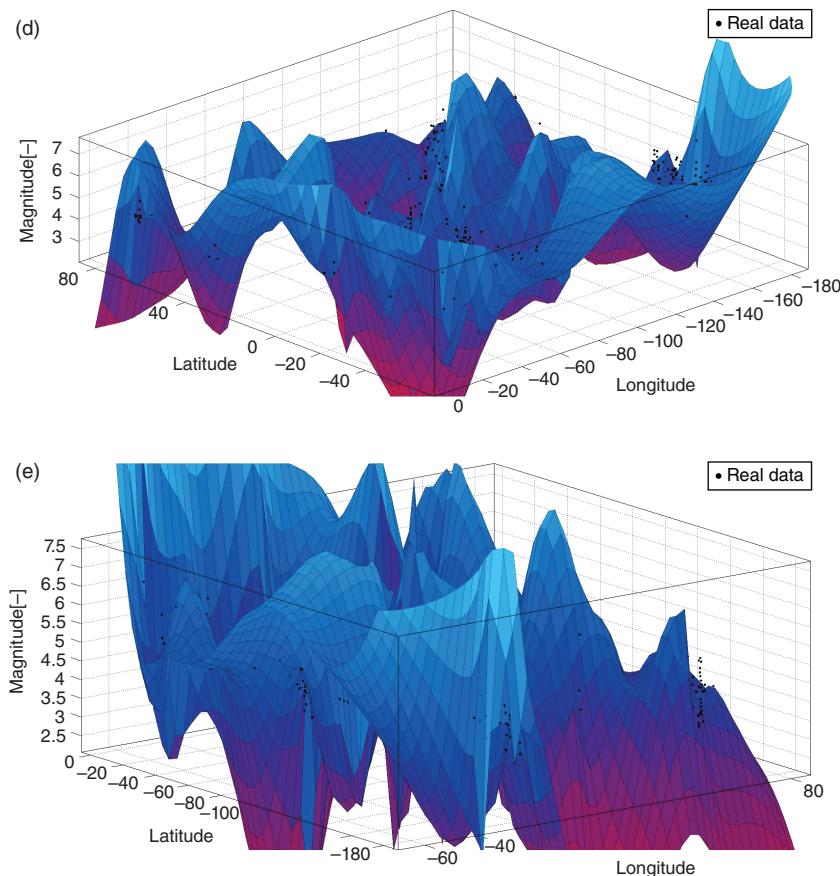


FIGURE 9.11 (Continued)

The result of bilinear interpolation is independent of which axis is interpolated first and which second. If we had first performed the linear interpolation in the  $y$ -direction and then in the  $x$ -direction, the resulting approximation would be the same. The obvious extension of bilinear interpolation to three dimensions is called trilinear interpolation; this process needs no arithmetic operations and is very fast. It has a discontinuity at each value and its bounds are fixed at extreme points.

### 9.3.3 BICUBIC INTERPOLATION

In mathematics, bicubic interpolation is an extension of cubic interpolation for interpolating data points on a two-dimensional regular grid. The

interpolated surface is smoother than corresponding surfaces obtained by bilinear interpolation interpolation or nearest-neighbor interpolation. Bicubic interpolation can be accomplished using Lagrange polynomials, cubic splines, or cubic convolution algorithms.

Suppose that the function values of  $f$ , and its derivatives  $f_x, f_y$ , and  $f_{xy}$  are known at the four corners  $(0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$  of the unit square. The interpolated surface can then be written as:

$$p(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j$$

The interpolation problem consists of determining the 16 coefficients  $a_{ij}$ . Matching  $p(x, y)$  with the function values yields four equations,

$$\begin{aligned} f(0, 0) &= p(0, 0) = a_{00} \\ f(1, 0) &= p(1, 0) = a_{00} + a_{10} + a_{20} + a_{30} \\ f(0, 1) &= p(0, 1) = a_{00} + a_{01} + a_{02} + a_{03} \\ f(1, 1) &= p(1, 1) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} \end{aligned}$$

All the directional coefficients can be determined by the following identities:

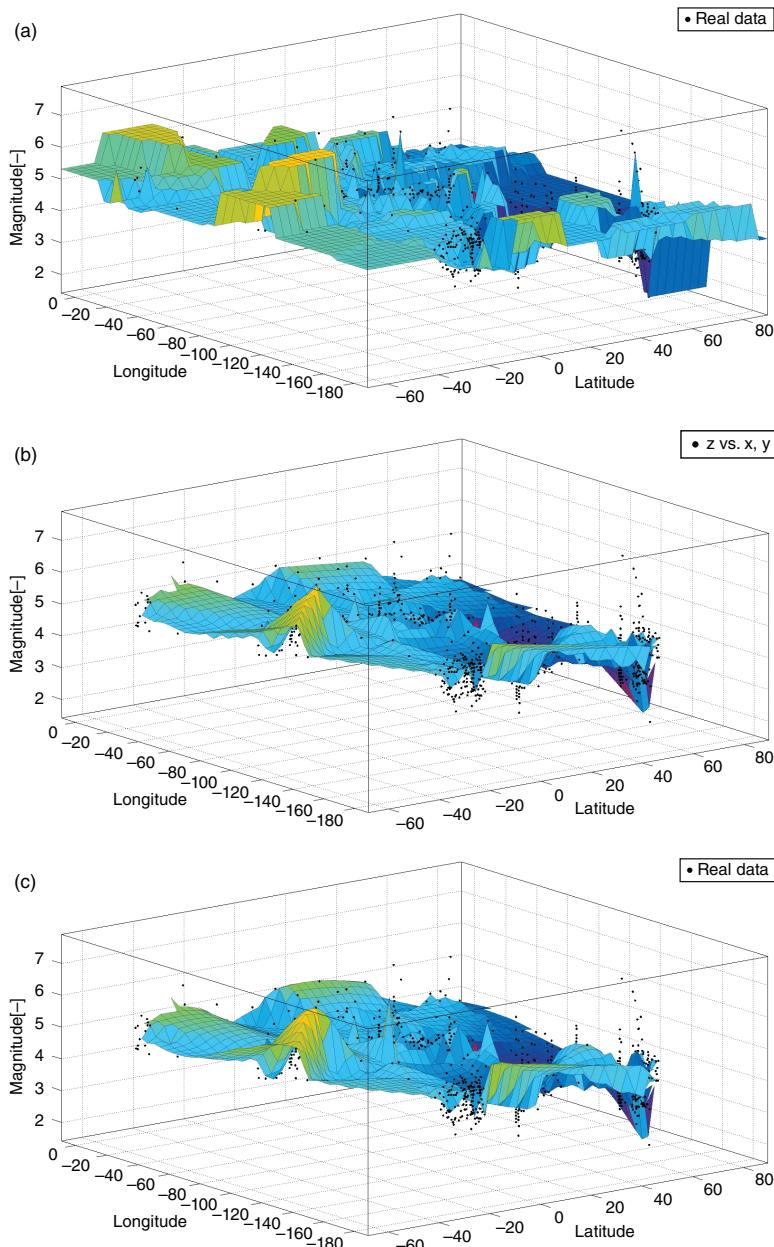
$$\begin{aligned} f_x(x, y) &= p_x(x, y) = \sum_{i=1}^3 \sum_{j=0}^3 a_{ij} i x^{i-1} y^j \\ f_y(x, y) &= p_y(x, y) = \sum_{i=0}^3 \sum_{j=1}^3 a_{ij} j x^i y^{j-1} \\ f_{xy}(x, y) &= p_{xy}(x, y) = \sum_{i=1}^3 \sum_{j=1}^3 a_{ij} i j x^{i-1} y^{j-1}. \end{aligned}$$

This procedure yields a surface  $p(x, y)$  on the unit square  $[0, 1] \times [0, 1]$ , which is continuous and with continuous derivatives. Bicubic interpolation on an arbitrarily sized regular grid can then be accomplished by patching together such bicubic surfaces, ensuring that the derivatives match on the boundaries. If the derivatives are unknown, they are typically approximated by the function values at points neighboring the corners of the unit square, for example, using finite differences. The unknowns coefficients  $a_{ij}$  can be easily found out by solving a linear equation.

### 9.3.4 BIHARMONIC INTERPOLATION

Polynomial splines in  $\mathbb{R}^3$  are functions of the form

$$S(x) = p(x) + \sum_{i=1}^N d_i |x - x_i|^{2v-1}, \quad (9.1)$$



**FIGURE 9.12** Simulation results for the earthquake that happened in the months of January–May in 1979. For all the simulations, we have used 1139 data points that contain the magnitude of the earthquake collected from different locations. The Interpolant estimation surface generated by the following: (a) nearest neighborhood method; (b) bilinear method; (c) bicubic method; (d) biharmonic method; and (e) TPS.

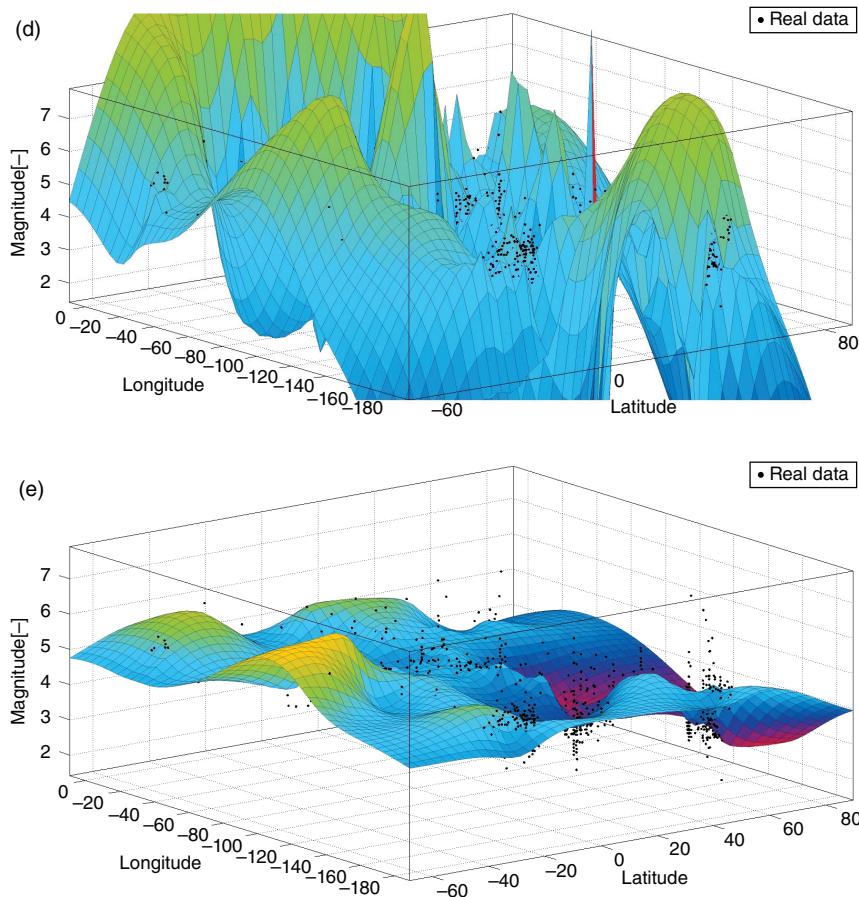
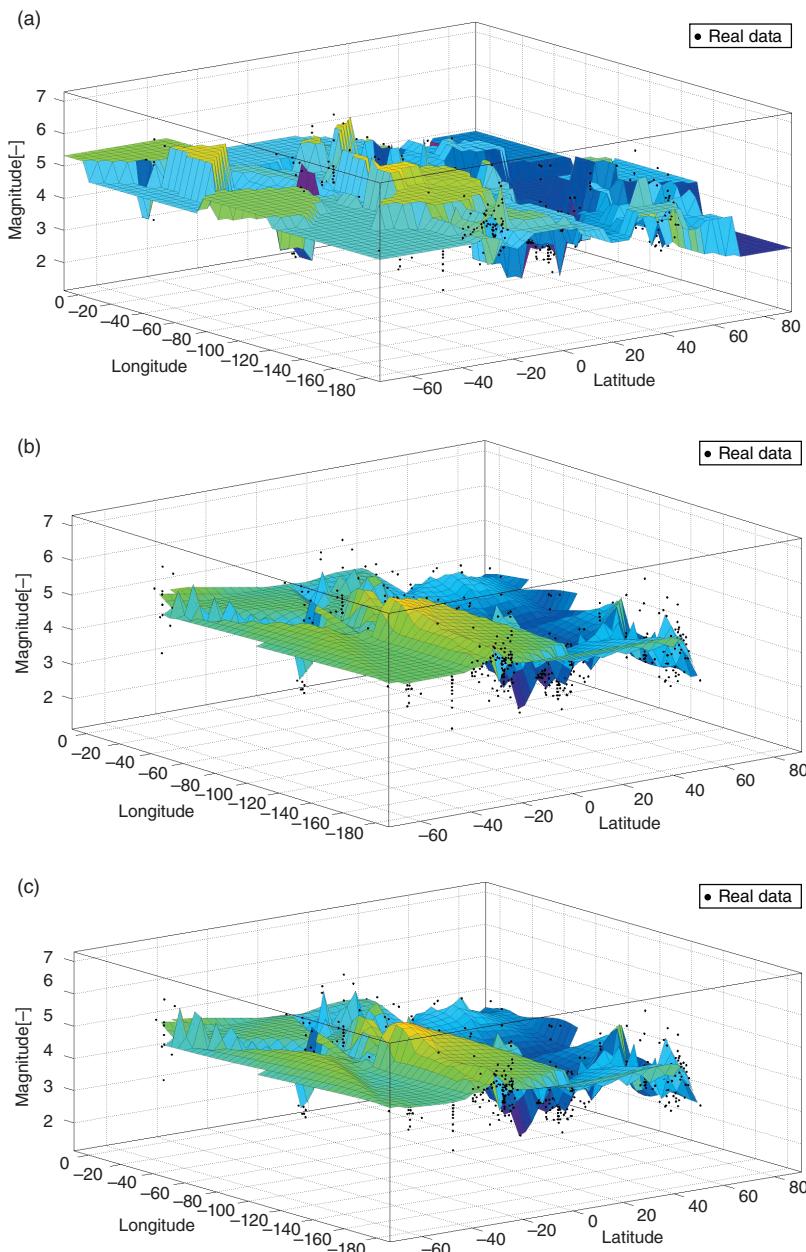


FIGURE 9.12 (Continued)

with  $v$  a positive integer and  $p$  being a polynomial of degree at most equal to  $v$ . One reason for the name of polyharmonic spline is that  $|x|^{2v-1}$  is a multiple of the fundamental solution  $\Phi$  to the distributional equation

$$\Delta^{v+1}\Phi = \delta_0,$$

where the Laplacian is denoted by  $\Delta$  and  $\delta_0$  is the Dirac measure at origin. The main advantage of using polyharmonic splines is their smoothing interpolation property. Focusing on the  $\mathbb{R}^3$  case, given a set of distinct points  $\{x_i\}_{i=1}^N \in \mathbb{R}^3$  unisolvant for  $\pi_v^3$ , and corresponding functional values



**FIGURE 9.13** Simulation results for the earthquake that happened in the months of April–June in 1988. For all the simulations, we have used 700 data points that contain the magnitude of the earthquake collected from different locations. The Interpolant estimation surface generated by the following: (a) nearest neighborhood method; (b) linear method; (c) cubic method; (d) biharmonic method; and (e) TPS.

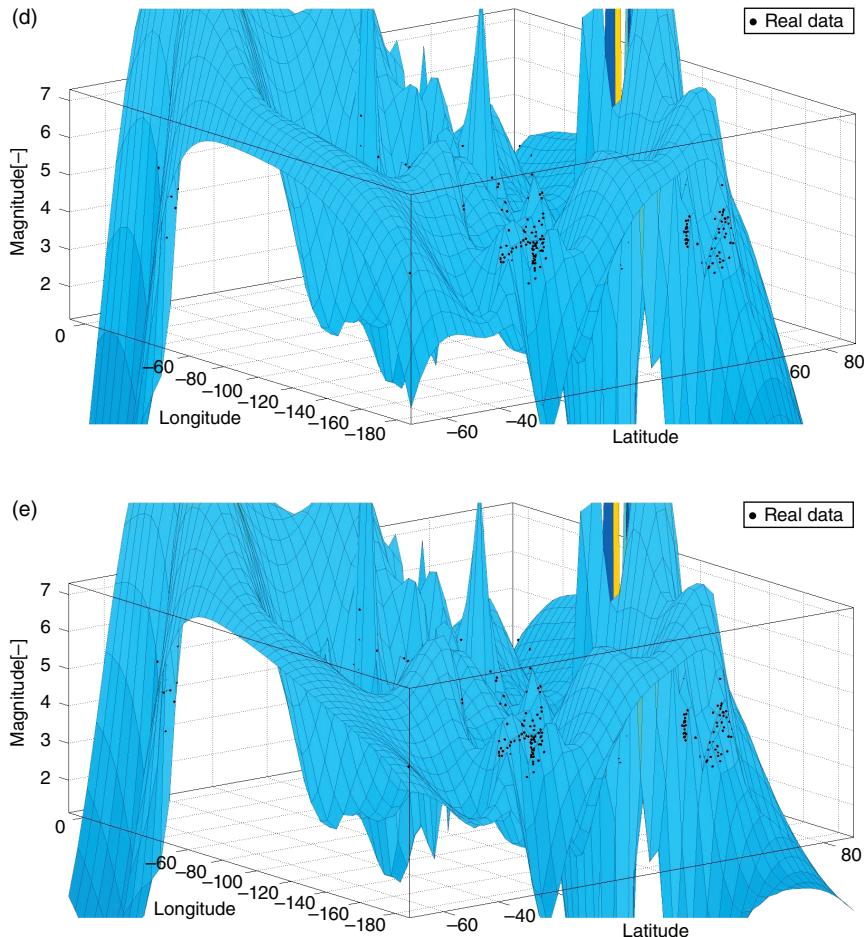


FIGURE 9.13 (Continued)

for  $f_i \in \mathbb{R}$ , there is a unique  $(v + 1)$ -harmonic splines  $S$  of the form (9.1) satisfying the interpolation conditions

$$S(x_i) = f_i, \quad i = 1, 2, \dots, N,$$

and the side conditions

$$\sum_{i=1}^N d_i q(x_i) = 0, \quad \forall q \in \pi_v^3.$$

**Biharmonic** case is a special case for  $v = 1$  in equation (9.1).

**TABLE 9.7 Goodness-of-fit parameters for the data of the year 1973.**

Method	SSE	R-square
“Nearest-neighbor”	0.31	0.999
“Bilinear”	0.31	0.999
“Bicubic”	0.31	0.999
“Biharmonic”	0.31	0.999
“Thin-plate Spline”	0.31	0.999

**TABLE 9.8 Goodness-of-fit parameters for the data of the year 1979.**

Method	SSE	R-square
“Nearest-neighbor”	14.14	0.9825
“Bilinear”	14.14	0.9825
“Bicubic”	14.14	0.9825
“Biharmonic”	14.14	0.9825
“Thin-plate Spline”	251	0.6897

### 9.3.5 THIN PLATE SPLINES

Thin plate splines (TPSs) are an interpolation and smoothing technique, the generalization of splines so that they may be used in two or more dimensions.

#### 9.3.5.1 Physical analogy

The name TPS refers to a physical analogy involving the bending of a thin sheet of metal. Just as the metal has rigidity, the TPS fit resists bending also, implying a penalty involving the smoothness of the fitted surface. In the physical setting, the deflection is in the  $z$  direction, orthogonal to the plane. To apply this idea to the problem of coordinates transformation, one interprets the lifting of the plate as a displacement of the  $x$  or  $y$  coordinates within the plane. In the two-dimensional cases, given a set of  $K$  corresponding points, the TPS warp is described by  $2(K + 3)$  parameters, which include SIX global affine motion parameters and  $2K$  coefficients for correspondences of the control points. These parameters are computed by solving a linear system; in other words, TPS has a closed-form solution.

### 9.3.5.2 Smoothness measure

The TPS arises on considering the square of the second derivative integral—this forms its smoothness measure. In the case where  $x$  is two-dimensional, for interpolation, the TPS fits a mapping function  $f(x)$  between corresponding point-sets  $y_i$  and  $x_i$  that minimizes the following energy function:

$$E = \int \int \left[ \left( \frac{\delta^2 f}{\delta x_1^2} \right)^2 + 2 \left( \frac{\delta^2 f}{\delta x_1 \delta x_2} \right) + \left( \frac{\delta^2 f}{\delta x_2^2} \right)^2 \right] dx_1 dx_2.$$

The smoothing variant, correspondingly, uses a tuning parameter  $\lambda$  to control how nonrigid is allowed for the deformation, balancing the aforementioned criterion with the measure of goodness of fit, thus minimizing

$$\begin{aligned} E_{TPS}(f) = \sum_{i=1}^K \|y_i - f(x_i)\|^2 + \lambda \int \int & \left[ \left( \frac{\delta^2 f}{\delta x_1^2} \right)^2 + 2 \left( \frac{\delta^2 f}{\delta x_1 \delta x_2} \right) \right. \\ & \left. + \left( \frac{\delta^2 f}{\delta x_2^2} \right)^2 \right] dx_1 dx_2. \end{aligned}$$

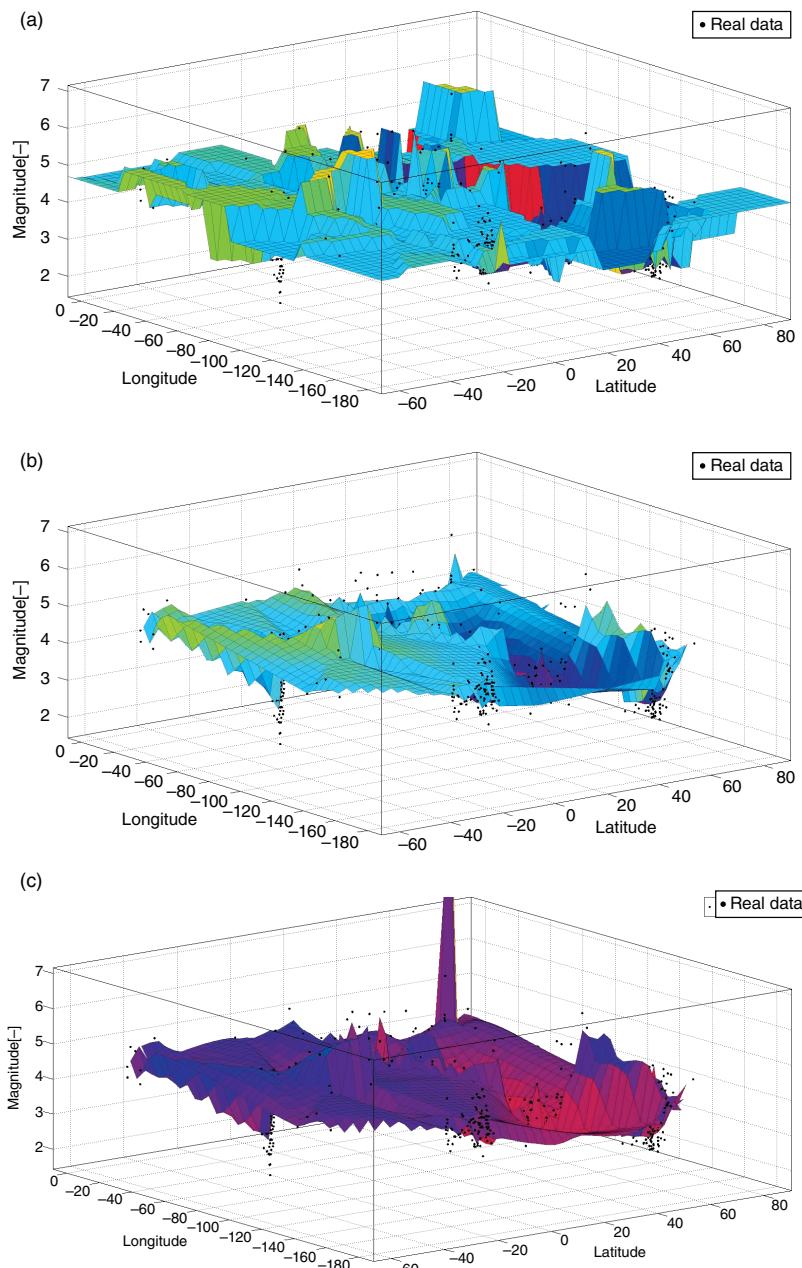
For this variational problem, it can be shown that there exists a unique minimizer  $f$ . The finite element discretization of this variational problem, the method of elastic maps, is used for data mining and nonlinear dimensionality reduction.

### 9.3.5.3 Radial basis function

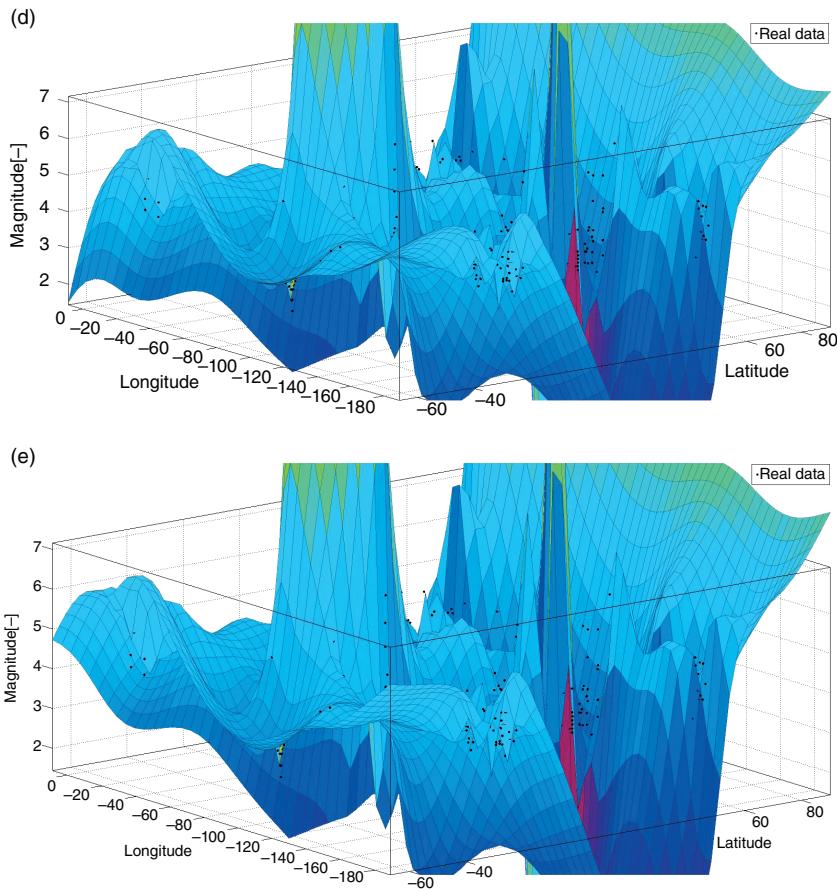
The TPS has a natural representation in terms of radial basis functions. Given a set of control points  $\{w_i, i = 1, 2, \dots, K\}$ , a radial basis function basically defines a spatial mapping that maps any location  $x$  in space to a new location  $f(x)$ , represented by

$$f(x) = \sum_{i=1}^K c_i \psi(\|x - w_i\|)$$

where  $\|\cdot\|$  denotes the usual Euclidean norm and  $\{c_i\}$  is a set of mapping coefficients. The TPS corresponds to the radial basis kernel  $\varphi(r) = r^2 \log r$ .



**FIGURE 9.14** Simulation results for the earthquake that happened in the months of September–October in 1996. For all the simulations, we have used 560 data points that contain the magnitude of the earthquake collected from different locations. The Interpolant estimation surface generated by the following: (a) nearest neighborhood method; (b) bilinear method; (c) bicubic method; (d) biharmonic method; and (e) TPS.



**FIGURE 9.14 (Continued)**

### 9.3.6 NUMERICAL APPLICATIONS

We study the efficiency and accuracy of the different interpolation techniques on the geophysical data set, and we have applied five different interpolation processes to the same data set. Moreover, we calculated the best fit parameters such as SSE and  $R$ -square. These parameters indicate how well fitted the surfaces are with respect to the given data set.

In this numerical exploration of the data set, we used curve fitting toolbox in Matlab to draw all the interpolation surfaces and calculate the parameters of best fit. Results are presented for five randomly selected years where the magnitude for the earthquake in different locations is available.

**TABLE 9.9 Goodness-of-fit parameters for the data of the year 1988.**

Method	SSE	R-square
“Nearest-neighbor”	4.18	0.994
“Bilinear”	4.2	0.993
“Bicubic”	4.2	0.993
“Biharmonic”	4.21	0.993
“Thin-plate Spline”	4.18	0.994

**TABLE 9.10 Goodness-of-fit parameters for the data of the year 1996.**

Method	SSE	R-square
“Nearest-neighbor”	0	1
“Bilinear”	1.059 e-25	1
“Bicubic”	2.078 e-27	1
“Biharmonic”	7.259 e-12	1
“Thin-plate Spline”	1.435 e-10	1

**TABLE 9.11 Goodness-of-fit parameters for the data of the year 2008.**

Method	SSE	R-square
“Nearest-neighbor”	0	1
“Bilinear”	9.263 e-26	1
“Bicubic”	3.253 e-28	1
“Biharmonic”	2.414 e-12	1
“Thin-plate Spline”	3.942 e-11	1

In Figures 9.1, 9.2, 9.3, 9.4, and 9.5 below, we display typical results for the earthquake estimation surface simulated by the five interpolation techniques in some areas of California. We used data from 1973, 1979, 1988, 1996, and 2008 for certain range of months. Real value data were used to draw the estimation surface. The data for these figures was measured in the Western Hemisphere (i.e.,  $-180^\circ$  in longitude). The entire set of earthquakes analyzed (from 1973, 1978, 1988, 1996, and 2008) is presented in Tables 9.1, 9.2, 9.3, 9.4, and 9.5, respectively. Each table lists the parameters of goodness of fit, like SSE and R-square, that were obtained

by using different interpolation methods for five separate years. These parameters are an excellent indicator of how good our surface of fitness is.

The numerical results obtained by implementing our interpolation methods on the geophysical data set indicated that our estimated interpolation surface actually fitted very good with respect to the data set that was used. Looking at the goodness-of-fit parameters, such as SSE and *R*-square, all the data were very accurately and efficiently used to generate the interpolating surface. It is realized that goodness-of-fits are generally dependent on the number of data points used. In most of the cases it will give a near-zero SSE value and an *R*-square value close to 1 that are considered to be an excellent measure of fit. Although when we applied our interpolating techniques on the data set from 1979, the result did not come out very good as we used a larger number of data points. This is only an indication that these methods are very efficient with local data.

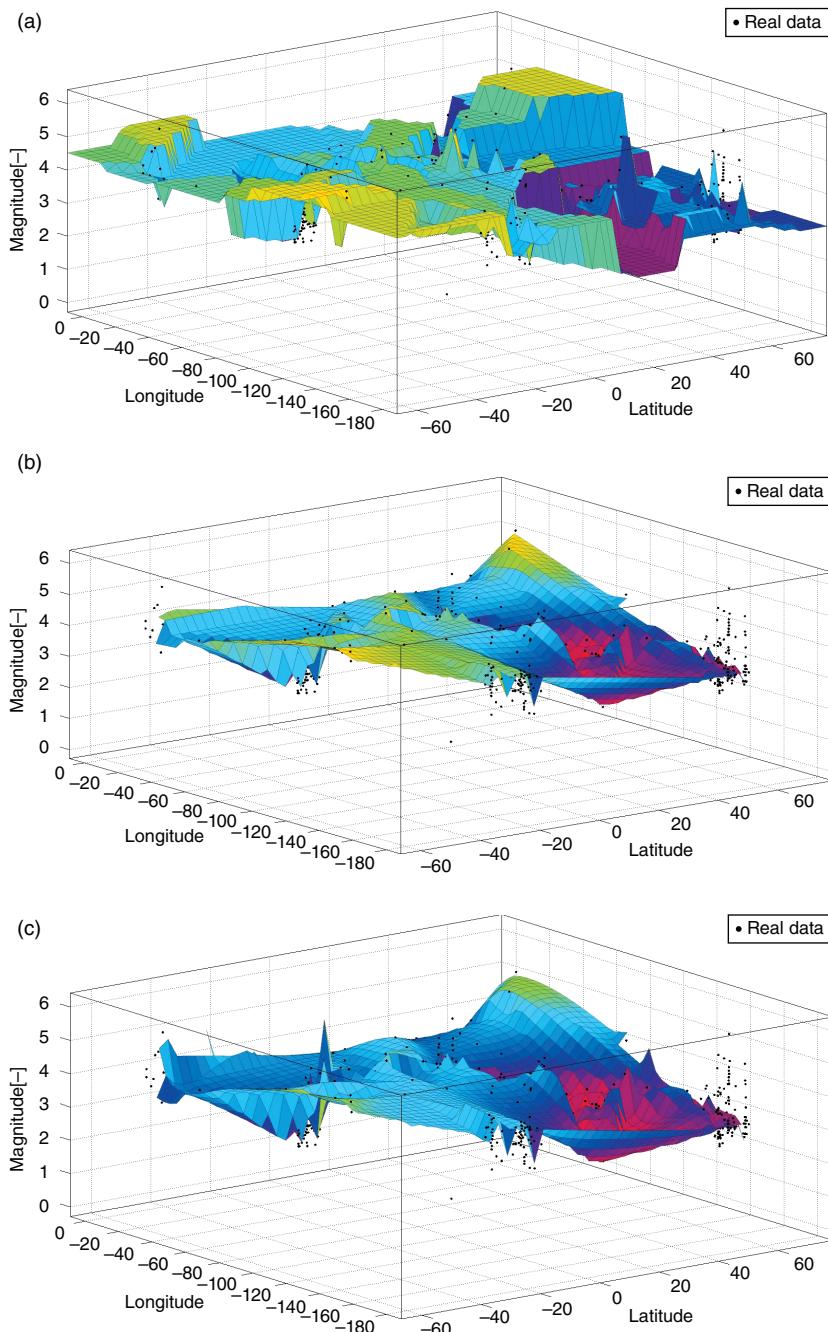
On the application of our methods to data corresponding to the years 1973 and 1988, we noticed that the parameters of goodness of fit came out same by applying different interpolation techniques. So we concluded that these methods are highly data dependent and work best when applied locally. Some more investigation is needed to realize the true connection between these processes and the data set. But overall, we can conclude that the interpolation techniques can serve as a very powerful and robust methodology to process spatial data.

---

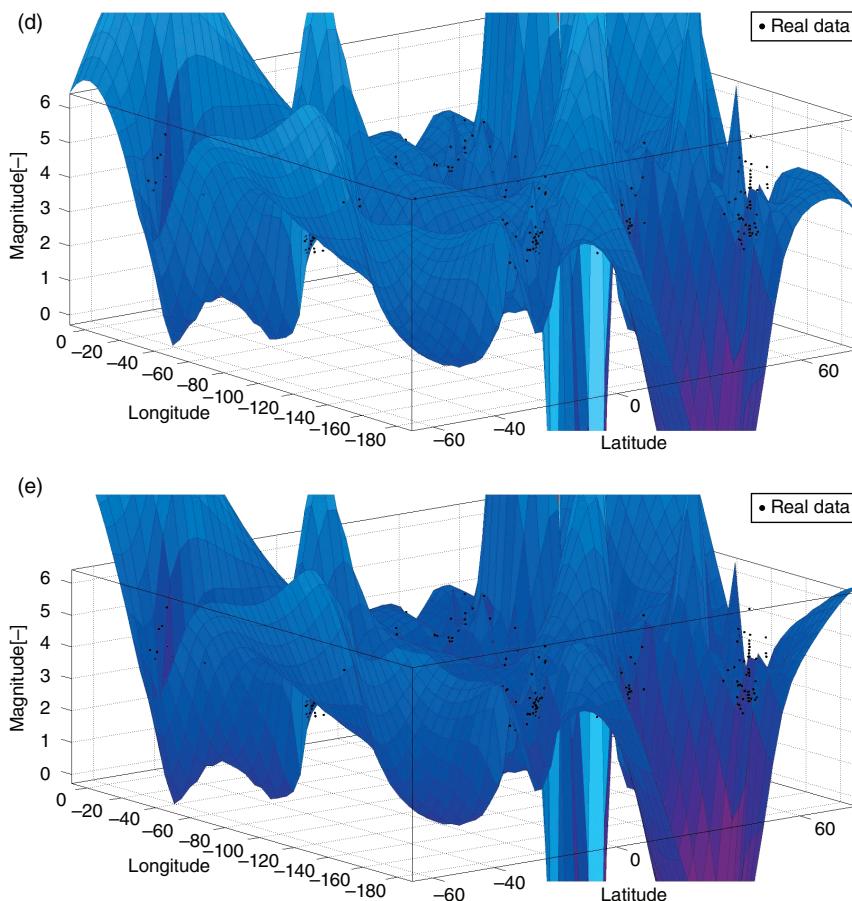
## 9.4 Conclusion

---

As discussed in Section 9.2, the biggest advantage that the Lowess/Loess methods have over many other methods is the fact that they do not require the specification of a function to fit a model over the sampled global data. Instead, an analyst has to provide a smoothing parameter value and the degree of the local polynomial. Moreover, the flexibility of this process makes it ideal for modeling complex processes for which no theoretical model exists. Also, the simplicity to execute the methods makes these processes very popular among the modern era regression methods that fit the general framework of least squares regression, but having a complex deterministic structure. Although they are less obvious than some of the other methods related to linear least squares regression, Lowess/Loess also enjoy most of the benefits generally shared by the other methods, the most important of those is the theory for computing uncertainties for prediction,



**FIGURE 9.15** Simulation results for the earthquake that happened in the months of November–December in 2008. The Interpolant estimation surface generated by the following: (a) nearest neighborhood method; (b) bilinear method; (c) bicubic method; and (d) biharmonic method.



**FIGURE 9.15 (Continued)**

estimation, and calibration. Many other tests and processes used for validation of least square models can also be extended to Lowess/Loess. The major drawback of Lowess/Loess is the inefficient use of data compared with other least square methods. Typically they require fairly large, densely sampled data sets to create good models, the reason behind is that the Lowess/Loess relies on the local data structure when performing the local fitting, thus proving less complex data analysis in exchange of increased computational cost. The Lowess/Loess methods do not produce a regression function that is represented by a mathematical formula, what may be a disadvantage: At times it can make really difficult to transfer the results of an analysis to other researchers, to transfer the regression

function to others, they would need the data set and the code for the Lowess/Loess calculations. In nonlinear regression, on the other hand, it is only necessary to write down a functional form in order to provide estimates of the unknown parameters and the estimated uncertainty. On the basis of the application, this could be either a major or minor setback of using Lowess/Loess. In particular, the simple form of Lowess/Loess cannot be applied for mechanistic modeling where fitted parameters specify particular physical properties of the system. Finally, it is worth mentioning the computation cost associated with this procedure, although this should not be a problem in the modern computing environment, unless the data sets being used are very large. Lowess/Loess methods also have a tendency to be affected by the outliers in the data set, like any other least square methods. There is an iterative robust version of Lowess/Loess (see [1]) that can be applied to reduce sensitivity to outliers, but if there are too many extreme outliers, this robust version also fails to produce the desired results. Analyzing earthquake data sets is not always a very easy modeling procedure, as many different factors can be involved in these phenomena. If we analyze the time series data in order to estimate parameters corresponding to some extreme earthquakes, the modeling technique has to be dependent on traditional stochastic procedure. As we performed spatial analysis of the data with a freezing time, the deterministic behavior can be taken into account. We observe that the results were more dependent on the nature of the data and how the different locations (where the magnitude of the earthquake is given) are close to each other, if we consider data for locations that are sparsely located, then the local regression model will not work up to our satisfaction; we have considered the locations that are geographically closer. We conclude that somehow Lowess has been proved to be a better estimator than the other process. That may be due to some data trends in the data set. Overall, this method has been proven to be a better deal to get numerical estimations for spatial analysis.

The high-frequency data arising from financial market were treated with different smoothing techniques and the best curve fitted provided a very good estimation to the data. The robust version of the weighted local regression technique was much more desirable than its original version. The current work shows that these modeling methods may be applied to high-frequency data and to individual equity data.

In previous works that were done with these geophysical data sets, Ising-type models, Lévy models, and scale invariance properties were used (see [12, 15] and the references therein) and the authors provide estimations

of a “critical phenomenon” by using the time series data. Similar work has been done in [3, 6].

With our weighted local regression-type model, we fixed the time (in this case the year) and used the magnitude of the earthquake from different locations within the time frame to estimate the magnitude of the earthquake at locations whose data were not used. In other words, our model performed a spatial analysis with the given geophysical data. As an extension of this work, we have applied different interpolation methods to the same geophysical data sets and obtained very promising results. Although these are all deterministic models and in general earthquake data are stochastic, we have plans to somehow merge these deterministic models with a strong Lévy’s model to see any possible modeling approach that can open a perspective to deal with these data in future. Generally, a Lévy process consists of three essential components: (i) deterministic part, (ii) continuous random Brownian part, and (iii) discontinuous jump part. For spatial analysis using geophysical data, the third part does not play a big role so modified the deterministic approach can be considered an efficient one to deal this phenomenon.

To model high-frequency data, we fitted a curve of best fit in the time series data where returns of the stock price were given for every minute for five different financial institutions. Time series modeling (exponential smoothing and ARIMA) is a well-established methodology to address questions of prediction in financial time series. Many works have been done to address these questions, but our aim is to show the usefulness of the Local Regression models with some modification applied to such time-dependent data set. In literature, there are numerous such fits like the one we presented here, but our fit is very appropriate and efficient to apply for local data. As a matter of fact, when stock prices are, in general, locally influenced, this fit will act as better than many others. Overall, we conclude that our approach is a very powerful and easy-to-apply method that produces numerical results with excellent efficiency.

On a final note, we believe that this manuscript deals with a relevant problem: the estimation of parameters associated with major events in geophysics. We believe that the methodology that we present can also be used to estimate and predict parameters associated with major/extreme events in econophysics, for example, phase transition. The analogy between phase transition and financial modeling can be easily done when considering the original one-dimensional Ising model in phase transition, this simple model has been used in physics to describe the ferromagnetism. Isings

model considers a lattice composed of  $N$  atoms that interact with their immediate lattice neighbors. Likewise, the financial model will consider a lattice composed of  $N$  traders (each trader can also represent a cluster of traders) that interact in a similar manner. In the model for ferromagnetism, a material evidences a net magnetization below a critical parameter, when all spins were arranged on the same direction. In a similar way, in the model for a market crash, the crash happens when all the traders in the market start to sell.

## Acknowledgments

---

The authors thank the anonymous reviewer(s) for the careful reading of the manuscript and the fruitful suggestions.

## References

1. Cleveland, William S. (1979), Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, **74** (368), Pages 829–836.
2. Cleveland, William S., Devlin, Susan J. (1988) ‘Locally-weighted regression: An approach to regression analysis by local fitting’, *Journal of the American Statistical Association*, **83** (403), Pages 596–610.
3. Fonseca R. C. B., Figueiredo A., Castro M. T., Mendes F. M. (2013), Generalized Ornstein–Uhlenbeck process by Doob’s theorem and the time evolution of financial prices, *Physica A: Statistical Mechanics and Its Applications*, **392**, Pages 1671–1680.
4. Fox, J. (2000a), Nonparametric simple regression: Smoothing scatterplots. *Sage*, Thousand Oaks, CA.
5. Fox, J. (2000b), Multiple and generalized nonparametric regression. *Sage*, Thousand Oaks, CA.
6. Habtemicael, S., SenGupta, I. (2014), Ornstein–Uhlenbeck processes for geophysical data analysis, *Physica A: Statistical Mechanics and Its Applications*, **399**, Pages 147–156.
7. Hainzl, S., Ogata, Y. (2005), Detecting fluid signals in seismicity data through statistical earthquake modeling, *Journal of Geophysical Research: Solid Earth*, **110**, B05S07.

8. Ising, E. (1925), Beitrag zur Theorie des Ferromagnetismus, *Zeitschrift fur Physik*, **31**, Pages 253–258.
9. Johansen, A., Sornette, D. (1997), Large financial crashes, *Physica A*, **245** (N3-4), Pages 411–422.
10. Johansen, A., Ledoit, O., Sornette, D. (2000), Crashes as critical points, *International Journal of Theoretical and Applied Finance*, **3** (1), Pages 219–255.
11. Mariani, M. C., Basu, K. (2014), Local regression type methods applied to the study of geophysics and high frequency financial data, *Physica A: Statistical Mechanics and Its Applications*, **410**, Pages 609–622.
12. Mariani, M. C., Florescu, I., Sengupta, I., Beccar Varela, M. P., Bezdek, P., Serpa, L. (2013), Lévy models and scale invariance properties applied to geophysics, *Physica A: Statistical Mechanics and Its Applications*, **392**, Pages 824–839.
13. Mantegna, R. N., Stanley, H. E. (1999), *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge.
14. Mariani, M. C., Liu, Y. (2006), A new analysis of intermittence, scale invariance and characteristic scales applied to the behavior of financial indices near a crash, *Physica A*, **367**, Pages 345–352.
15. Mariani, M. C., Bezdek, P., Serpa, L., Florescu, I. (2011), Ising type models applied to geophysics and high frequency market data, *Physica A*, **390**, 23/24, Pages 4396–4402.
16. Nason, G. P., Silberman, B. W. (1994), The discrete wavelet transform in S, *Journal of Computational and Graphical Statistics*, **3**, Pages 163–191.
17. Nason, G. P., Silberman, B. W. (2000), Wavelets for regression and other statistical problems, In *Smoothing and Regression: Approaches, Computation, and Application*, Wiley, New York.
18. Ogata, Y. (1988), Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association*, **83**, Pages 9–27.
19. Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics*, **50**, Pages 379–402.

20. Reasenberg, P. A., Jones L. M. (1989), Earthquake hazard after a mainshock in California, *Science*, **243**, Pages 1173–1176.
21. Stein, R. S. (1999), The role of stress transfer in earthquake occurrence, *Nature*, **402**, Pages 605–609.
22. Toda, S., Stein, R. S., Reasenberg, P. A., Dieterich, J. H., Yoshida, A. (1998), Stress transferred by the 1995  $M_w = 6.9$  Kobe, Japan, shock: Effect on aftershocks and future earthquake probabilities, *Journal of Geophysical Research: Solid Earth*, **103**, Pages 24543–24565.
23. The USGS Earthquake Magnitude Working Group, USGS Earthquake Magnitude Policy, USGS. 29 March 2010. Web November 26, 2010. <[http://earthquake.usgs.gov/aboutus/docs/020204mag\\_policy.php](http://earthquake.usgs.gov/aboutus/docs/020204mag_policy.php)>

## Chapter Ten

# Study of Volatility Structures in Geophysics and Finance Using Garch Models

Maria C. Mariani<sup>1</sup>, F. Biney<sup>1</sup>, and I. SenGupta<sup>2</sup>

<sup>1</sup>*Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX, USA*

<sup>2</sup>*Department of Mathematics, North Dakota State University, Fargo, ND, USA*

### 10.1 Introduction

With the availability of high-frequency data for financial market analysis, there has been an increase in the studies dealing with the persistence of shocks in both the mean and variance of financial instruments return. Several studies report evidence of persistence (long-memory) behavior in squared returns or empirical volatilities; see Breidt et al. [1], Robinson [2], Shephard [3], Lobato and Savin [4], and Baillie [5]. Similar features have been observed in data from other fields such as physics and geophysics. In physics, the presence of strong autocorrelation in the squares of differences in velocity of the mean wind direction has been explored by Barndorff-Nielsen and Shephard [6].

The pioneering work of Box et al. [7] in the area of autoregressive moving average models paved the way for related work in the area of volatility

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

modeling with the introduction of ARCH and then GARCH models by Engle [8] and Bollerslev [9], respectively. In terms of the statistical framework, these models provide motion dynamics for the dependency in the conditional time variation of the distributional parameters of the mean and variance, in an attempt to capture such phenomena as autocorrelation in returns and squared returns. Extensions to these models have included more sophisticated dynamics such as the threshold model (TGARCH) [Zakoian (1994)] to capture the asymmetry in the news impact, the NGARCH model [10], the EGARCH models [11], the stochastic volatility models [12], the FIGARCH and FIEGARCH models [13, 14], and the long-memory generalized autoregressive conditionally heteroskedastic (LMGARCH) models [2, 15, 16] as well as distributions other than the normal to account for the skewness and excess kurtosis observed in practice such as Student's  $t$  test, Generalized error, Generalized Hyperbolic [17], the Normal inverse gaussian distribution, and Johnson's SU distribution [18].

In the literature, the most popular GARCH model is the GARCH(1,1), where the persistence parameter is less than 1 to ensure strictly and covariance stationarity. It turns out that on modeling using GARCH, most often the persistence parameter is approximately 1 but yet the model does not adequately capture the persistence in volatility. This fact motivated the introduction of the integrated GARCH (IGARCH) model where Bollerslev and Engel [19] allowed for unit persistence in the GARCH model; that is, the persistence was set to 1. The IGARCH model has some structural complication in the sense that its unconditional variance does not exist. Baillie et al. [13] extended the IGARCH to the fractional IGARCH (FIGARCH) by allowing for high persistence (long memory) directly in the conditional variance, while avoiding the complications of IGARCH; that is, they allowed the integration coefficient to vary between [0,1].

Our main interest is to investigate the underlying volatility process in earthquake series, explosive series, high-frequency financial data, and financial indices and examines the applicability of a range of GARCH specifications for modeling volatility of these series to identify similarities and differences in the volatility structure.

This chapter is organized as follows: In Sections 10.2 and 10.3, we review short memory processes (ARMA and GARCH models) and long memory processes (ARFIMA, ARFIMA-GARCH, and FIGARCH), respectively. In Section 10.4, we present methods for detecting and estimating long memory processes such as ADF test, KPSS test, and Whittle approximation. In Section 10.5, we discuss the data and empirical results. Finally, in Section 10.6, we give the conclusion.

## 10.2 Short memory models

In this section, we present a brief introduction of short memory processes. A stationary process has a short memory if its autocorrelation functions decay to zero exponentially fast ( $\rho_k \rightarrow 0$  exponentially fast as  $k \rightarrow \infty$ ) [20] and the autocorrelations are absolutely summable [21]. That is, there exist a  $D > 0$  and  $r < 1$  such that

$$\rho_k \sim D|r|^k \quad (10.1)$$

and

$$\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty \quad (10.2)$$

### 10.2.1 ARMA(p,q) MODEL

The general ARMA(p, q) model is given by

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{i=1}^q \theta_i a_{t-i} \quad (10.3)$$

where  $a_t$  is a white noise series and  $p$  and  $q$  are nonnegative integers. The autoregressive (AR) and moving average (MA) models are special cases of the ARMA( $p, q$ ) model. Using the back-shift operator, the model can be written as

$$(1 - \phi_1 B - \cdots - \phi_p B^p)r_t = \phi_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t \quad (10.4)$$

The polynomial  $1 - \phi_1 B - \cdots - \phi_p B^p$  is the AR polynomial of the model. Similarly,  $1 - \theta_1 B - \cdots - \theta_q B^q$  is the MA polynomial. It is required that there are no common factors between the AR and MA polynomials; otherwise, the order( $p, q$ ) of the model can be reduced [22].

### 10.2.2 GARCH(p,q) MODEL

The standard GARCH model [23] may be written as:

$$\begin{aligned} r_t &= \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t \epsilon \\ \sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 \end{aligned} \quad (10.5)$$

## 298 CHAPTER 10 Study of Volatility Structures in Geophysics and Finance

---

where  $\sigma_t^2$  denoting the conditional variance,  $\omega$  the intercept,  $\varepsilon_t$  is a sequence of iid random variables with mean 0 and variance 1,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_i \geq 0$  and

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1.$$

The GARCH order is defined by (q,p)(ARCH, GARCH), and  $\hat{P}$  is the persistence defined below:

$$\hat{P} = \sum_{j=1}^q \alpha_j + \sum_{j=1}^p \beta_j \quad (10.6)$$

### 10.2.3 IGARCH(1,1) MODEL

From (5), when  $p = q = 1$  and  $\alpha_1 + \beta_1 = 1$  the GARCH(1,1) model becomes

$$\sigma_t^2 = \omega + (1 - \beta_1)\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \quad (10.7)$$

This model, first developed by Engle and Bollerslev [24], is referred to as an Integrated GARCH model or as an IGARCH model. Squared shocks are persistent, so the variance follows a random walk with a drift  $\omega$ . It is known that a GARCH model is analogous to an ARMA model and the IGARCH model where the variance process has a unit root is analogous to an ARIMA model. Because of unit persistence, its unconditional variance does not exist. Nelson [25] showed that while the process in the GARCH(1,1) model is covariance stationary, strictly stationary, and ergodic, the IGARCH(1,1) model is not covariance stationary but is still strictly stationary and ergodic, distinguishing it from the random walk with drift case.

## 10.3 Long memory models

---

In this section, we define long memory model and look at some processes that have long memory. A process has long memory if its autocovariances are not absolutely summable, that is:

$$\sum_{k=-\infty}^{\infty} |\gamma_k| = \infty \quad (10.8)$$

and the autocovariances decay hyperbolically or at a polynomial rate (slow rate), that is:

$$\gamma_k \sim ck^{2d-1} \quad (10.9)$$

as  $k \rightarrow \infty$ , where  $d$  is the long-memory parameter and  $c > 0$  [21].

### 10.3.1 ARFIMA(p,d,q) MODEL

The autoregressive fractionally integrated moving-average (ARFIMA) process is a class of long-memory models introduced by Granger, Joyeux [26] and Hosking [27]. An ARFIMA process  $Y_t$  may be defined by

$$\phi(B)Y_t = \theta(B)(1 - B)^{-d}\varepsilon_t, \quad (10.10)$$

where  $(B)$  is the back shift operator,  $\phi(B) = 1 + \phi_1B + \dots + \phi_pB^p$  and  $\theta(B) = 1 + \theta_1B + \dots + \theta_pB^p$  are the autoregressive and moving-average polynomials, respectively;  $\phi(B)$  and  $\theta(B)$  have no common roots,  $(1 - B)^{-d}$  is a fractional differencing operator defined by the binomial expansion

$$(1 - B)^{-d} = \sum_{j=0}^{\infty} \eta_j(B)^j \quad (10.11)$$

where

$$\eta_j = \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)}, \quad (10.12)$$

with  $d < 1, d \neq 0, -1, -2, \dots$ ,  $\varepsilon_t$  a white noise sequence with finite variance and  $\Gamma$  the gamma function.

### 10.3.2 ARFIMA(p,d,q)-GARCH(r,s)

An  $ARFIMA(p, d, q) - GARCH(r, s)$  process is defined by equation

$$\phi(L)Y_t = \theta(L)(1 - L)^d\varepsilon_t, \quad (10.13)$$

$$\varepsilon_t = \epsilon\sigma_t \quad (10.14)$$

$$\sigma_t^2 = \omega + \sum_{j=1}^r \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, \quad (10.15)$$

where  $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$  are the past observations,

$$\sigma_t^2 = E[Y_t^2] \quad (10.16)$$

the conditional variance of the process  $\{Y_t\}$ , the GARCH coefficients  $\alpha_1, \alpha_2, \dots, \alpha_r$  and  $\beta_1, \beta_2, \dots, \beta_s$  are positive, and satisfy:

$$\sum_{j=1}^r \alpha_j + \sum_{j=1}^s \beta_j < 1, \quad (10.17)$$

$\{\epsilon_t\}$  is a sequence of independent and identically distributed zero mean and unit variance random variables. Although  $\epsilon_t$  is often assumed to be Gaussian, in some cases it may be specified by a  $t$ -distribution, double exponential distribution, or a generalized error distribution, among others. These distributions have a greater flexibility to accommodate a possible heavy tail behavior of some financial time series [28].

### 10.3.3 INTERMEDIATE MEMORY PROCESS

The ARFIMA – GARCH process defined in (13) – (17) has been employed for modeling many financial series exhibiting long-range dependence. Nevertheless, the squares of an ARFIMA-GARCH process have only intermediate memory for  $d \in (0, 1/4)$ . A process is said to have *intermediate memory* if for a large lag  $k$  its auto correlation function (ACF) behaves like  $\rho(k) \sim c|k|^{2d-1}$  with  $d < 0$  and  $c > 0$ . Thus, the ACF decays to zero at an hyperbolic rate but it is absolutely summable [28], that is:

$$\sum_{k=0}^{\infty} |\rho(k)| < \infty \quad (10.18)$$

### 10.3.4 FIGARCH MODEL

The previous section showed that the ARFIMA model can be used directly to model the long memory behavior observed in the financial asset returns. However, sometimes we may want to model the dynamics of the asset returns, together with its volatility. In those situations, the GARCH class models provide viable alternatives for volatility modeling. As illustrated in Section 10.2, two components GARCH models can be used to capture the high persistence in volatility by allowing a highly persistent long-run component but yet the squared GARCH process have only short memory.

This subsection shows how GARCH models can be allowed directly for long memory and high persistence in volatility. From the lag operation representation, the GARCH(p,q) model can be written as

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2 \quad (10.19)$$

where

$$Y_t = \mu + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma_t^2) \quad (10.20)$$

and  $L$  is the lag operator (polynomial),

$$\alpha(L) = \sum_{i=1}^q \alpha_i L^i, \quad \beta(L) = \sum_{j=1}^p \beta_j L^j \quad (10.21)$$

The stationarity of this process is achieved when  $\alpha(L) + \beta(L) < 1$ .

Let  $v_t = \varepsilon_t^2 - \sigma_t^2$ , GARCH(p,q) model can be written as an ARMA(m,q) process in terms of squared residuals as

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t \quad (10.22)$$

with  $m = \max\{p, q\}$ . From this formulation, allowing for the presence of a unit root in  $[1 - \alpha(L) - \beta(L)]$ , Bollerslev and Engel (1996) defined the integrated GARCH(*IGARCH*( $p, q$ ) process:

$$(1 - L)\phi(L)\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t \quad (10.23)$$

where  $\phi(L) = \sum_{i=1}^{m-1} \phi_i L^i$  and as it is of order  $m - 1$ , then the unconditional variance does not exist [25]. To allow for high persistence and long memory in the conditional variance while avoiding the complications of *IGARCH* models, Baillie et al. [13] extended the *IGARCH* case allowing for the integration coefficient to vary between  $[0, 1]$  by using parallel argument from *ARMA*( $p, q$ ) process to *ARFIMA*( $m, d, q$ ) processes. The *FIGARCH*( $p, d, m$ ) process is defined as follows:

$$\phi(L)(1 - L)^d \varepsilon_t^2 = \omega + [1 - \beta(L)]v_t \quad (10.24)$$

where  $\phi(L) = \sum_{i=1}^{m-1} \phi_i L^i$  and is of order  $m - 1$  and all the roots of  $\phi(z) = 0$  and  $\beta(L) = 0$  lie outside the unit circle. When  $d = 0$ , this reduces to the usual *GARCH* model, which has short memory; when  $d = 1$ , this becomes the *IGARCH* model, which has infinite persistence and when  $0 < d < 1$ , the fractionally differenced squared residuals,  $(1 - L)^d \varepsilon_t^2$ , follows a stationary

$ARMA(m, q)$  process. The  $ARMA$  representation of the  $FIGARCH$  process is given by

$$\varepsilon_t^2 = \omega + [1 - \beta(L)][(1 - L)^d \phi(L)]^{-1} v_t \quad (10.25)$$

and it is defined in terms of the conditional variance as

$$\sigma_t^2 = \omega + [1 - (1 - \beta L)^{-1}(1 - L)^d(1 - \phi L)] \varepsilon_t^2. \quad (10.26)$$

Baillie [13] referred to the equation (10.26) model as the fractionally integrated  $GARCH$ , or  $FIGARCH(p, d, m)$  model. When  $0 < d < 1$ , the coefficients in  $\phi(L)$  and  $\beta(L)$  capture the short-run dynamics of volatility, while the fractional difference parameter  $d$  models the long-run characteristics of volatility.

## 10.4 Detection and estimation of long memory

In this section, we discuss some of the methods that are used to test and estimate long memory parameter. First, the unit root tests (ADF and KPSS) are discussed followed by the semiparametric, Whittle estimation of long memory long parameters.

### 10.4.1 AUGMENTED DICKEY–FULLER TEST(ADF TEST)

The Dickey–Fuller test is based on the  $AR(1)$  models

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad (10.27)$$

and

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t. \quad (10.28)$$

where  $\varepsilon_t$  denotes the error term. The null hypothesis is that there is a unit root; that is,  $H_0 : \phi_1 = 1$ , and the alternative  $H_1 : \phi_1 < 1$ , which is stationarity or fractional integration, assuming that  $\phi > -1$ . This is the well-known unit-root testing problem; see Dickey and Fuller [29]. The test statistic is the  $t$  ratio of the least squares (LS) estimate of  $\phi_1$  under the null hypothesis. For (27), the LS method gives

$$\hat{\phi}_1 = \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2}, \quad \hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^T (y_t - \hat{\phi}_1 \hat{y}_{t-1})^2}{T - 1} \quad (10.29)$$

where  $y_0 = 0$  and  $T$  is the sample size. The  $t$  ratio is given by:

$$DF \equiv t \text{ ratio} = \frac{\hat{\phi}_1 - 1}{std(\hat{\phi}_1)} = \frac{\sum_{t=1}^T y_{t-1} \varepsilon_t}{\sigma_\varepsilon \sqrt{\sum_{t=1}^T y_{t-1}^2}}, \quad (10.30)$$

which is commonly referred to as the Dickey-Fuller (DF) test. If  $\{\varepsilon_t\}$  noise series with finite moments of order slightly greater than 2, then the DF statistic converges to a function of the standard Brownian motion as  $T \rightarrow \infty$  (see Chan and Wei [30] and Phillips [31]). If  $\phi_0 = 0$  but (27) is employed, then the resulting  $t$  ratio for testing  $\phi_1 = 1$  will converge to another nonstandard asymptotic distribution. In either case, simulation is used to obtain critical values of the test statistics for selected critical values [32]. If  $\phi_0 \neq 0$  and (28) is used, then the  $t$  ratio for testing  $\phi_1 = 1$  is asymptotically normal. However, large sample sizes are needed for the asymptotic normal distribution to hold.

#### 10.4.2 KPSS TEST

The ADF unit root tests are for the null hypothesis that a time series  $y_t$  is  $I(1)$ . Stationarity tests, on the other hand, are for the null that  $y_t$  is  $I(0)$ . The most commonly used stationarity test, the KPSS test, is due to Kwiatkowski, Phillips, Schmidt, and Shin [33]. The test is based on the equations

$$y_t = \beta D_t + \mu_t \quad (10.31)$$

$$\mu_t = \mu_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

where  $D_t$  contains deterministic components (constant or constant plus time trend),  $u_t$  is  $I(0)$ , may be heteroskedastic and is a pure random walk with innovation variance  $\sigma_\varepsilon^2$ . The null hypothesis that  $y_t$  is  $I(0)$  is formulated as  $H_0 : \sigma_\varepsilon^2 = 0$ , which implies that  $\mu_t$  is a constant. Although not directly apparent, this null hypothesis also implies a unit moving average root in the ARMA representation of  $\Delta y_t$ . The KPSS test statistic is the Lagrange multiplier (LM) or score statistic for testing  $\sigma_\varepsilon^2 = 0$  against the alternative that  $\sigma_\varepsilon^2 > 0$  and is given by

$$KPSS = \frac{(T^{-2} \sum_{t=1}^T \hat{S}_t^2)}{\hat{\lambda}^2} \quad (10.32)$$

**TABLE 10.1 Quantiles of the distribution of the KPSS statistic.**

	Right	Tail	Quantiles	
Distribution	0.90	0.925	0.950	0.975
$\int_0^1 V_1(r) dr$	0.349	0.396	0.446	0.592
$\int_0^2 V_2(r) dr$	0.120	0.133	0.149	0.184

where  $\hat{S}_t = \sum_{j=1}^t \hat{\mu}_j$ ,  $\hat{\mu}_j$  is the residual of a regression of  $y_t$  on  $D_t$  and  $\hat{\lambda}^2$  is a consistent estimate of the long-run variance of  $\mu_t$  using  $\hat{\mu}_t$ . Under the null that  $y_t$  is  $I(0)$ , Kwiatkowski, Phillips, Schmidt, and Shin showed that KPSS converges to a function of standard Brownian motion that depends on the form of the deterministic terms  $D_t$  but not the coefficient values of  $\beta$ . If  $D_t = 1$  then

$$\text{KPSS} \xrightarrow{d} \int_0^1 V_1(r) dr \quad (10.33)$$

where  $V_1(r) = W(r) - rW(1)$  and  $W(r)$  is a standard Brownian motion for  $r \in [0, 1]$ . If  $D_t = (1, t)'$  then

$$\text{KPSS} \xrightarrow{d} \int_0^1 V_2(r) dr \quad (10.34)$$

where  $V_2(r) = W(r) + r(2 - 3r)W(1) + 6r(r^2 - 1) \int_0^1 W(s) d(s)$ . Critical values from the asymptotic distributions (33) and (34) are obtained by simulation methods, and these are summarized in Table 10.1 below.

The stationary test is a one-sided right-tailed test so that we reject the null of stationarity at the  $100\alpha\%$  level if the KPSS test statistic (32) is greater than the  $100(1 - \alpha)\%$  quantile from the appropriate asymptotic distribution (33) or (34).

### 10.4.3 WHITTLE METHOD

This methodology gives approximate maximum-likelihood estimates of the long memory parameters based on the calculation of the periodogram by means of the *fast Fourier transform* (FFT) and the use of the Whittle approximation of the Gaussian log-likelihood function.

Suppose that the sample vector  $Y = (y_1, \dots, y_n)$  is normally distributed with zero mean and variance  $\Gamma_\theta$ . Then, the log-likelihood function divided by the sample size is given by

$$\mathcal{L}(\theta) = -\frac{1}{2n} \log \det \Gamma_\theta - \frac{1}{2n} Y' \Gamma_\theta^{-1} Y. \quad (10.35)$$

where the variance–covariance matrix  $\Gamma_\theta$  may be expressed in terms of the spectral density of the process  $f_\theta(\cdot)$  as follows:

$$(\Gamma_\theta)_{ij} = \gamma_\theta(i - j),$$

where

$$\gamma_\theta(k) = \int_{-\pi}^{\pi} f_\theta(\lambda) \exp(i\lambda k) d\lambda.$$

Two approximations are made to obtain the Whittle method. Since

$$\frac{1}{n} \log \det \Gamma_\theta \longrightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \log[2\pi f_\theta(\lambda)] d\lambda,$$

as  $n \rightarrow \infty$  the first term of (39) is approximated by

$$\frac{1}{2n} \log \det \Gamma_\theta \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \log[2\pi f_\theta(\lambda)] d\lambda.$$

and the second approximated by

$$\frac{1}{2n} Y' \Gamma_\theta^{-1} Y \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\theta(\lambda)} d\lambda,$$

where

$$I(\lambda) = \frac{1}{2\pi n} \left| \sum_{j=1}^n y_j \exp(i\lambda j) \right|^2$$

is the periodogram of the series  $\{y_t\}$ . Thus, the log-likelihood function is approximated, up to a constant, by

$$\mathcal{L}(\theta) = -\frac{1}{4\pi} \left[ \int_{-\pi}^{\pi} \log f_\theta(\lambda) d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\theta(\lambda)} d\lambda \right], \quad (10.36)$$

with a discrete version given by

$$\mathcal{L}(\theta) = -\frac{1}{2n} \left[ \sum_{j=1}^n \log f_\theta(\lambda_j) + \sum_{j=1}^n \frac{I(\lambda_j)}{f_\theta(\lambda_j)} \right]. \quad (10.37)$$

## 10.5 Data collection, analysis, and result

The analysis that follows is focused on the daily prices of the Dow Jones Industrial Average (DJIA). Volatility analysis includes fitting various GARCH specification to study the volatility structure of the return series. The sample data includes daily prices, obtained for the period March 14, 2003, to October 10, 2011, comprising 2164 data points (Yahoo Finance). We concentrated on this duration because we wanted to include a period that involves a crash (September 2008).

### 10.5.1 ANALYSIS ON DOW JONES INDEX (DJIA) RETURNS

The time plot of the series in Figure 10.1 shows that the mean of the series changes over time with maximum value of 14165 on 10-9-2007 and minimum value of 6547 on 3-9-2009. The histogram and normal Q–Q plot indicate that the empirical distribution of the series is nonnormal and skewed to the right. From Figure 10.1 we can see that the log of the prices changes over time, which suggests that the series is nonstationary. By performing the unit root test on the series, we found that the ADF test statistic ( $-2.2801$ ) is higher than the critical value at a 5% significance level ( $-2.86431$ ), indicating that we accept the null hypothesis that there is a unit root in the series. This is also supported by a  $p$ -value of 0.4598.

To eliminate the unit root, we found the first difference in  $\ln(Prices)$ , that is,  $\text{returns} = \ln(\text{Price})_t - \ln(\text{Price})_{t-1}$ , and did the test again. ADF test statistic for the return series is ( $-12.3213$ ) with a  $p$ -value of 0.001, indicating that we reject the null hypothesis of a unit root in the series. Hence we conclude that the returns series is stationary.

From Figure 10.2, the return series appears to be stationary around its mean, the histogram looks symmetric with high kurtosis with heavy tails, and the normal q–q plot indicates nonnormal series with outliers. The autocorrelation and the partial autocorrelation functions (ACF/PACF) for the returns are illustrated in Figure 10.3.

From Figure 10.3, the ACF and PACF showed dependency in the return series that required correlation structure in the conditional mean.

### 10.5.2 MODEL SELECTION AND SPECIFICATION: CONDITIONAL MEAN

From Table 10.2, ARMA(2,2) has the minimum  $AIC(-12908.44)$  and  $AIC_c(-12908.39)$  with four significant parameter estimates shown in Table 10.3.

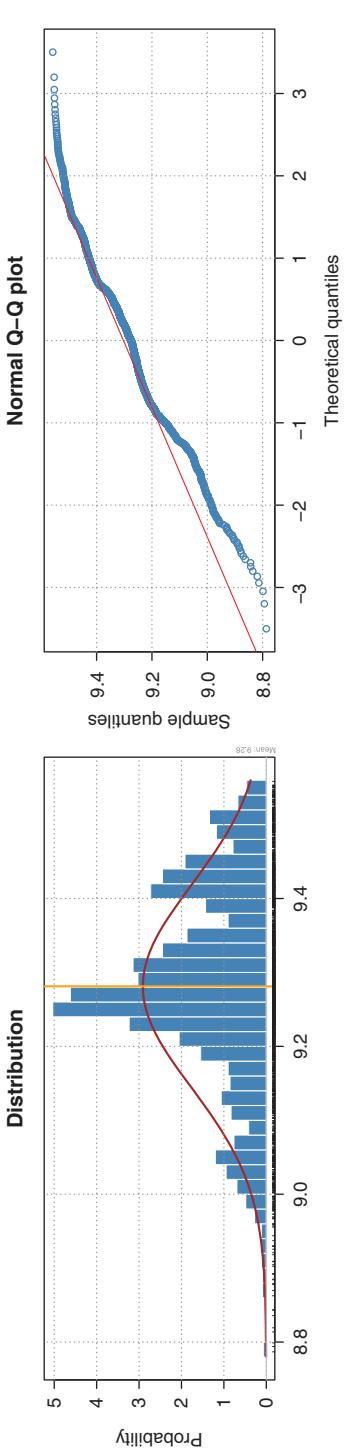
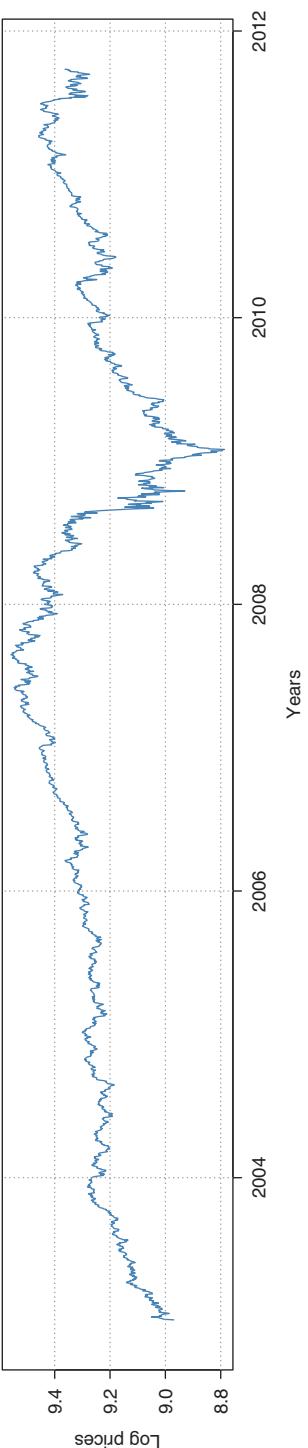


FIGURE 10.1 Log daily price, distribution, and normal Q-Q plot of the DJIA prices from March 14, 2003, to October 10, 2011.



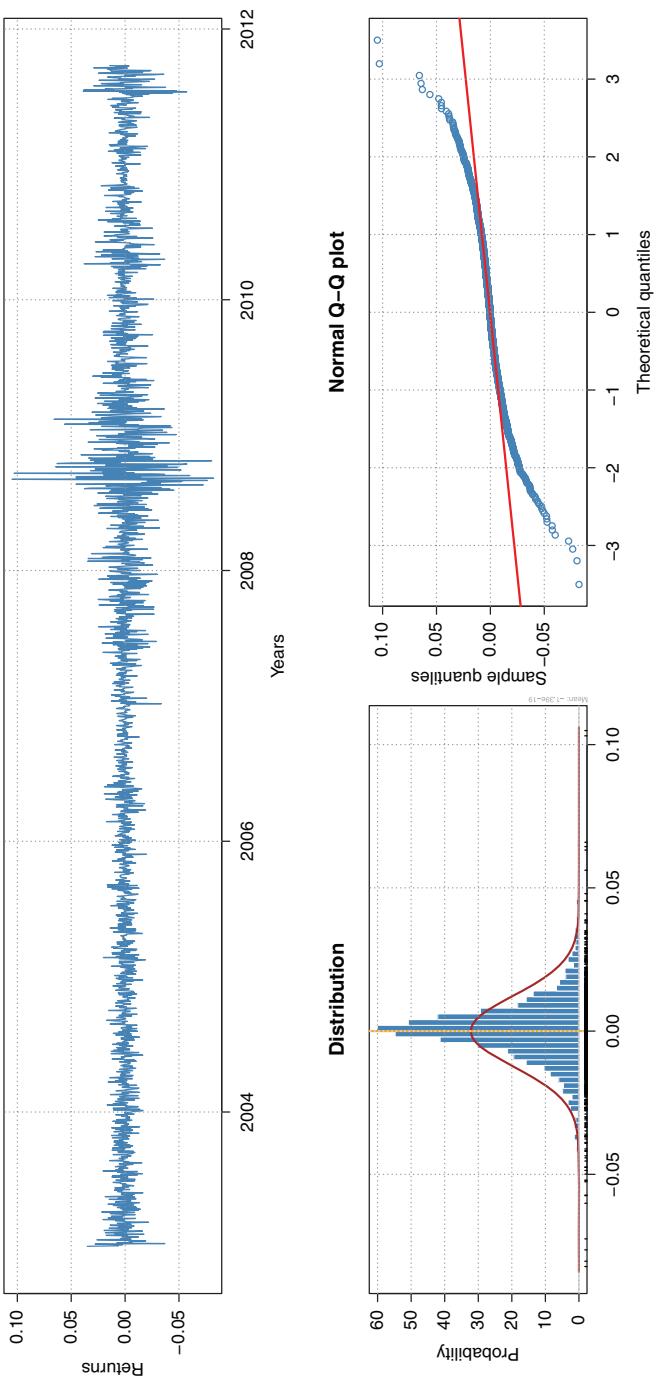


FIGURE 10.2 First difference of the log prices (returns series).

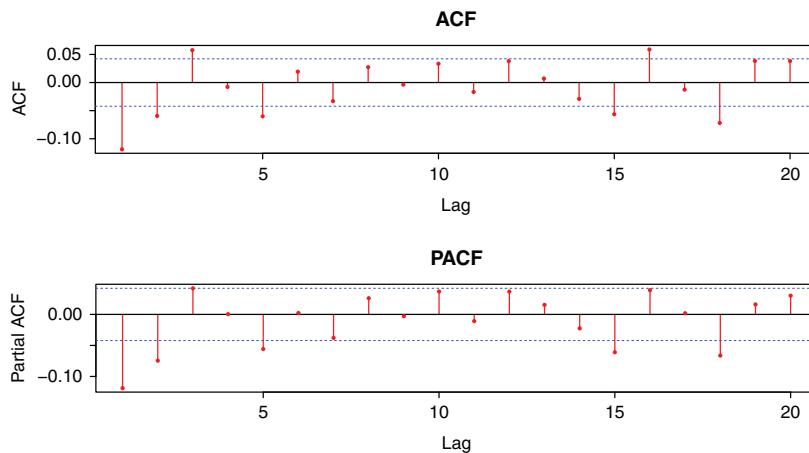


FIGURE 10.3 ACF and PACF of returns.

### 10.5.3 CONDITIONAL MEAN MODEL (RETURNS)

The  $ARMA(p, q)$  model states that the current value of some series  $r_t$  depends linearly on its own previous values plus a combination of current and previous values of a white noise error term  $\varepsilon_t$ .

Our model for the conditional mean of the DJIA returns is  $ARMA(2, 2)$  given by

$$r_t = -0.4035r_{t-1} - 0.3948r_{t-2} + 0.2810\varepsilon_{t-1} + 0.2955\varepsilon_{t-2} + \varepsilon_t \quad (10.38)$$

### 10.5.4 MODEL DIAGNOSTICS: $ARMA(2, 2)$

The time plot of the standardized residuals in Figure 10.4 shows changing volatility and outliers in the series. The ACF of the standardized residuals

TABLE 10.2 Model selection by  $AIC_c$ .

$ARMA(p,q)$	$AIC$	$AIC_c$	$BIC$
$ARMA(1,0)$	-12,896.19	-12,896.18	-12,879.16
$ARMA(0,1)$	-12,900.45	-12,900.44	-12,883.42
$ARMA(1,1)$	-12,901.27	-12,901.25	-12,878.55
$ARMA(2,1)$	-12,906.68	-12,906.65	-12,878.28
$ARMA(2,2)$	-12,908.44	-12,908.39	-12,874.35
$ARMA(3,0)$	-12,908.11	-12,908.08	-12,879.71
$ARMA(0,3)$	-12,906.29	-12,906.26	-12,877.89

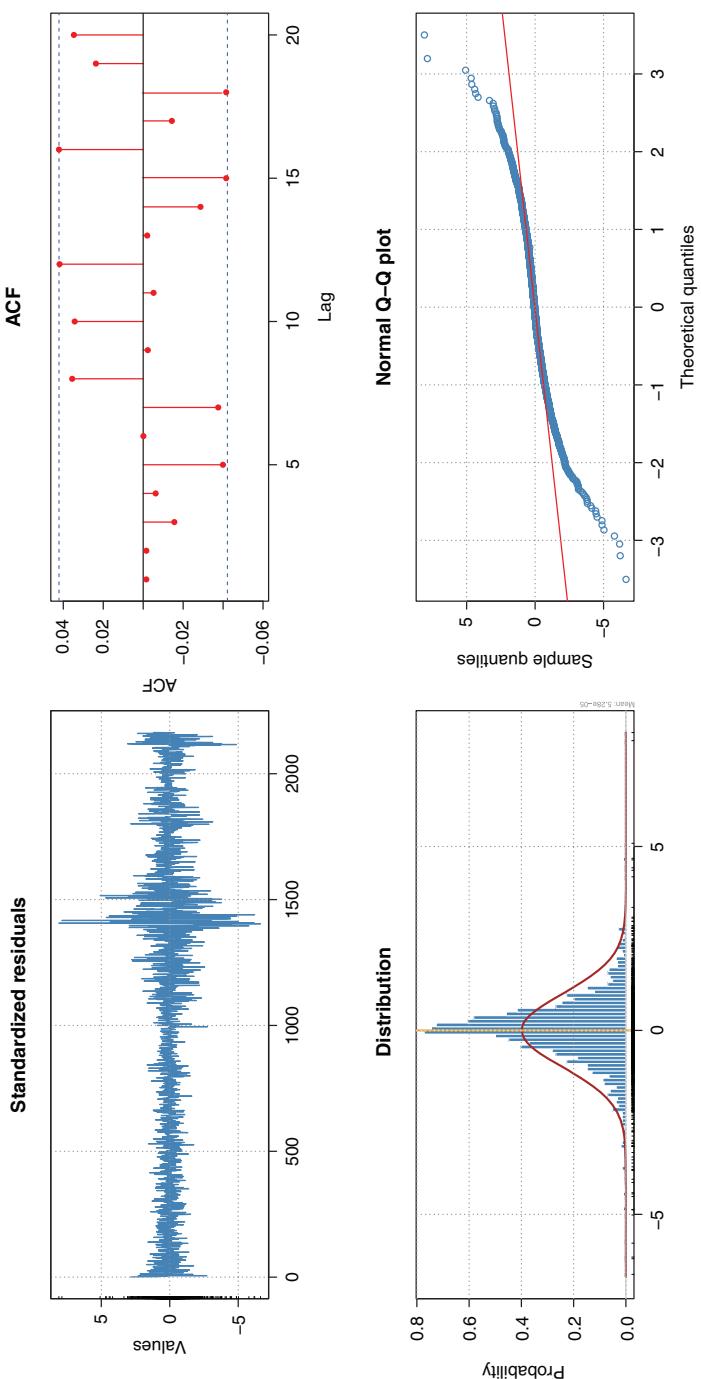


FIGURE 10.4 Time plot, ACF, histogram, and normal q-q plot of standardized residuals.

**TABLE 10.3 ARMA(2,2) model's parameter estimates and standard errors.**

Variable	Coeff	Std. Error	t-Stat	Prob
AR(1)	-0.4035	0.1860	-2.437	0.01479
AR(2)	-0.3948	0.1215	-2.930	0.00339
MA(1)	0.2810	0.127	2.213	0.02700
MA(2)	0.2955	0.1283	2.071	0.03835

AIC : -12912    $\sigma = 0.0122$ .

shows no apparent departure from the model assumptions. The normal q–q plot of the residuals shows departure from normality at the tails due to the outliers and the changing volatility in the return series. The empirical distribution of residuals indicated excess kurtosis (9.34) and skewness (-0.254) with Jarque–Bera higher than 5.99 (7908) at 5% significance. This indicates nonnormality of standardized residuals.

The model appears to fit well except for the fact that a distribution with heavier tails than the normal distribution should be employed. This is our motivation for fitting GARCH to the model residuals.

### 10.5.5 TEST FOR ARCH EFFECT

The analysis continued with a test for an ARCH effect presence in the specified model *ARMA*(2, 2). First, we looked at ACFs of the squared residuals and squared returns. Figure 10.5 shows the ACF of the squared residuals of the fitted model and squared returns, respectively. The ACFs showed dependence in both the squared residuals and squared returns. We also saw that the residuals are not normally distributed, which also suggest

**TABLE 10.4 Statistics of standardized residual.**

Statistics	Value	Statistics	Value
Mean	0.0000	SD	1.000
Median	0.0514	Skew	-0.254
Minimum	-6.632	Kurtosis	9.340
Maximum	8.097	Jarque–Bera	7908(5.99)

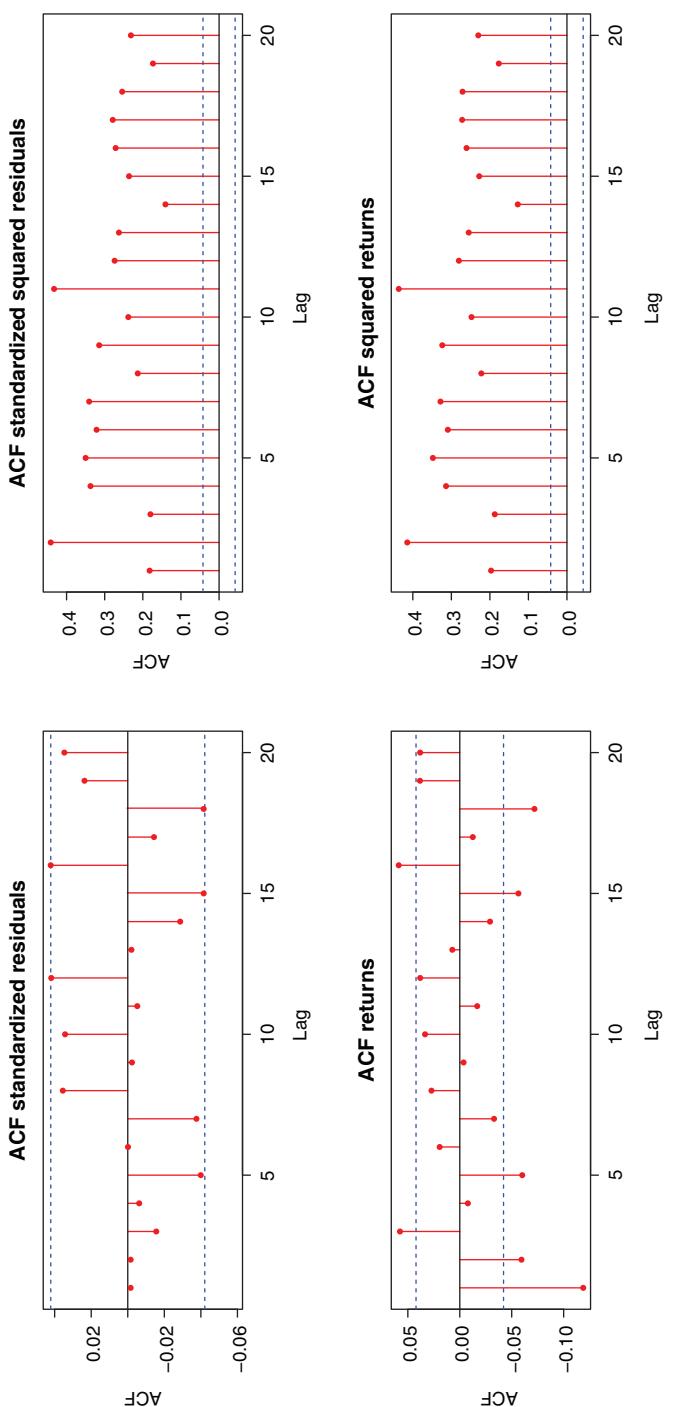


FIGURE 10.5 ACF and PACF of the squared residuals and squared returns.

**TABLE 10.5 Model selection by AIC.**

Model	AIC	$d$
<i>GARCH</i> (1, 1)	-6.545	0
<i>IGARCH</i> (1, 1)	-6.544	1
<i>FIGARCH</i> (0, $d$ , 0)	-6.507	0.215
<i>FIGARCH</i> (1, $d$ , 1)	-6.554	0.602
<i>GARCH</i> (2, 1)	-6.552	0
<i>IGARCH</i> (2, 1)	-6.551	1

the presence of ARCH effect in the series. This is confirmed by the Box-Ljung test statistics X-squared of 1994.54 with 0.00  $p$ -value for the squared residuals, and 1921.53 with 0.00  $p$ -value for the squared returns.

### 10.5.6 MODEL SELECTION AND SPECIFICATION: CONDITIONAL VARIANCE

The model selected by AIC criterior is *FIGARCH*(1,  $d$ , 1) with AIC(-6.554) and the long memory parameter  $d = 0.602$ . Table 10.5 shows the model specification, AIC, and the long memory parameter. The four significant parameter estimates and their standard errors are shown in Table 10.6.

The estimates of the FIGARCH(1,d,1) are  $\omega(0.03816)$ ,  $\phi(0.05652)$ ,  $\beta(0.66672)$ , and  $d(0.60164)$  with respective probabilities of 0.012, 0.035, 0.00, and 0.00. The assumed conditional volatility model for the return series is given by

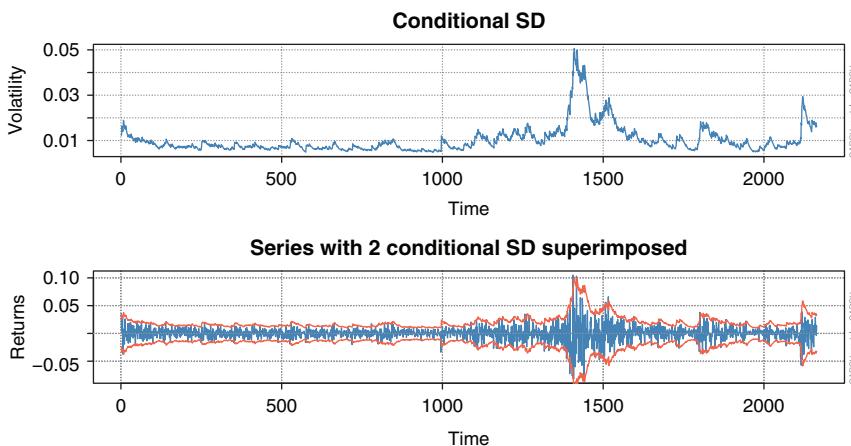
$$r_t = \varepsilon_t \quad \varepsilon_t \sim GED(0, \sigma_t^2) \quad (10.39)$$

$$\sigma_t^2 = 0.038 + [1 - (1 - 0.667L)^{-1}(1 - L)^{0.602}(1 - 0.057L)] \varepsilon_t^2 \quad (10.40)$$

**TABLE 10.6 FIGARCH(1,d,1) model's parameter estimates and standard errors.**

Variable	Coeff	Std. error	t-stat	Prob
omega	0.03816	0.01523	2.505	0.012
ARCH(1)	0.05652	0.02680	2.109	0.035
GARCH(1)	0.66672	0.08811	7.567	0.000
d	0.60164	0.11648	5.165	0.000

AIC : -6.554.



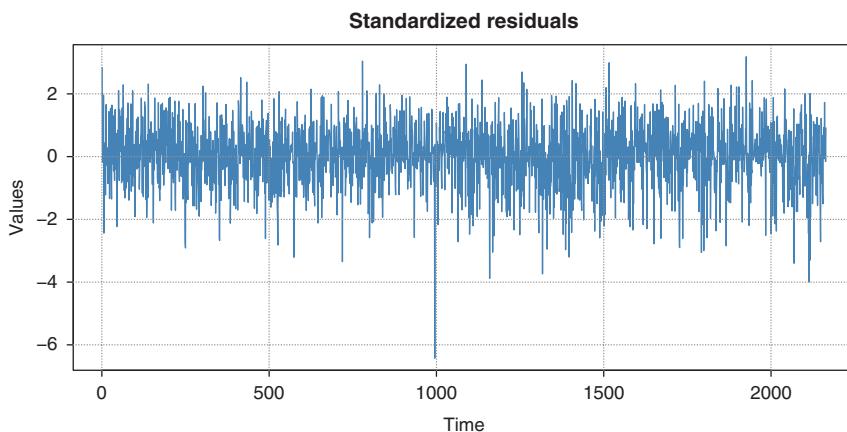
**FIGURE 10.6** Conditional standard deviation superimposed on the returns.

### 10.5.7 STANDARDIZED RESIDUALS TEST

The Jarque–Bera test for normality was 265.50 and the ARCH LM of 13.041 with respective  $p$ -values of 0.000 and 0.4068, indicating that there is no ARCH in the standardized squared residuals. The Ljung–Box statistics of standardized residuals for autocorrelation at lag 15 = 24.03 with a  $p$ -value of 0.065, and the standardized squared residuals at lag 15 = 13.04 with a  $p$ -value of 0.445, which indicates no serial correlation standardized squared residual. These values indicate that the model is adequate in describing the dynamic volatility. Figure 10.6 shows the conditional standard deviation (top) that changes over time and the conditional standard deviation superimposed on the returns (bottom). From the last figure, we can see that the assumed model captures well the volatility of the returns.

### 10.5.8 MODEL DIAGNOSTICS

The time plot of the standardized residuals in Figure 10.7 shows no obvious patterns, but we notice a spike around the 1000th observation. The ACF of the standardized residuals and squared standardized residuals in Figure 10.8 shows no apparent departure from the model assumptions. The histogram and generalized error q–q plot of the standardized residuals show no departure from model assumptions (i.e., the assumed conditional



**FIGURE 10.7 Time plot of standardized residual.**

distribution captured the high kurtosis and the heavy tails of the residuals). This suggests that the residuals are independent generalized error distribution (GED); hence, the model is adequate to describe the changing volatility of the returns.

### 10.5.9 RETURNS AND VARIANCE EQUATION

We finally considered ARMA with FIGARCH variance. The assumed mean model for the DJIA return series was ARMA(2,2) and that of the variance FIGARCH(1,d,1); hence, we first looked at ARMA(2,2)+FIGARCH(1,d,1). Its AIC was  $-6.559$ , but three of the parameter estimates were not significant as shown in Table 10.8, so we considered a simpler model with similar AIC  $-6.558$ , the MA(1)+FIGARCH(1,1), where all the parameter estimates were significant as shown in Table 10.9.

**TABLE 10.7 Standardized residuals test.**

Residual tests	Variable	Test stat	Test value	Prob
Jarque–Bera test	$R$	$\chi^2$	265.50	0.000
Ljung–Box test	$R$	$Q(15)$	24.03	0.065
Ljung–Box test	$R^2$	$Q(15)$	13.04	0.445
LM Arch test	$R$	$TR^2$	13.041	0.4068

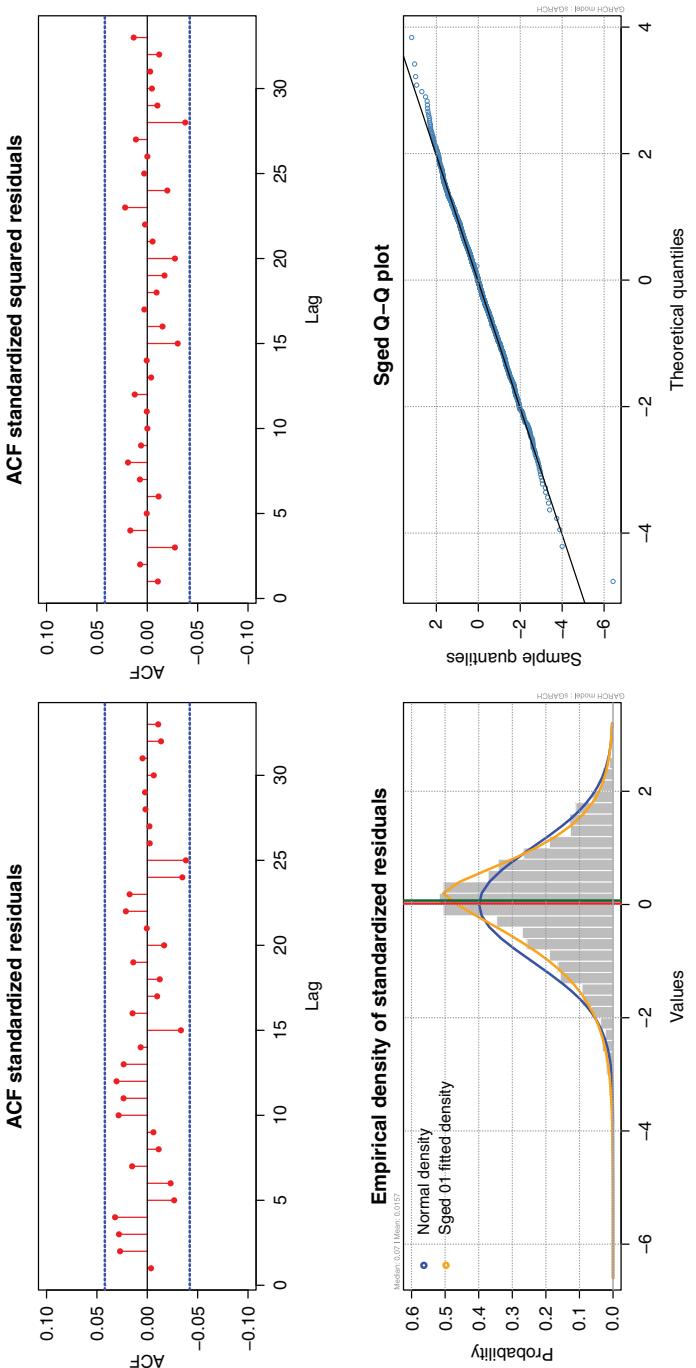


FIGURE 10.8 ACF and distribution of the squared residuals.

**TABLE 10.8 ARMA(2,2)+FIGARCH(1,d,1) model's parameter estimates.**

Variable	Coeff	Std. error	t-stat	Prob
AR(1)	-0.302479	0.17610	-1.718	0.0860
AR(2)	-0.536389	0.20150	-2.662	0.0078
MA(1)	0.250994	0.17683	1.419	0.1559
MA(2)	0.511098	0.20387	2.507	0.0123
omega	0.040077	0.016379	2.447	0.0145
d-Figarch	0.598114	0.11836	5.053	0.0000
ARCH(1)	0.043609	0.052057	0.8377	0.4023
GARCH(1)	0.659513	0.092828	7.105	0.0000

AIC : -6.559.

The assumed conditional mean model with conditional variance is  $MA(1) + FIGARCH(1, d, 1)$  with generalized error given by

$$r_t = -0.0596\epsilon_{t-1} + \varepsilon_t \quad \varepsilon_t \sim GED(0, \sigma_t^2) \quad (10.41)$$

$$\sigma_t^2 = 0.039 + [1 - (1 - 0.664L)^{-1}(1 - L)^{0.597}(1 - 0.0498L)] \varepsilon_t^2 \quad (10.42)$$

The MA(1) parameter estimate is -0.059553 with 0.00 p-values. The estimates of the FIGARCH(1,d,1) are  $\omega(0.039181)$ ,  $\phi(0.049773)$ ,  $\beta(0.663531)$ , and  $d(0.597023)$  with respective probabilities 0.012, 0.040, 0.00, and 0.00. The estimated parameters and their standard error are shown in Table 5.8.

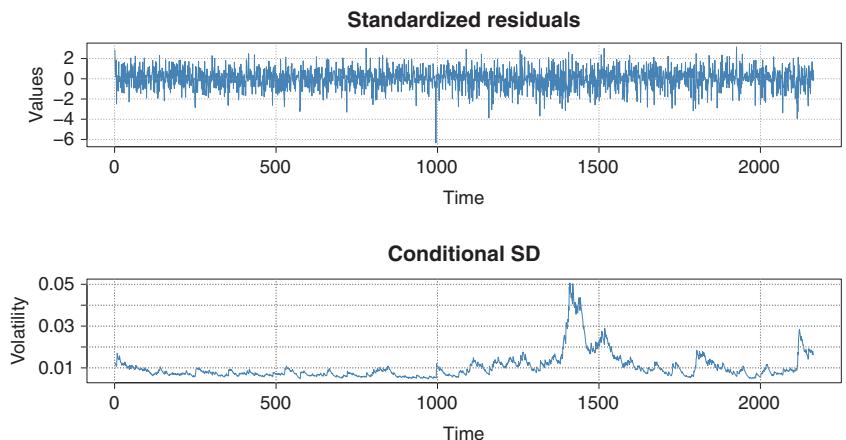
### 10.5.10 STANDARDIZED RESIDUALS TEST

The Jarque–Bera test for normality was 237.12 and the ARCH LM of 7.75041 with respective p-values of 0.000 and 0.6532, which

**TABLE 10.9 ARMA(0,1)+FIGARCH(1,d,1) model's parameter estimates.**

Variable	Coeff	Std. error	t-stat	Prob
MA(1)	-0.059553	0.018602	-3.201	0.0014
omega	0.039181	0.015568	2.517	0.0119
d-Figarch	0.597023	0.11449	5.215	0.0000
ARCH(Phi1)	0.049773	0.052101	0.9553	0.0395
GARCH(Beta1)	0.663531	0.087988	7.541	0.0000

AIC : -6.558.



**FIGURE 10.9** The conditional standard deviation and standard residuals of MA(1)+FIGARCH(1,d,1).

indicate nonnormality and no ARCH in the standardized residuals. The Ljung–Box statistics of standardized residuals for autocorrelation at lag 15 = 13.114 with a  $p$ -value of 0.5176, and the standardized squared residuals at lag 15 = 10.259 with a  $p$ -value of 0.7430, which indicate no serial correlation in both the standardized residuals and the standardized squared residual. These values indicate that the model is adequate in describing the returns series and the dynamic volatility.

### 10.5.11 MODEL DIAGNOSTIC OF CONDITIONAL RETURNS WITH CONDITIONAL VARIANCE

The time plot of the standardized residuals in Figure 10.9 shows no obvious patterns, but we notice a spike closed to the 1000th observation. The ACFs of the standardized residuals and standardized squared residuals show no apparent departure from the model assumptions as shown in Figure 10.10. The histogram and the GED-QQ plot in Figure 10.10 show that the residuals are iid GED errors. Hence, the model seems to be adequate for the data. Consequently, the  $MA(1) + FIGARCH(1, d, 1)$  with *GED* errors model in (41), (42) are adequate for describing the conditional mean of the returns series and prediction at the 5% significance level.

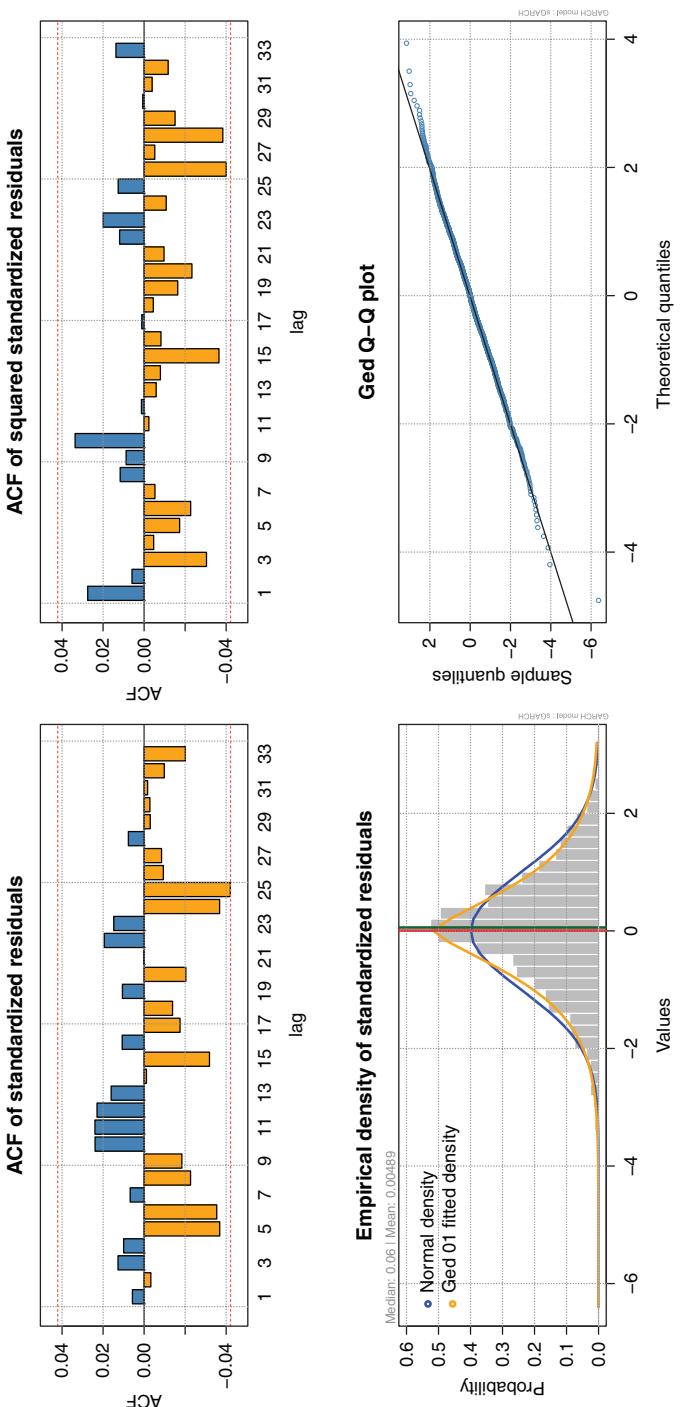


FIGURE 10.10 MA(1)+FIGARCH(1,d,1) standardized residuals distribution.

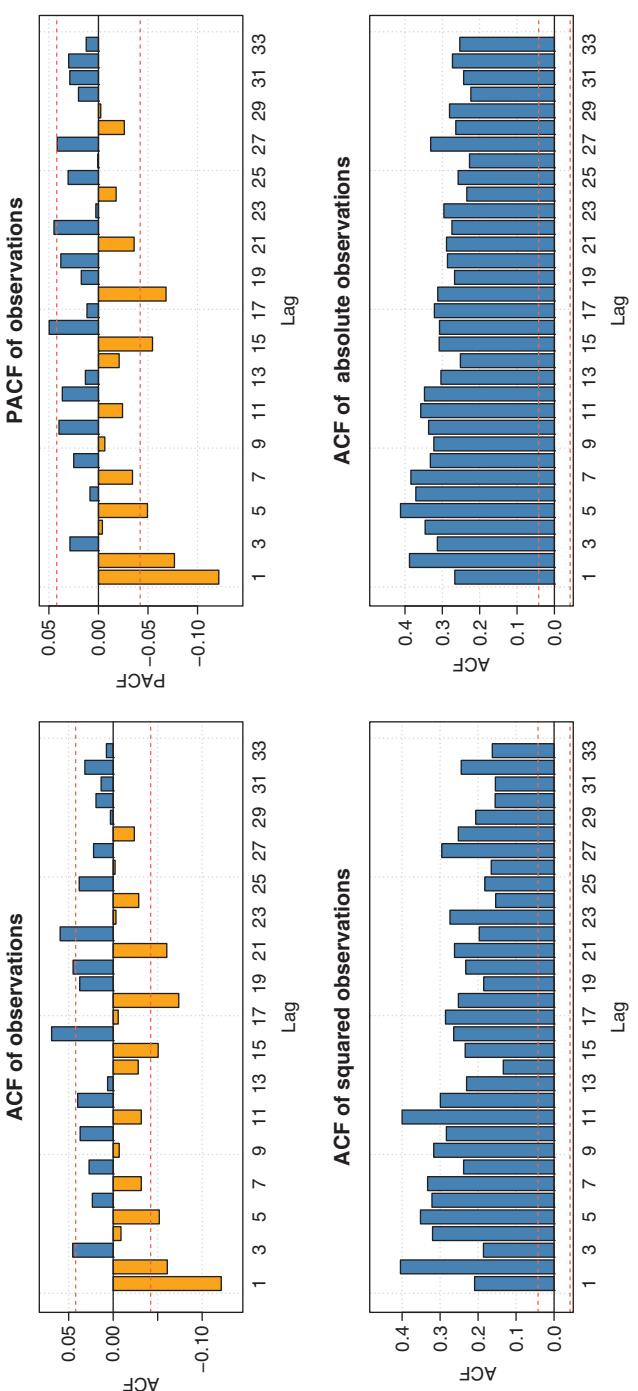
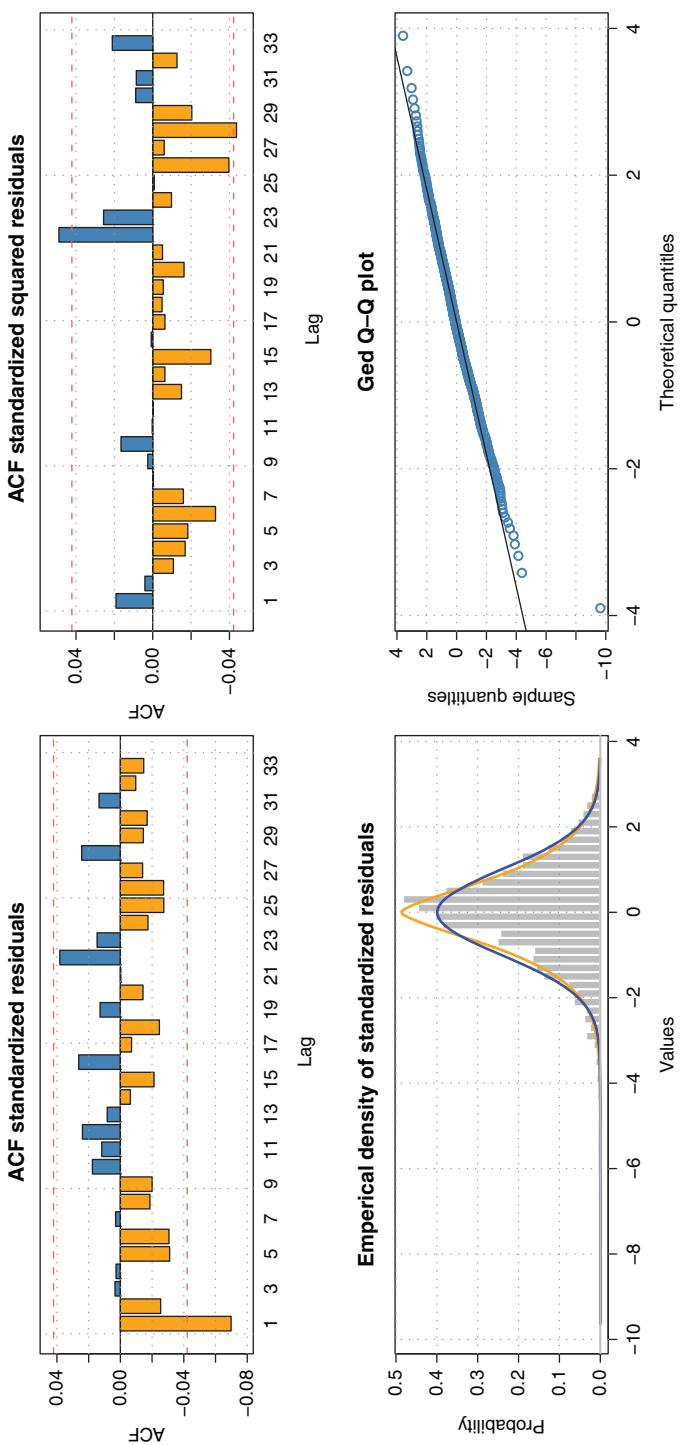


FIGURE 10.11 S&P 500 returns and squared and absolute returns ACF and PACF.



**FIGURE 10.12** S&P 500 volatility model diagnostics: Top are the ACF of the standardized and standardized squared residual. Bottom are the distribution and GED Q-Q plot for the fitted model.

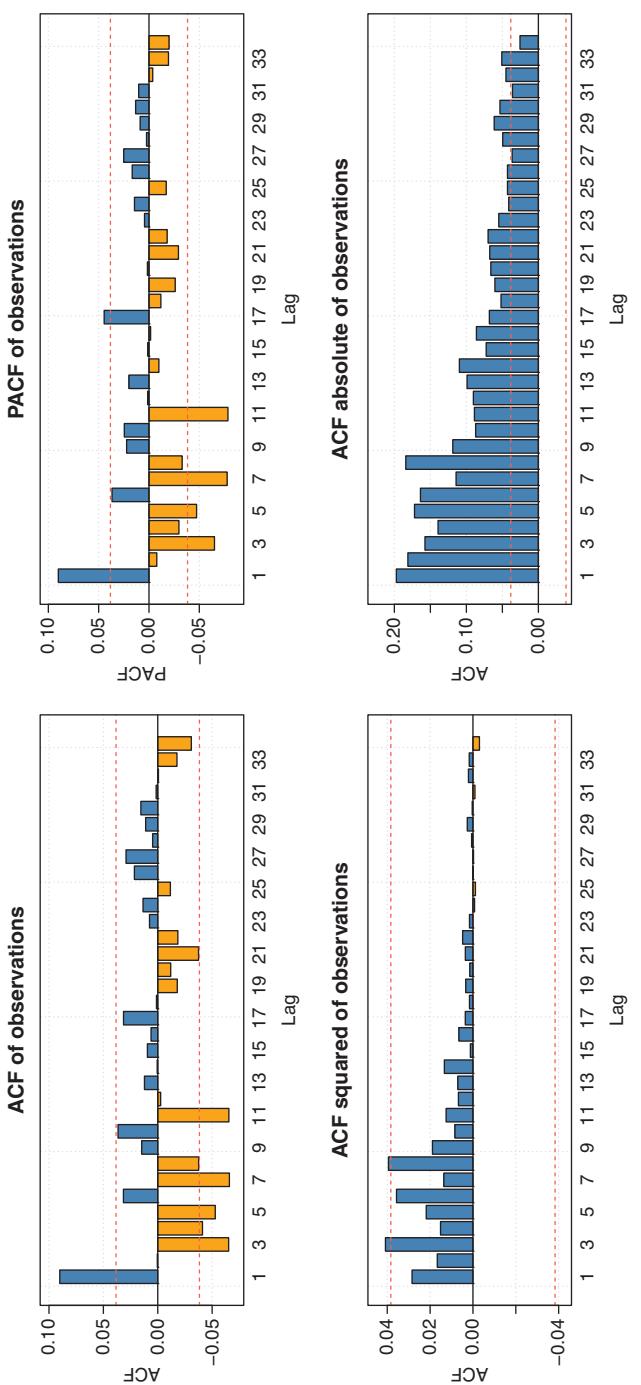
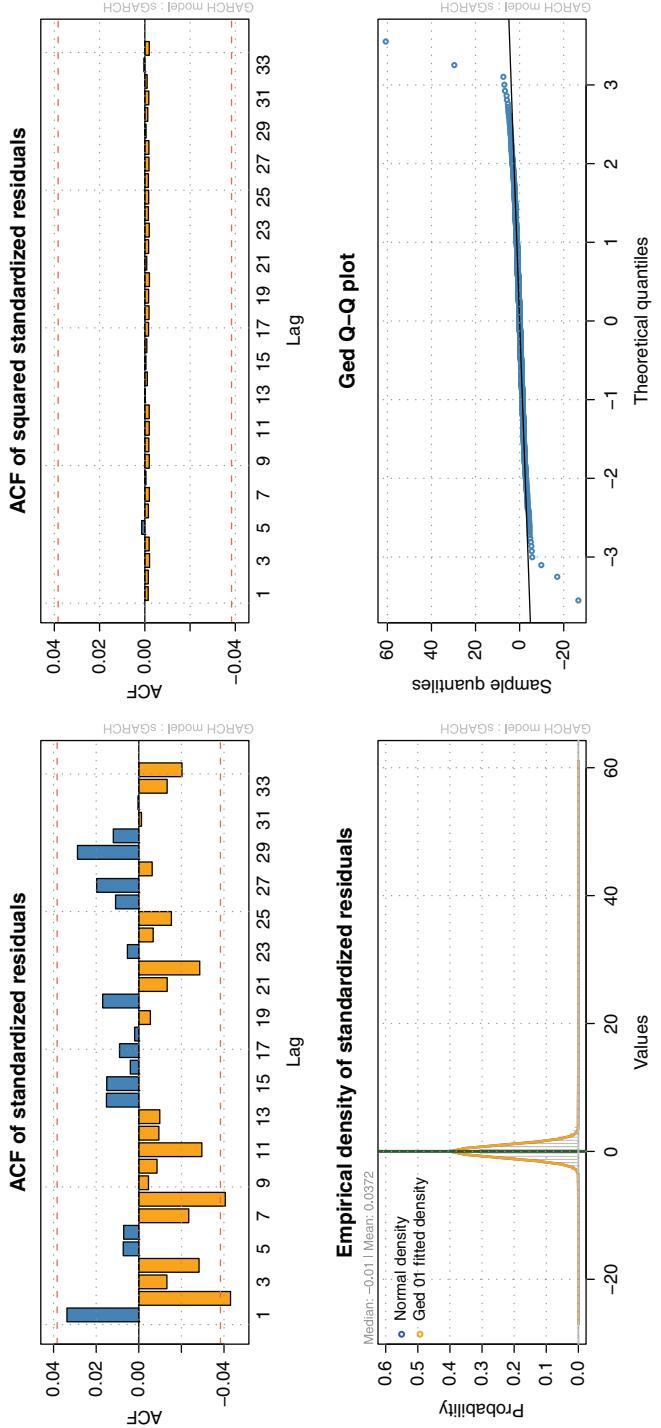


FIGURE 10.13 BAC high-frequency returns and squared and absolute returns ACF and PACF.



**FIGURE 10.14** BAC conditional volatility model diagnostics: Top are the ACF of the standardized and standardized squared residual. Bottom are the distribution and GED Q–Q plot for the fitted model.

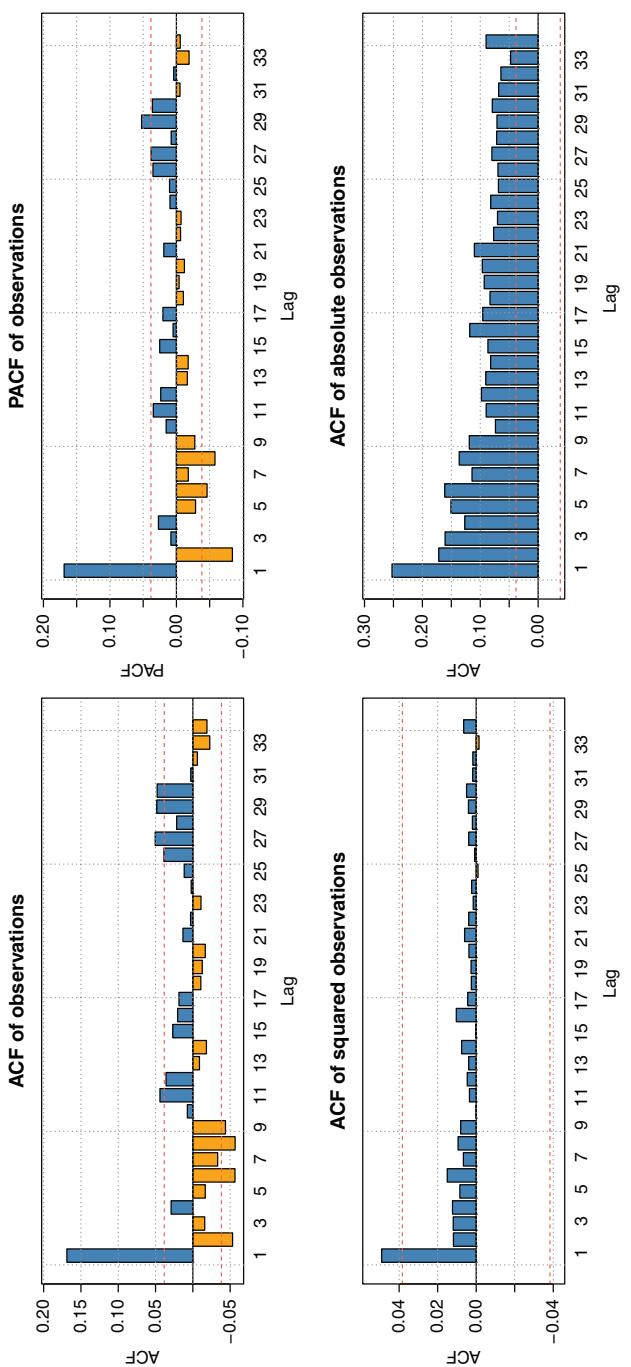
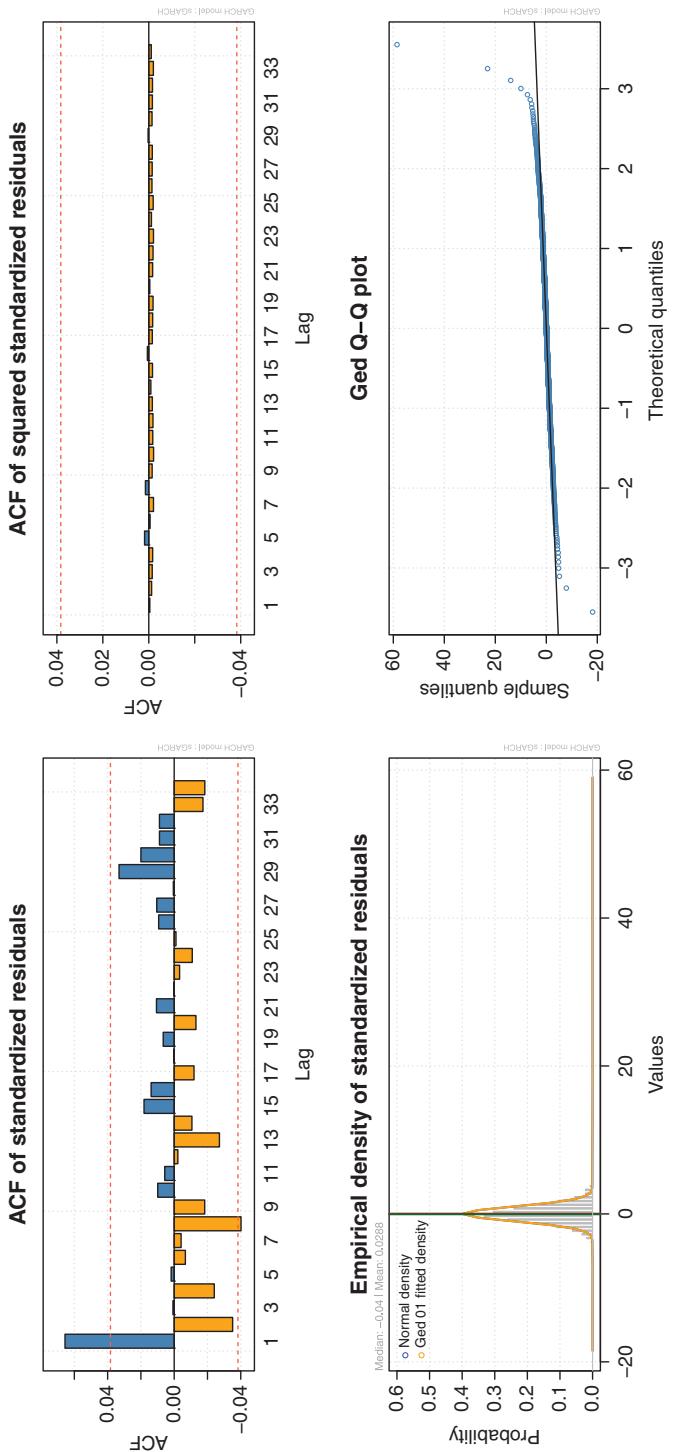


FIGURE 10.15 JPM high-frequency returns and absolute returns ACF and PACF.



**FIGURE 10.16** JPM conditional volatility model diagnostics: Top are the ACF of the standardized and standardized squared residual. Bottom are the distribution and GED Q-Q plot for the fitted model.

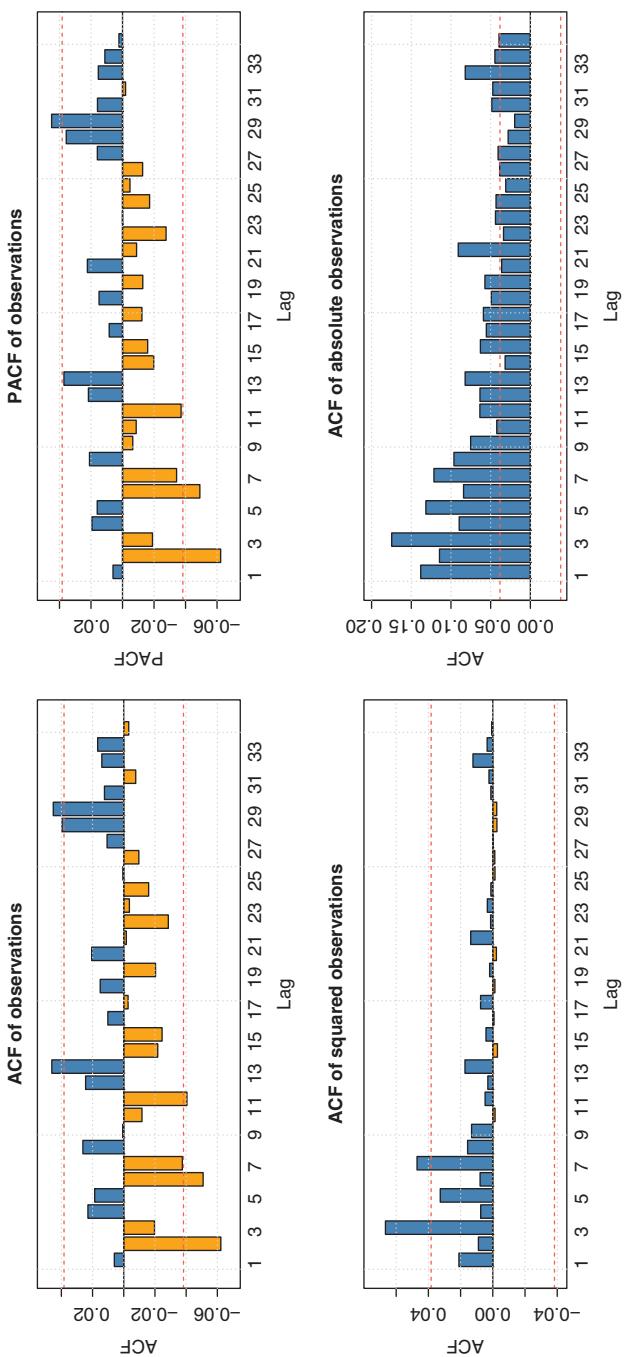
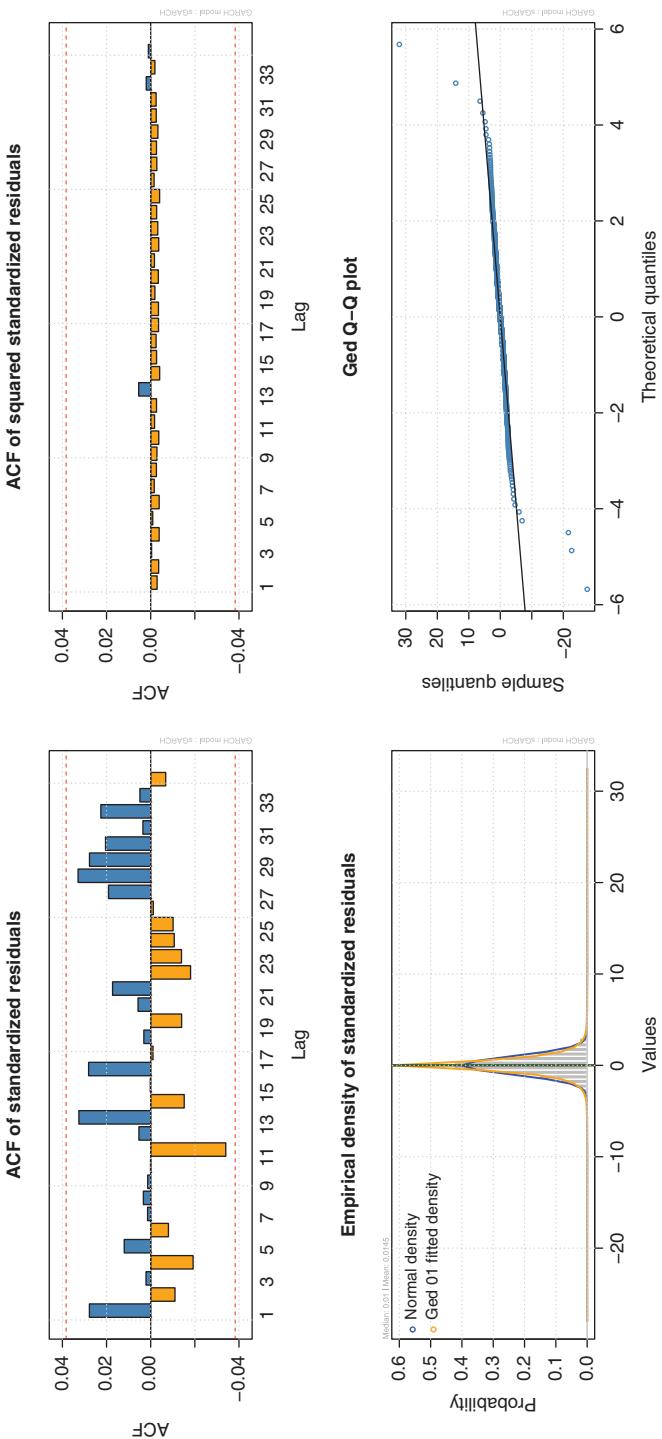
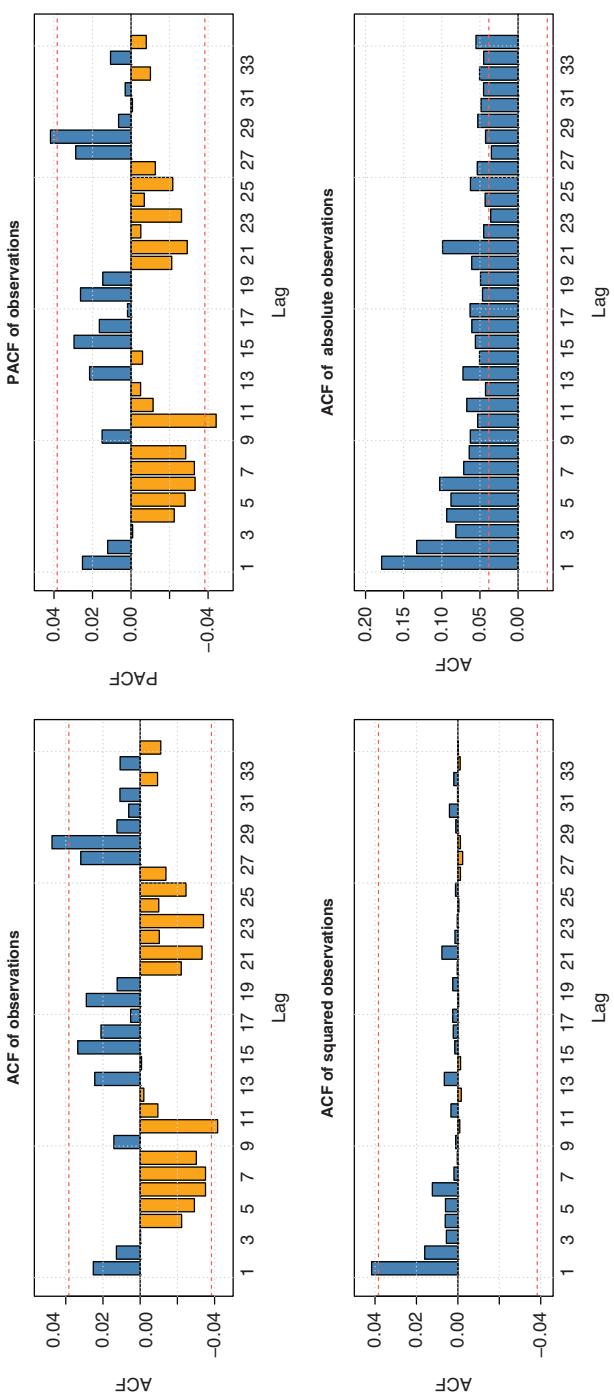


FIGURE 10.17 IBM high-frequency returns and squared and absolute returns ACF and PACF.



**FIGURE 10.18** IBM conditional volatility model diagnostics: Top are the ACF of the standardized and standardized squared residual. Bottom are the distribution and GED Q-Q plot for the fitted model.



**FIGURE 10.19** WMT high-frequency returns and squared and absolute returns ACF and PACF.

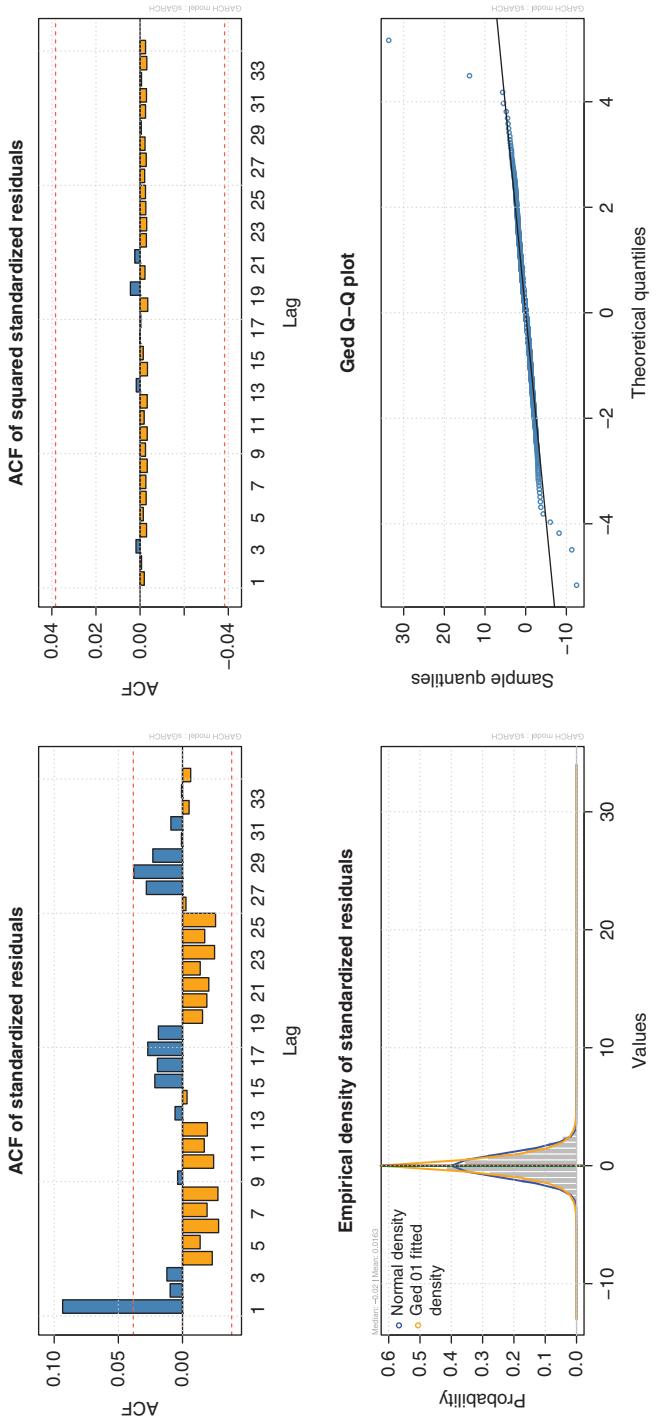


FIGURE 10.20 WMT high-frequency conditional volatility model diagnostics.

**TABLE 10.10 Standardized residuals test.**

Residual tests	Variable	Test stat	Test value	Prob
Jarque–Bera test	R	$\chi^2$	237.12	0.000
Ljung–Box test	R	Q(15)	13.114	0.5176
Ljung–Box test	$R^2$	Q(15)	10.259	0.7430
LM Arch test	R	$TR^2$	7.75041	0.6532

### 10.5.12 ONE-STEP AHEAD PREDICTION OF LAST 10 OBSERVATIONS

We used the MA(1) with FIGARCH(1,d,1) variance model for the returns to predict the last 10 observations by reconstructing a model each time with one-step-ahead predictions of the next observation. The last observed return was  $-0.0218941$ . The result is given in Table 10.11 below.

### 10.5.13 ANALYSIS ON HIGH-FREQUENCY, EARTHQUAKE, AND EXPLOSIVES SERIES

Similar volatility analysis was done on S&P500 index, Bank of America Corp (BAC), JPM, International Business Machines Corp (IBM), and Wal-Mart Stores (WMT), high-frequency data (HFD), earthquake series (EQ2), and EXP series. The results are shown in Tables 10.12 to 10.18 and Figures 10.11 to 10.22.

**TABLE 10.11 Actual returns and prices to forecasted returns and prices.**

Actual	Mean forecast	Standard dev	Prices	Estimate prices
-0.024019	0.001408	0.01737	10,655	10,929
0.014208	0.001533	0.01778	10,809	10,672
0.011982	-0.0009184	0.01873	10,940	10,799
0.016536	-0.0007733	0.01806	11,123	10,931
-0.001906	-0.001063	0.01748	11,103	11,112
0.029206	0.0001227	0.01738	11,433	11,104
-0.001565	-0.001886	0.01632	11,416	11,412
0.008855	0.0001012	0.01816	11,519	11,417
-0.003629	-0.0005748	0.01702	11,478	11,512
0.014302	0.0002351	0.01621	11,644	11,481

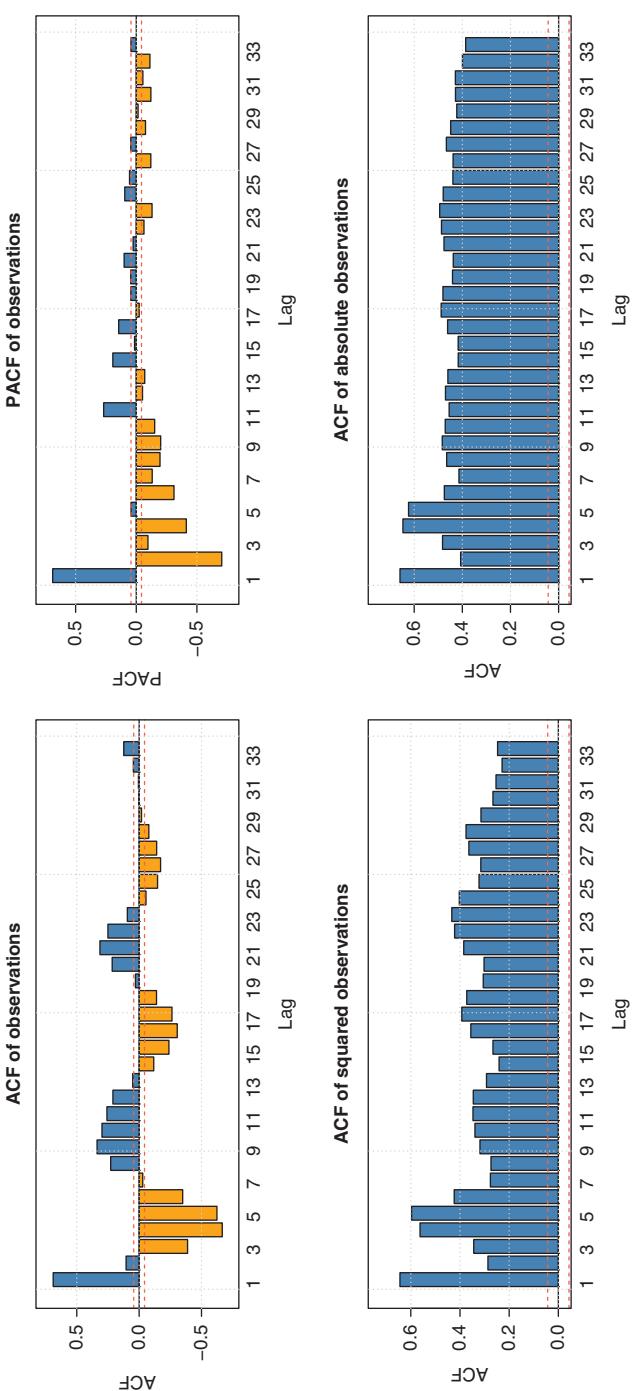


FIGURE 10.21 Explosive series and squared and absolute series ACF and PACF.

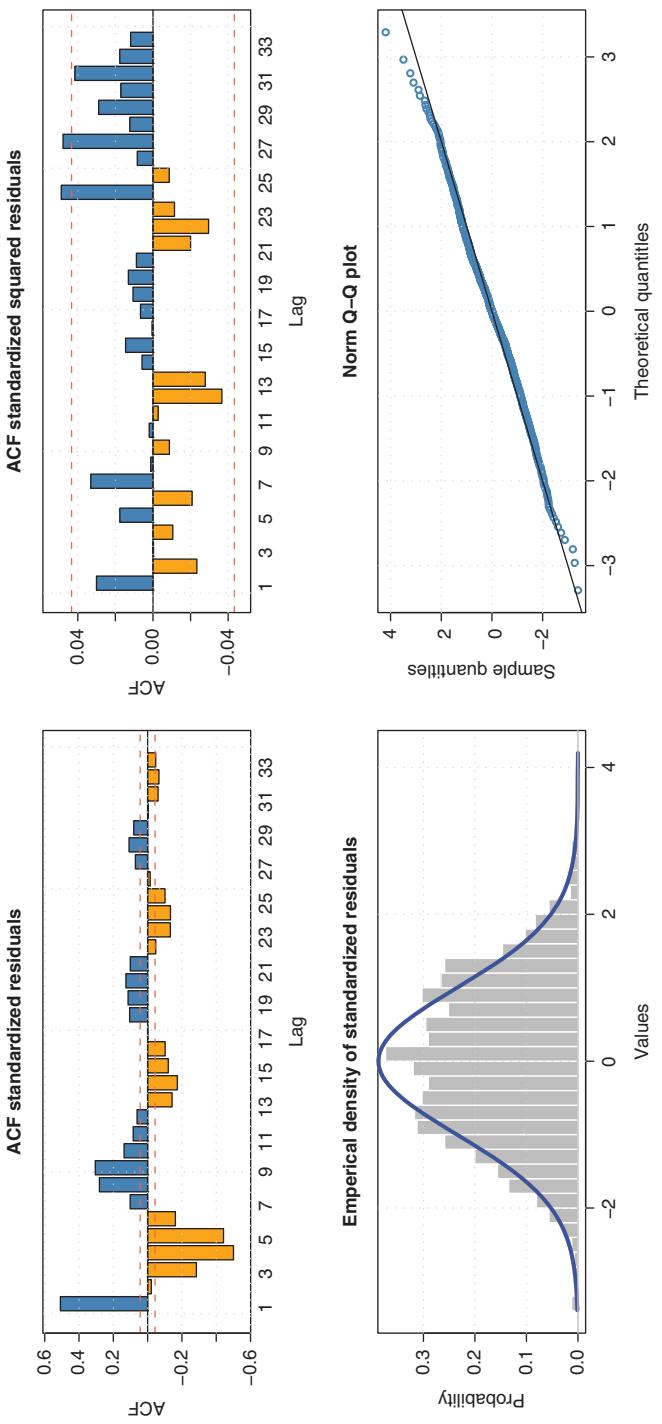


FIGURE 10.22 Explosive series conditional volatility model diagnostics.

The studied HFD corresponds to the collapse of the Bear Stearns in March 2008. The data used consists of the week (five trading days) March 10–14, 2008, before the merging announcement over the weekend as well as the two following trading days March 17 and 18. On Friday, March 14, 2008, at about 9:14 a.m., JP Morgan Chase (JPM) and the Federal Reserve Bank of New York announced an emergency loan to Bear Stearns (of about 29 billion, terms undisclosed) to prevent the firm from becoming insolvent. This bailout was declared to prevent the very likely crash of the market as a result of the fall of one of the biggest investment banks at the time. This measure proved to be insufficient to keep the firm alive and 2 days later, on Sunday March 16, 2008, Bear Stearns signed a merger agreement with JPM essentially selling the company for \$2 a share (price revised on March 24 to \$10/share). The same stock traded at \$172 in January 2007 and \$93 a share in February 2007. Today, this collapse is viewed as the first sign of the risk management meltdown of investment bank industry in September 2008 and the subsequent global financial crisis and recession.

The behavior of stocks representing the financial institutions that were affected by the crisis such as JPM and BAC was studied. We also looked at other institutions, such as IBM and WMT, that should not have been much affected. All the stocks were sampled with the period  $T = 1$  min. During the days considered, the BAC and JPM return fluctuated more than 10% while that of WMT and IBM less than 1%.

The data on earthquakes magnitude was obtained from the US Geological Survey (USGS) from January 1, 1973, to November 9, 2010. The downloaded data contains information about the date, longitude, latitude, and the magnitude of each recorded earthquake. The location of the major earthquake and its distribution defines the studied area. The earthquake magnitude is the data used in the analysis. The sample data is from July 24, 2010, to November 9, 2010, containing 3000 observations with a maximum (7.1), which occurred at August 12, 2011.

We then continue our analyses with the study of seismic traces of a mining explosion. A seismic trace is a plot of the earth's motion over time. The data presented are measurements of the earth's vertical displacement where the recording frequency is 40 per second. The data sets are from a recording station in Scandinavia and are reported by Shumway and Stoffer [34].

**10.5.13.1 ADF test****TABLE 10.12 Augmented Dickey–Fuller (ADF) unit root test values.**

$H_0: I(1)$ , critical values: **-2.567 (1%); -1.941 (5%); -1.616 (10%)** for the indices, HFD, earthquake series, and the explosive series.

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
Return	-12.321	-12.493	-14.75	-14.328	-14.752	-14.325	-11.352	-18.232
Sqd. Ret	-4.7648	-4.833	-12.34	-13.001	-12.455	-12.456	-11.367	-5.628
Abs. Ret	-4.8911	-4.845	-9.193	-9.6733	-10.032	-10.032	-11.352	-4.361

**10.5.13.2 KPSS test****TABLE 10.13 Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) unit root test values.**  $H_0: I(0)$ , critical values: **0.739 (1%); 0.463 (5%); 0.347 (10%)** for the indices, HFD, earthquake series, and the explosive series.

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
Return	0.166	0.208	0.167	0.146	0.089	0.089	6.641	0.007
Sqd. Ret	1.549	1.724	0.108	0.099	0.075	0.075	6.850	0.867
Abs. Ret	2.544	2.968	0.979	1.271	0.598	0.598	6.641	1.233

**10.5.13.3 ARCH test****TABLE 10.14 AutoRegressive Conditional Heteroscedastic test for correlation in squared and absolute series. Critical values: 32.91 (0.1%); 26.22 (1%); and 21.03 (5%)** for the indices, HFD, earthquake series, and the explosive series.

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
ARCH LM	731.136	703.342	15.454	8.198	21.741	5.649	177.227	1299
Ljung. Box	2508.581	2539.780	18.621	8.838	23.145	5.938	364.196	4266
Pierce. Box	2498.418	2529.511	18.570	8.824	23.088	5.929	363.255	4251

**10.5.13.4 Whittle estimate of d****TABLE 10.15 Whittle estimate of the long-memory parameter ( $d$ ) for the indices, HFD, earthquake series, and the explosive series, where  $d \in (-0.5, 0.5)$ .**

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
Return	-0.083*	-0.085*	0.031	0.091**	-0.019	0.005	0.136**	0.499**
Sqd. Ret	0.199**	0.202**	0.042	0.042	0.038	0.035	0.140**	0.419**
Abs. Ret	0.218**	0.215**	0.183**	0.204**	0.145**	0.159**	0.136**	0.385**

\* significant at 5%; \*\*significant at 1%.

### 10.5.13.5 AIC

**TABLE 10.16 AIC of models fitted to the indices, HFD, earthquake series, and the explosive series. The model with minimum AIC is selected.**

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
GARCH(1,1)	-6.544	-6.417	-11.1720	-11.042	-12.632	-12.713	2.508	-3.473
ARFIMA+GAR	-6.545	-6.420	-11.1783	-11.051	-12.623	-12.694	2.484	-3.484
FIGARCH(0,d,0)	-6.506	-6.375	-11.1750	-11.044	-12.624	-12.695	2.517	-3.472
FIGARCH(0,d,1)	-6.541	-6.417	-11.1780	-11.043	-12.623	-12.694	2.518	-3.471
FIGARCH(1,d,1)	-6.553	-6.429	-11.1781	-11.041	-12.622	-12.693	2.510	-3.657
IGARCH(1,1)	-6.544	-6.417	-11.1730	-11.043	-12.624	-12.695	2.507	-3.472

### 10.5.13.6 BIC

**TABLE 10.17 BIC of models fitted to the indices, HFD, earthquake series, and the explosive series. The model with minimum BIC is selected.**

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
GARCH(1,1)	-6.534	-6.407	-11.162	-11.031	-12.612	-12.682	2.517	-3.461
ARFIMA+GAR	-6.533	-6.407	-11.163	-11.038	-12.619	-12.701	2.495	-3.644
FIGARCH(0,d,0)	-6.499	-6.367	-11.169	-11.035	-12.615	-12.686	2.523	-3.463
FIGARCH(0,d,1)	-6.531	-6.406	-11.167	-11.032	-12.612	-12.682	2.526	-3.459
FIGARCH(1,d,1)	-6.541	-6.416	-11.165	-11.032	-12.609	-12.679	2.521	-3.469
IGARCH(1,1)	-6.537	-6.409	-11.164	-11.034	-12.614	-12.685	2.514	-3.464

### 10.5.13.7 *d* Estimate

**TABLE 10.18 Long memory volatility models and their estimated long memory parameter for the indices, HFD, earthquake series, and the explosive series.**

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
ARFIMA+GAR	-0.032*	-0.045**	0.10**	.236**	.107**	0.152*	0.092**	0.861**
FIGARCH(0,d,0)	0.215**	0.214**	0.00	1.24	0.158	0.00	0.097	0.390**
FIGARCH(0,d,1)	0.307**	0.313**	0.00	-0.12	0.246	0.00	0.115	0.390**
FIGARCH(1,d,1)	0.602**	0.634**	0.00	0.00	0.443	0.00	0.331	0.782**

\*significant at 5%; \*\*significant at 1%.

## 10.6 Discussion and conclusion

The underlying volatility processes in earthquake series, high-frequency financial data, financial indices, and explosive data were explored using various GARCH models in this thesis. The GARCH models applied include basic GARCH, IGARCH, ARFIMA (0,d,0)-GARCH, and FIGARCH specifications. The methodology is not new; however, the major

contribution of this work comes in the realm of applications. The methodology was applied to three domains: geophysics (earthquake data), finance (high-frequency financial data and indices), and explosives data. In all the applications, the methodology provides insight into features of the series volatility.

The results show that the indices (DJIA and S&P 500) returns and the explosives (EXP) series volatility had the highest persistence that were best described by using FIGARCH a long memory model. This result is in line with the previous works of Breidt et al. [1], Mariani et al. [35], and Mike So [36] for the indices volatility. In his work, Mike So [36] applied the modified rescaled range test (R/S) proposed by Lo [37] and the semiparametric test (GPH) proposed by Geweke and Porker-Hudak [38] to detect the existence of long-term dependence in volatility in the S & P 500 index and Dow Jones Industrial Average index. He used three proxies of the variability of returns to achieve this result: the absolute mean deviation, the squared mean deviation, and the logarithm of the absolute mean deviation. Mariani et al. [35], in their paper, used the Hurst exponent and the Detrended fluctuation analysis methodology to show the existence of long-memory effects in the international Market indices, that is, Morgan Stanley Capital International Europe, Australasia, and Far East index and the Emerging market index, and compare their result with S& P 500. They also found that immediately before a crisis the estimate of the long memory increases, while during the time of the crisis the stocks behave randomly. Finally Breidt et al. [1] also used their proposed long-memory stochastic volatility model to indicate the existence of long memory in financial indices. Hence, our results of long-memory in the indices volatility reinforced previous results by different methodology.

The ARFIMA(0,d,0)-GARCH specification was preferred for BAC and JPM high-frequency series whose volatility was found to be intermediate. The intermediate memory found in the high-frequency data is consistent with previous results of Barany et al. [39] and Mariani et al. [40] except with the WMT and IBM high-frequency data volatility that were best described using the GARCH model, which has short memory. Mariani et al. [40] employed the relationship between the Hurst parameter ( $H$ ) and the Detrended fluctuation analysis parameter ( $\alpha$ ) and compared with the value 0.5 to investigate the memory behavior of BAC and JPM high-frequency data from March 10–18, 2008 (Bear Stearns financial crisis). Their results gave Hurst estimate of 0.63 and 0.62 for BAC and JPM, respectively, which implies long memory. Our values for the fractional

difference parameter estimate 0.1 and 0.236 for the BAC and JPM, respectively, also showed long memory.

The earthquake series was divided into two regions: symmetrically (EQ1) and nonsymmetrically (EQ2) distributed. Both regions showed intermediate memory. Our analysis indicated that the earthquake series showed long memory while the explosive series showed short memory. On the other hand, both the explosive and the earthquake series volatility showed long memory, but the persistence in the explosive volatility was higher than that of the earthquake. The order of persistence (memory) in series volatility from highest to lowest is:

1. Indices (DJIA and S&P 500) and explosives data
2. Earthquake data
3. BAC and JPM high-frequency financial data
4. WMT and IBM high-frequency financial data

The predictions made from the  $MA(1) - FIGARCH(1, d, 1)$  model for the DJIA index offered good results since all the actual observations were within the prediction limits. The earthquake series predictions from our assumed model were fairly accurate since we had 8 out of 10 earthquake directions correctly predicted.

The outliers observed from the generalized error Q–Q plot for the indices correspond to significant drops in prices in a short time period (1 or 2 days). This is what happened on October 15, 2008, for both DJIA and S&P 500 indices.

## References

1. F. J. Breidt, N. Crato, and P. De Lima (1998). *The detection and estimation of long memory in stochastic volatility*. Journal of Econometrics, 83, 325–348.
2. P. M. Robinson (1991). *Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression*. Journal of Econometrics, 47, 67–84.
3. N. Shephard (1996). *Statistical aspects of ARCH and stochastic volatility*. In D. R. Cox, D. B. Hinkley, and O. E. Barndorff-Nielsen, editors. Time Series Models: In Econometrics, Finance and Other Fields. Chapman Hall, London.

4. I. N. Lobato and N. E. Savin (1998). *Real and spurious long-memory properties of stock-market data*. Journal of Business & Economic Statistics, 16, 261–283.
5. R. T. Baillie (1996). *Long memory processes and fractional integration in econometrics*. Journal of Econometrics, 73, 5–59.
6. O. E. Barndorff-Nielsen and N. Shephard (2001). *Modelling by Levy processes for financial econometrics*. In Levy Processes. Birkhauser, Boston, MA, pp. 283–318.
7. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel (1994). *Time series analysis: forecasting and control*. Prentice Hall, 1994.
8. R. F. Engle (1982). *Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation*. Econometrica, 50(4), 987–1007.
9. T. Bollerslev (1986). *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, 31, 307–327.
10. R. F. Engle and V.K. Ng (1991). *Measuring and testing the impact of news on volatility*. Journal of Finance, 48(5), 1749–1778.
11. D. B. Nelson (1991). *Conditional heteroskedasticity in asset returns: a new approach*. Econometrica, 59, 347–370.
12. A. C. Harvey, E. Ruiz, and N. Shephard (1994). *Multivariate stochastic variance models*. Review of Economic Studies, 61, 247–265.
13. R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen (1996). *Fractionally integrated generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, 74, 3–30.
14. T. Bollerslev and H. Mikkelsen (1996). *Modeling and pricing long-memory in stock market volatility*. Journal of Econometrics, 73, 151–184.
15. P. M. Robinson and M. Henry (1999). *Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels*. Econometric Theory, 15, 299–336.
16. M. Henry (2001). *Averaged periodogram spectral estimation with long-memory conditional heteroscedasticity*. Journal of Time Series Analysis, 22(43), 1459.
17. K. Aas and I. H. Haff (2006). *The generalized hyperbolic skew student's t-distribution*. Journal of Financial Econometrics, 4(2), 275309.

18. R. A. Rigby and D. M. Stasinopoulos (2005). *Generalized additive models for location, scale and shape*. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(3), 507–554.
19. R. Engle and T. Bollerslev (1996). *Modelling the persistence of conditional variances*. Econometric Reviews, 5, 1–50.
20. D. Ruppert (2011). *Statistics and data analysis for financial engineering*. Springer Science-Business Media.
21. W. Palma (2007). *Long-Memory Time Series (Theory and Methods)*. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New Jersey, 2007.
22. R. S. Tsay (2005). *Analysis of financial time series*. Vol. 543. John Wiley & Sons, Inc.
23. T. Bollerslev (1986). *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, 31, 307–327.
24. R. F. Engle and T. Bollerslev (1986). *Modelling the persistence of conditional variances*. Econometric Reviews, 5(1), 1–50.
25. D. B. Nelson (1990). *Stationarity and persistence in the GARCH ( $l, l$ ) model*. Econometric theory. Cambridge University Press.
26. C. Granger and R. Joyeux (1980). *An introduction to long memory time series models and fractional differencing*. Journal of Time Series Analysis, 1, 15–30.
27. J. Hosking (1981). *Fractional differencing*. Biometrika, 68, 165–176.
28. W. Palma (2007). *Long-memory time series: theory and methods*. Vol. 662. John Wiley & Sons, Inc.
29. D. A. Dickey and W. A. Fuller (1979). *Distribution of the estimators for autoregressive time series with a unit root*. Journal of the American Statistical Association, 74(366a), 427–431.
30. N. H. Chan and C. Z. Wei (1988). *Limiting distributions of least squares estimates of unstable autoregressive processes*. Annals of Statistics, 16, 367–401.
31. P. C. B. Phillips (1987). *Time series regression with unit roots*. Econometrica, 55, 277–301.
32. W. A. Fuller (1976). *Introduction to Statistical Time Series*. John Wiley & Sons, Inc., New York.
33. D. Kwiatkowski, P. Phillips, P. Schmidt, and Y. Shin (1992). *Testing the null hypothesis of stationarity against the alternative of a unit root*:

- How sure we are that economic time series have a unit root?* Journal of Econometrics, 54, 159–178.
34. R. H. Shumway and D. S. Stoffer (2010). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.
  35. M. Mariani, I. Florescu, M. P. B. Varela, and E. Ncheuguim (2010). *Study of memory effect in international market indices*. Physica A, 2010, 389(8), 1653–1646.
  36. M.P. So (2010). *Long-term memory in stock market volatility*. Applied Financial Economics, 10(5), 519–524.
  37. A. W. Lo (1991). *Long-term memory in stock market prices*. Econometrica, 59, 1279–1313.
  38. J. Geweke and S. Porter-Hudak (1983). *The estimation and application of long memory time series models*. Journal of Time Series Analysis, 4, 221–238.
  39. E. Barany and M. P. B. Varela (2011). *Stochastic Differential Equations and Levy Models with Applications to High Frequency Data*. Handbook of Modeling High-Frequency Data in Finance, pp. 327–346.
  40. M. Mariani, I. Florescu, M. P. B. Varela, and E. Ncheuguim (2009). *Long correlations and Levy models applied to the study of memory effects in high frequency (tick) data*. Physica A, 2009, 388(8), 1659–1664.

## Chapter Eleven

# Scale Invariance and Lévy Models Applied to Earthquakes and Financial High-Frequency Data

M. P. Beccar-Varela<sup>1</sup>, Ionut Florescu<sup>2</sup>, and I. SenGupta<sup>3</sup>

<sup>1</sup>*Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX, USA*

<sup>2</sup>*Financial Engineering Division, Stevens Institute of Technology, Hoboken, NJ, USA*

<sup>3</sup>*Department of Mathematics, North Dakota State University, Fargo, ND, USA*

### 11.1 Introduction

In recent years, we have observed renewed interest in describing *critical phenomena* ([1]) using several modeling techniques. Ising models have been used in [1–3], phase transition [4], fitting the data with exponential sequence [5] or using the so-called scale invariance property [6, 7]. In the current work, we generalize the scale invariance approach, and we further use a technique based on truncated Lévy models to estimate the first instance when a critical event may start to be announced.

In Section 11.2, we present a deterministic model and the governing equations based on general scale invariant functions. We apply this deterministic model to study geophysical data characteristic for earthquakes. In Section 11.3, we present a stochastic standardized Lévy flight model and

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

we apply this stochastic model to the same geophysical data. In Section 11.4, we apply the scale invariance model to financial high-frequency data corresponding to the Bear Stearns collapse in March 2008. In Section 11.5, we present a brief description of the codes. Finally, in Section 11.6, we summarize the results obtained when applying the two methodologies to both earthquake and financial data.

## 11.2 Governing equations for the deterministic model

**Definition 11.1.** *We say that a function  $A$  is scale invariant with a scale parameter  $\lambda$  if and only if for every  $t \in \text{Domain}(A)$ :*

$$A(\lambda t) = \mu A(t) \quad (11.1)$$

where  $\mu$  is a constant dependent only on the scale factor  $\lambda$ .

As a trivial example of such function,  $A(t) = t$  is scale invariant for any  $\lambda \in \mathbb{R}$  with the constant  $\mu = \lambda$ .

In our previous study [3], using geophysical data, we observed that smaller earthquakes preceding a major event have a possible log-periodic structure. In the cited work [3], we approximate the periods between earthquakes up to the major earthquake, using an equation obtained from a temporal energy evolution in a diamond Ising model:

$$\ln T_N = \xi N + \ln(\alpha)$$

In this chapter, we extend this work and use the deterministic scale invariance property (Definition 11.1) to estimate the time of the major event.

We define the following function  $A$ :

$$A(t) = \begin{cases} 1 & \text{when } \ln t = C_1 k + C_2, k \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases} \quad (11.2)$$

It is not hard to see that the function defined in (11.2) is scale invariant under the transformation

$$t \rightarrow (e^{C_1})^n t, n \in \mathbb{Z}$$

for  $n \in \mathbb{Z}$ . Using the notations in (11.1), we see that  $\mu = 1$  and  $\lambda = (e^{C_1})^n, n \in \mathbb{Z}$ .

We next introduce the following function:

$$B(t) = A(t_c - t), \quad (11.3)$$

where  $t_c$  is the critical time of the major event, which we do not know at this point.

Note that  $B(t)$  alternates between 0 and 1, and we define the set  $S$ :

$$S = \{t | B(t) = 1\},$$

The critical time  $t_c$  is defined as the limit of the accumulation points of the set; that is, for  $C_1 \neq 0$ , the set  $S$  has an accumulation point at  $t_c$ .

A particular type of scale invariance is the one arising in the existence of intermittences or stationary intervals, constant in the logarithm of the independent variable [6, 7].

$$\begin{aligned} f^F(t) &= \beta e^{\alpha F(\log_a t)}, \alpha > 0 \\ f^C(t) &= \beta e^{\alpha C(\log_a t)} \end{aligned} \quad (11.4)$$

where  $F$  is the floor function and  $C$  is the ceiling function. From the form of equations (11.4), it is possible to see that (11.4) is scale invariant under the transformation:

$$t \rightarrow a^n t, n \in \mathbb{Z}$$

since

$$\begin{aligned} f^F(a^n t) &= \beta e^{\alpha F(\log_a a^n t)} = \beta e^{\alpha F(\log_a t + n)} \\ &= \beta e^{\alpha(F(\log_a t) + n)} = \beta e^{\alpha F(\log_a t)} \cdot \underbrace{e^{\alpha n}}_{\mu} \end{aligned}$$

We can see that constant  $\mu = e^{\alpha n}$  for  $\lambda = a^n$ . Let us introduce new one-parametric scale invariant function that generalizes (11.4):

$$f_x(t) = \beta e^{\alpha(F(\log_a t - x) + x)}, x \in \mathbb{R} \quad (11.5)$$

It is easy to see that

$$\begin{aligned} f_0(t) &= f^F(t) \\ f_1(t) &= f^C(t) \end{aligned}$$

### 11.2.1 APPLICATION TO GEOPHYSICAL (EARTHQUAKE DATA)

The geological data was obtained from U.S. Geological Survey (USGS) from January 1, 1973, to November 9, 2010. Downloaded data contains information about the date, longitude, latitude, and magnitude of each recorded earthquake in the region.

The location of the major earthquake chosen defines the area studied. The area chosen cannot be too small (lack of data) or too big (noise from unrelated events). The data is obtained using a *square* centered at

the coordinates of the major event. The sides of the square were usually chosen as  $\pm 0.1^\circ$ – $0.2^\circ$  in latitude and  $\pm 0.2^\circ$ – $0.4^\circ$  in longitude. A segment  $0.1^\circ$  of latitude at the equator is  $\approx 11.11$  km  $\approx 6.9$  miles in length.

The earthquake magnitude is the recorded data used in the analysis. The policy of the USGS regarding recorded magnitude is the following [8]:

- Magnitude is a dimensionless number between 1 and 12.
- The reported magnitude should be moment magnitude, if available.
- The least complicated, and probably most accurate, terminology is to just use the term *magnitude* and to use the symbol  $M$ .

The coefficients  $C_1, C_2$  and time  $t_c$  for (11.2) are obtained by least square fitting of data. We illustrate the fit in Figures 11.2, 11.3, 11.4, 11.5, 11.6, 11.7, and 11.8. As we may observe from equation (11.2) and also Table 11.1, the set of coefficients  $C_1, C_2$  is not unique. For example, if  $C_1, C_2$  are valid coefficients, then  $C_1, C_2 + C_1$  or  $C_1, C_2 - C_1$  is also a valid set of coefficients. To make the selection unique, we select an arbitrary peak as a peak with  $k = 1$ , and this standardizes the data and makes the choice of the coefficients  $C_1$  and  $C_2$  unique.

### 11.2.2 RESULTS

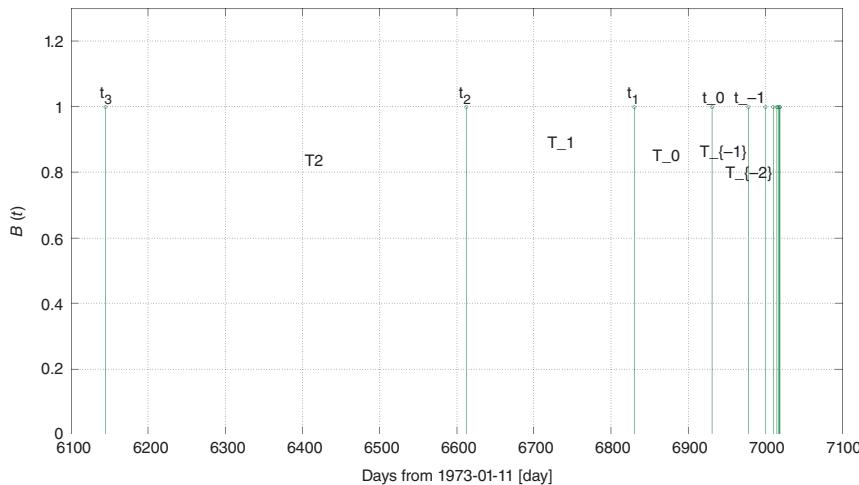
The least square fitting is performed in the following way. Let  $\vec{t} = (t_1, \dots, t_{k+1})$  denote the times of earthquakes preceding some major even. The times are reordered in decreasing order such that  $t_i > t_{i+1}$ . This is done since  $t_c - t$  is a decreasing sequence. The least square fitting is performed by minimizing the following error function:

$$e(C_1, C_2, \vec{t}) = \sum_{i=1}^k \left( \underbrace{(t_i - t_{i+1})}_{\tilde{T}_i} - \underbrace{\left| e^{C_1 i + C_2} - e^{C_1(i+1) + C_2} \right|}_{T_i} \right)^2 \quad (11.6)$$

where vector  $\vec{t}$  was held fixed and was corresponding to data obtained from USGS. The  $i^{\text{th}}$  period will be

$$T_i = t_i - t_{i+1} = \left| e^{C_1 i + C_2} - e^{C_1(i+1) + C_2} \right| = e^{C_1 i} e^{C_2} |1 - e^{C_1}|. \quad (11.7)$$

These time periods form an exponential sequence if the estimated  $C_1$  is positive.



**FIGURE 11.1** Times  $t_1, \dots, t_k$  need to correspond to the real earthquake events.

Then the accumulation point, the critical time  $t_c$ , will simply be:

$$t_c = t_{k+1} + \sum_{j=-\infty}^k T_j \quad (11.8)$$

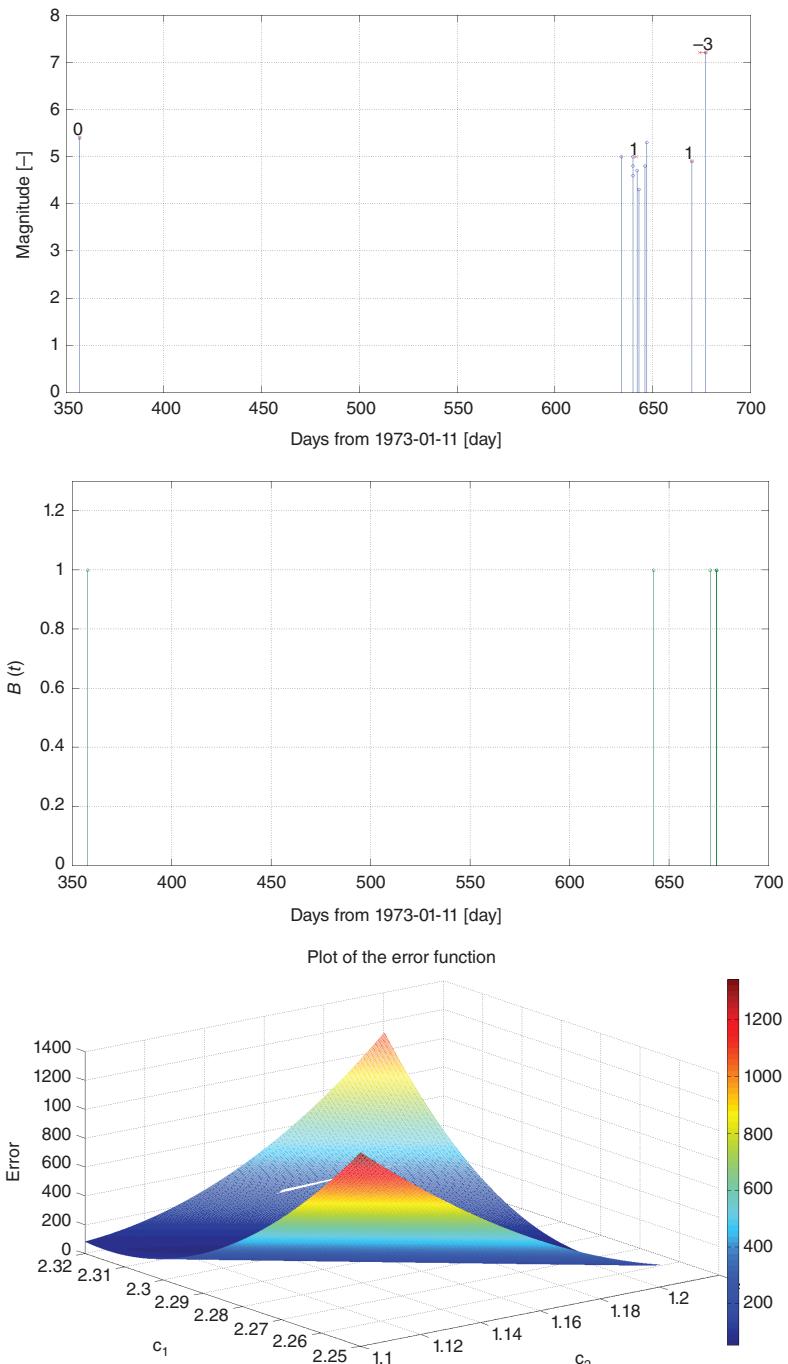
where  $t_i$ 's are the times corresponding to the real events right before the major one. Figure 11.1 next is illustrating the calculation, which is also illustrated in Figure 11.3.

In the following figures (Figures 11.2, 11.3, 11.4, 11.5, 11.6, 11.7, and 11.8), the first image plots the magnitude of the recorded earthquakes in time, the second image plots  $B(t)$  (as calculated using equation (11.3)) and finally the 3D plot is displaying the error function (as given in (11.6)).

Table 11.1 presents in a concise form the estimated critical times of the major event together with the estimated coefficient  $C_1$  and the actual earthquake day.

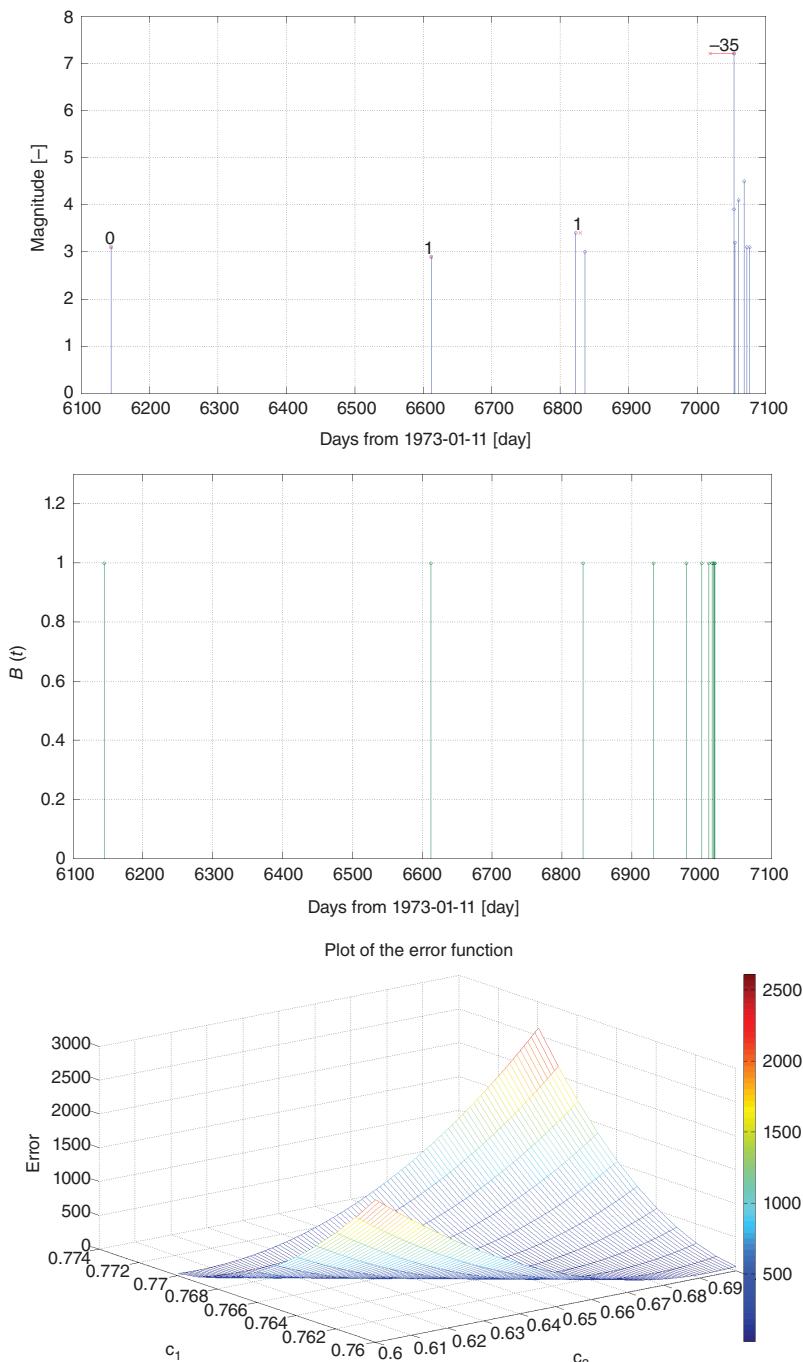
## 11.3 Lévy flights and application to geophysics

In this section, we compute the critical time described in the previous section, using a different method. Specifically, we consider the magnitude process as a stochastic process with increments modeled by a Lévy distribution. This distribution has the scale invariance property that was



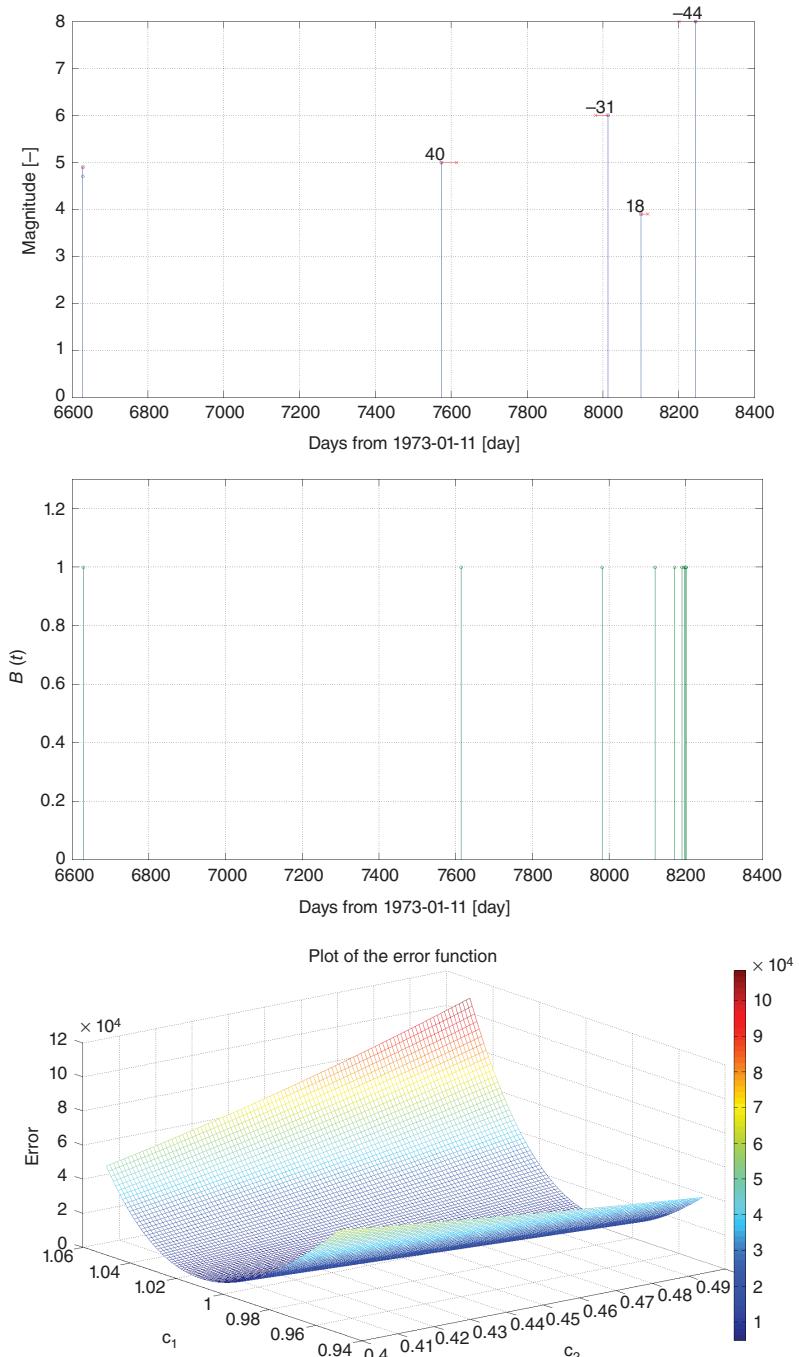
**FIGURE 11.2** Earthquake data and fitting with function (11.2).

- Latitude:  $-12.5^\circ \pm 0.17^\circ$
- Longitude:  $-77.78^\circ \pm 0.34^\circ$
- Time on the  $x$  axis is measured in days from 1973/01/11.



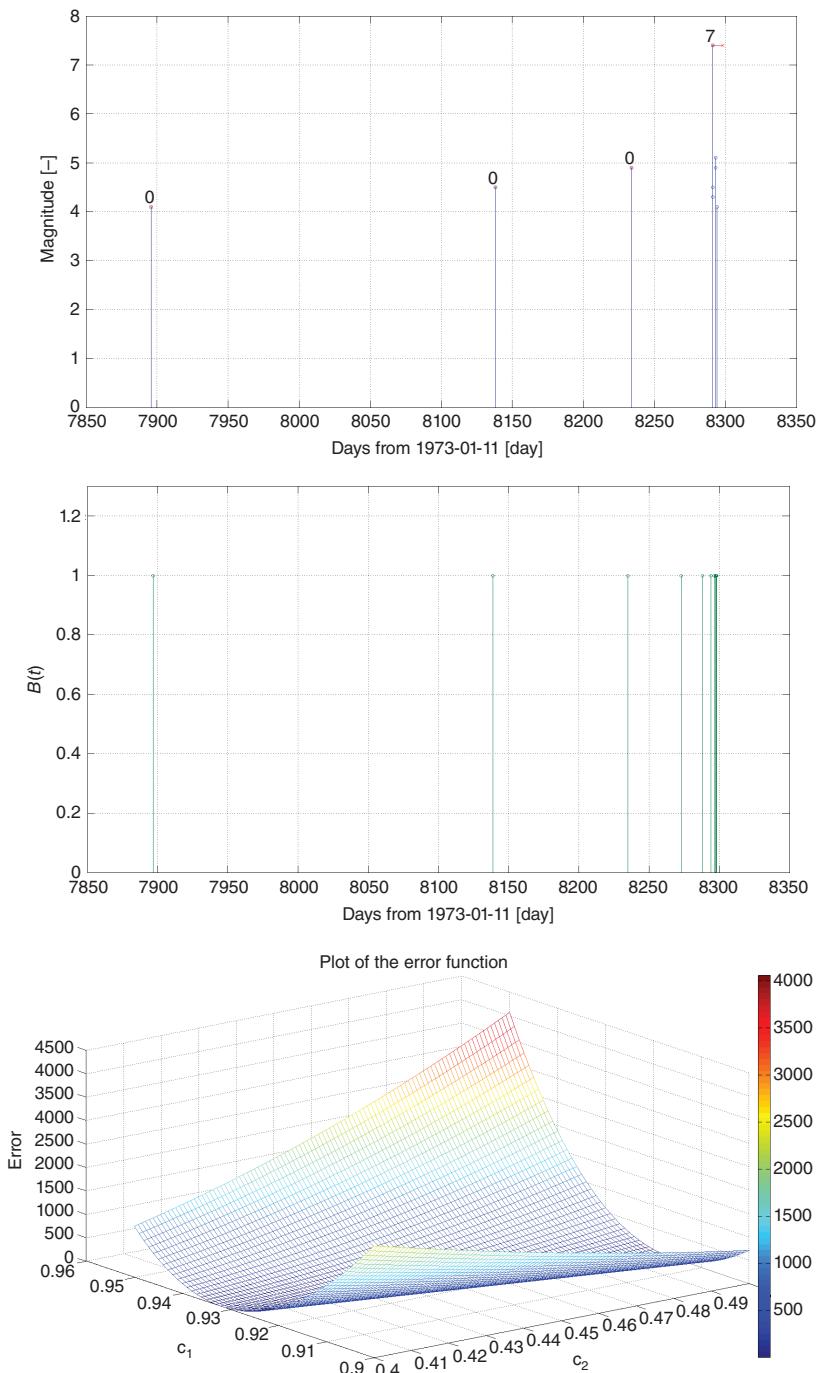
**FIGURE 11.3** Earthquake data and fitting with function (11.2).

- Latitude:  $40.37^\circ \pm 0.1^\circ$
- Longitude:  $-124.32^\circ \pm 0.2^\circ$
- Time on the x axis is measured in days from 1973/01/11.



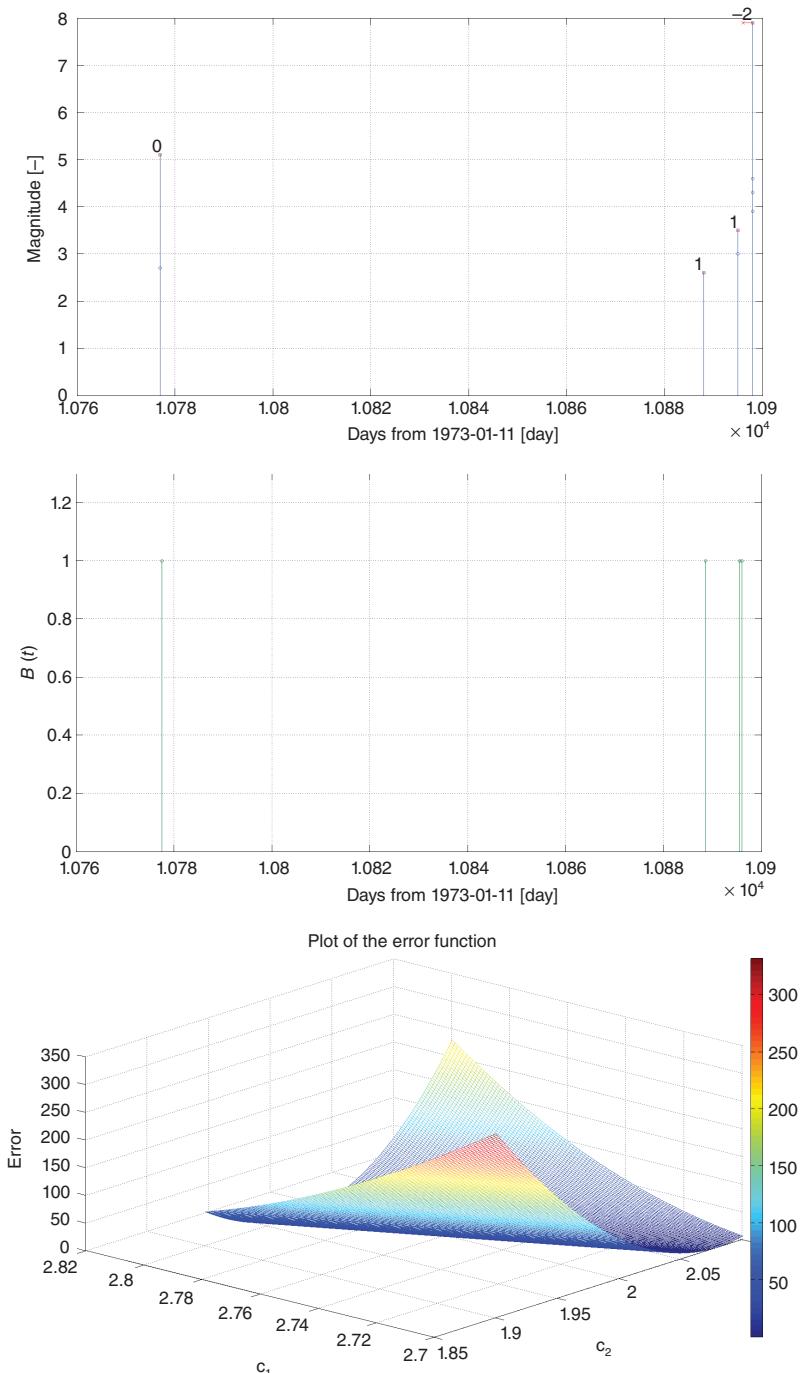
**FIGURE 11.4** Earthquake data and fitting with function (11.2).

- Latitude:  $-23.34^\circ \pm 0.2^\circ$
- Longitude:  $-70.30^\circ \pm 0.4^\circ$
- Time on the  $x$  axis is measured in days from 1973/01/11.



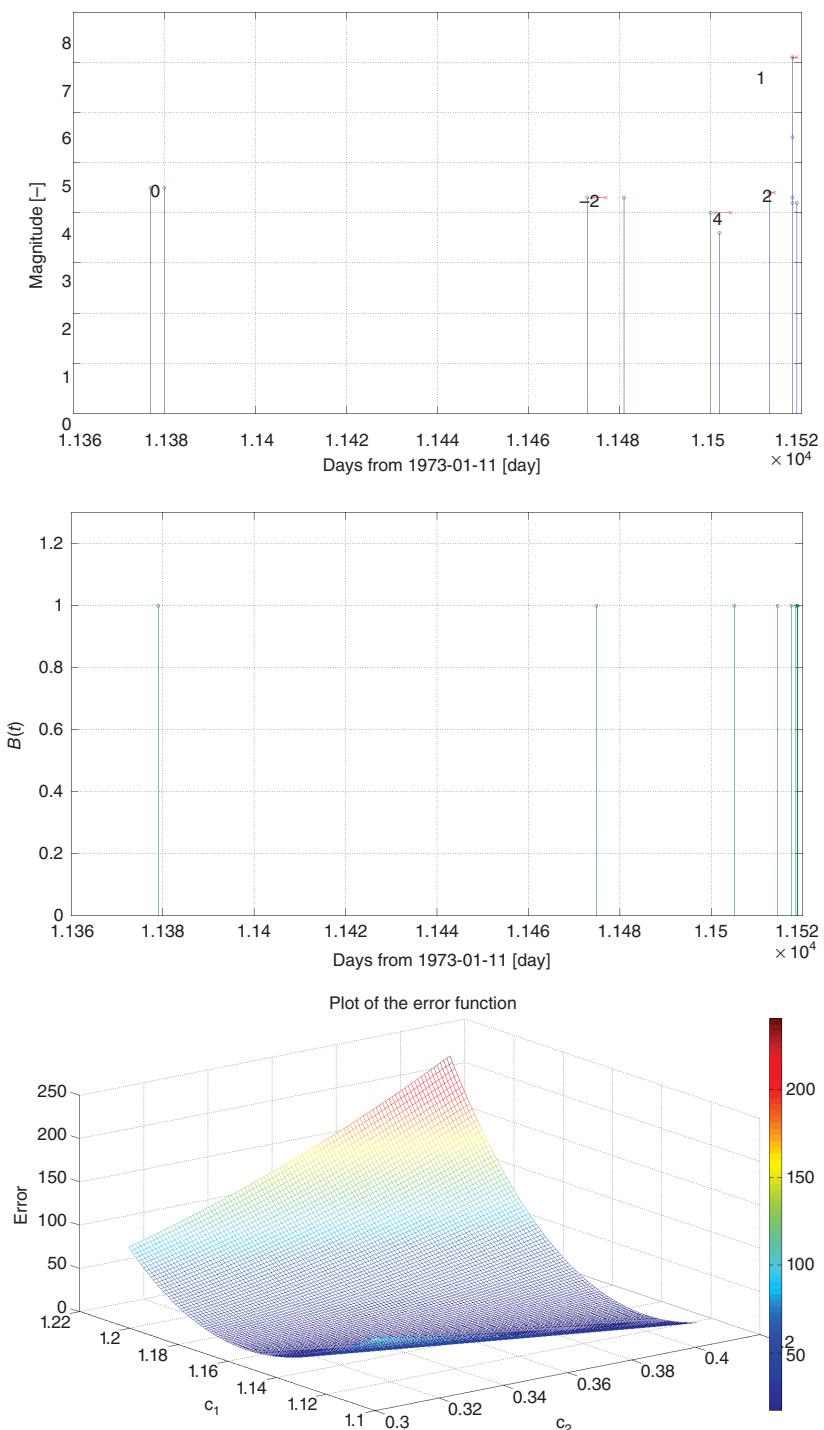
**FIGURE 11.5** Earthquake data and fitting with function (11.2).

- Latitude:  $16.78^\circ \pm 0.2^\circ$
- Longitude:  $-98.6^\circ \pm 0.4^\circ$
- Time on the x axis is measured in days from 1973/01/11.



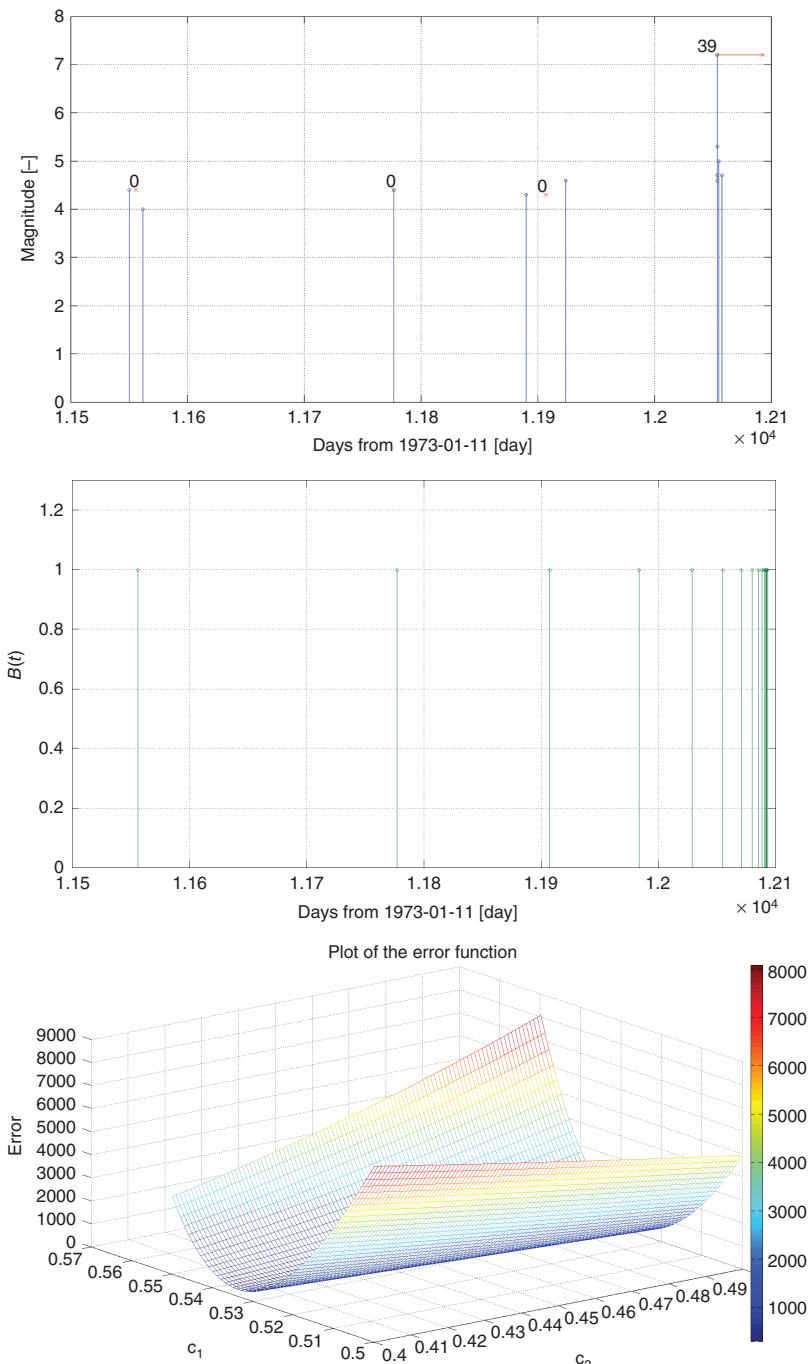
**FIGURE 11.6** Earthquake data and fitting with function (11.2).

- Latitude:  $63.52^\circ \pm 0.17^\circ$
- Longitude:  $-147.44^\circ \pm 0.34^\circ$
- Time on the  $x$  axis is measured in days from 1973/01/11.



**FIGURE 11.7** Earthquake data and fitting with function (11.2).

- Latitude:  $-17.66^\circ \pm 0.03^\circ$
- Longitude:  $-178.76^\circ \pm 0.06^\circ$
- Time on the x axis is measured in days from 1973/01/11.



**FIGURE 11.8** Earthquake data and fitting with function (11.2).

- Latitude:  $-19.93^\circ \pm 0.05^\circ$
- Longitude:  $-178.18^\circ \pm 0.1^\circ$
- Time on the  $x$  axis is measured in days from 1973/01/11.

**TABLE 11.1 Least square fitting with (11.2).**

Graph	$t_c$	Earthquake date	$C_1$
11.2	674	677	2.3008
11.3	7,019	7,054	0.7663
11.4	8,201	8,245	0.9858
11.5	8,298	8,291	0.9246
11.6	10,896	10,898	2.7636
11.7	11,519	11,518	1.1537
11.8	12,093	12,054	0.5306

crucial in the deterministic case. We fit the earthquake data, using a truncated Lévy flight process. We show that the fitted cumulative distribution curve has some outlining points. As observed in [9, 10], a market crash may be found as an outlining point in the cumulative probability distribution of the stochastic process described by the Lévy model. In the current work, we use a similar idea to obtain the critical time after which a *major* earthquake event may follow.

In the next section, we briefly describe the Truncated Lévy Flight distribution. Then, we approximate the observed data, using the appropriate truncated Lévy flight model. We locate the points in the data set that are outliers to the fitted Lévy flight in the cumulative probability distribution. If there are more than one such outliers, we only select the first such point. We then identify from the data set the time corresponding to this outlier. That is the estimated time ( $t_{\text{est}}$ ) after which the awareness level should be raised significantly that a major event will follow.

### 11.3.1 TRUNCATED LÉVY FLIGHT DISTRIBUTION

In this section, we briefly describe the theoretical setup for the computations done in the next section. For a review of stable distribution and for other definitions and notations, we refer to the Appendix.

The leptokurtic (heavy tails) property of the stable distributions for  $\alpha < 2$  is very desirable to model processes where extreme events have a greater (than normal) chance of appearing. Choosing to use Levy-Khintchine-type distributions seems like a natural idea; however, the infinite variance of these distributions for any  $\alpha < 2$  is an issue when working with real data. In order to avoid this problem, Mantegna [11] considers a Lévy-type distribution truncated at some parameter  $l$ , which obviously

has finite variance. This distribution was named the Truncated Lévy flight (henceforth *TLF*):

$$T(x) = cP(x)\mathbf{1}_{(-l,l)}(x)$$

where  $P(x)$  denotes any symmetric Lévy distribution obtained from its characteristic function in (11.14) when  $\beta = 0$ , and  $\mathbf{1}_A(x)$  is the indicator function of the set  $A$ .  $c$  is the constant that makes the definition of probability density. Clearly, as  $l \rightarrow \infty$ , one obtains the regular Lévy distribution characterized by  $P(\cdot)$  (a stable distribution). However, the *TLF* distribution itself is not stable for any finite truncation level  $l$ . Instead, this distribution has finite variance; thus, independent variables drawn from this distribution satisfy a regular central limit theorem. If the parameter  $l$  is large, the convergence to the limiting Gaussian may be very slow [11]. If the parameter  $l$  is small (so that the convergence is fast), the cut in its tails is very abrupt.

In order to have continuous tails, Koponen [12] considered a *TLF* in which the cut function is a decreasing exponential characterized by a separate parameter  $l$ . The characteristic function of this distribution, when  $\alpha \neq 1$ , can be expressed as follows:

$$\begin{aligned} \varphi(q) = \exp & \left\{ c_0 - c_1 \frac{(q^2 + 1/l^2)^{\frac{\alpha}{2}}}{\cos(\pi\alpha/2)} \cos(\alpha \arctan(l|q|)) \right. \\ & \left. \times (1 + il|q|\beta \tan(q \arctan l|q|)) \right\} \end{aligned}$$

with  $c_1$  a scale factor:

$$c_1 = \frac{2\pi \cos(\pi\alpha/2)}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} At$$

and

$$c_0 = \frac{l^{-\alpha}}{\cos(\pi\alpha/2)} c_1 = \frac{2\pi}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} Al^{-\alpha} t.$$

In the case of symmetric distributions  $\beta = 0$ , and in this case the variance can be calculated using the characteristic function:

$$\sigma^2(t) = - \left. \frac{\partial^2 \varphi(q)}{\partial q^2} \right|_{q=0} = t \frac{2A\pi(1-\alpha)}{\Gamma(\alpha) \sin(\pi\alpha)} l^{2-\alpha}$$

All the following discussion is in this symmetric case ( $\beta = 0$ ).

If we use time steps  $\Delta t$  apart, and  $T = N\Delta t$ , following the previous discussion, at the end of each interval, we must calculate the sum of

$N$  stochastic variables that are independent and identically distributed. Therefore, the characteristic function of the sum is:

$$\varphi(q, N) = \varphi(q)^N = \exp \left\{ c_0 N - c_1 \frac{N(q^2 + 1/l^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos(\alpha \arctan(l|q|)) \right\}.$$

The model can be improved by standardizing it. If the variance is given by:

$$\sigma^2 = - \left. \frac{\partial^2 \varphi(q)}{\partial q^2} \right|_{q=0}$$

we have that

$$-\left. \frac{\partial^2 \varphi(q/\sigma)}{\partial q^2} \right|_{q=0} = -\frac{1}{\sigma^2} \left. \frac{\partial^2 \varphi(q)}{\partial q^2} \right|_{q=0} = 1$$

Therefore, a standardized model is:

$$\begin{aligned} \ln \varphi_S(q) &= \ln \varphi \left( \frac{q}{\sigma} \right) = c_0 - c_1 \frac{((q/\sigma)^2 + 1/l^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos \left( \alpha \arctan \left( l \frac{|q|}{\sigma} \right) \right) \\ &= \frac{2\pi Al^{-\alpha} t}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} \left[ 1 - \left( \left( \frac{ql}{\sigma} \right)^2 + 1 \right)^{\alpha/2} \cos \left( \alpha \arctan \left( \frac{ql}{\sigma} \right) \right) \right]. \end{aligned}$$

Next we state a well-known theorem applicable in our case.

**Theorem 11.2.** Let  $\hat{p}(k)$  be the characteristic function (Fourier transform) of  $p(x)$ . Let  $\log \hat{p}(k) = f_a(k) + f_s(k)$ , where  $f_a(k)$  is analytic and  $f_s(k) \sim -a|k|^\alpha$  as  $k \rightarrow 0$  with  $a \in \mathbb{R}$  and finite. Then

$$p(x) \sim \frac{A}{|x|^{1+\alpha}}, \quad \text{as } x \rightarrow \infty,$$

where

$$A = \frac{a \sin(\alpha\pi/2)\Gamma(\alpha)\alpha}{\pi}.$$

Going back to the TLF distribution, for small  $q$ ,

$$\log \varphi_S(q) \sim \frac{2\pi Al^{-\alpha} t}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} \left[ 1 - \left( 1 + \frac{\alpha}{2} \left( \frac{ql}{\sigma} \right)^2 \right) \cos \left( \alpha \arctan \left( \frac{ql}{\sigma} \right) \right) \right].$$

Therefore, using the aforementioned theorem  $p(x) \sim \frac{1}{|x|^3}$ .

We can also justify this from the viewpoint of the cutoff function used in [12] given by:

$$f(x) = \begin{cases} A_- e^{-|x|/l} |x|^{-1-\alpha} & x < 0 \\ A_+ e^{-|x|/l} |x|^{-1-\alpha} & x \geq 0. \end{cases}$$

We are considering

$$p(x) = \begin{cases} f(x) & x > l \\ cL(x) & -l \leq x \leq l \\ f(x) & x < -l. \end{cases}$$

Here  $L(x)$  is the Lévy distribution.

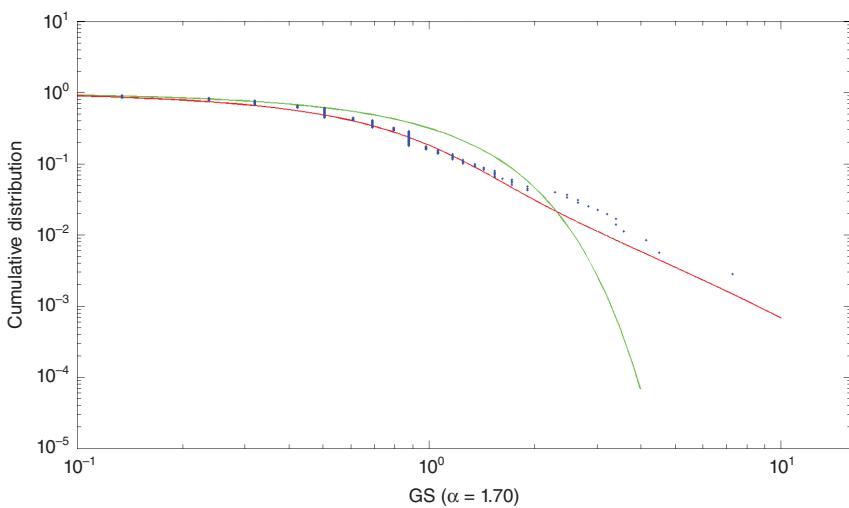
The decay is faster than the *heavy tail* decay due to the presence of the exponential function in  $f(x)$ . However, the main advantage of using this form of density  $p(x)$  is that we can keep the *heavy tail* of the Lévy as far as we want (by adjusting the parameter  $l$ ), and then have a sharp decay (faster than *heavy tail*). In other words, we are keeping the features of the heavy tail of the Lévy distribution as well as the finite variance feature given by the exponential decay.

To simulate the standardized truncated Lévy model, a Matlab module was developed. The parameter  $l$  is fixed at 1 and then the parameter  $A$  and the characteristic exponent  $\alpha$  are adjusted simultaneously in order to fit the cumulative function.

### 11.3.2 RESULTS

We use the model with TLF that we have already discussed in this chapter. We display the appropriate Lévy fitting in Figures 11.9, 11.10, 11.11, 11.12, and 11.13. The standardized model used is:

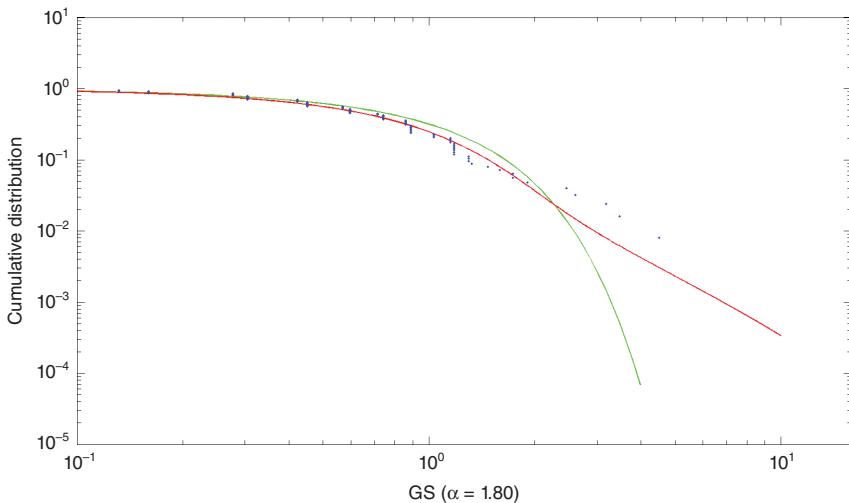
$$\begin{aligned} \ln \varphi_S(q) &= \ln \varphi\left(\frac{q}{\sigma}\right) = c_0 - c_1 \frac{((q/\sigma)^2 + 1/l^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos\left(\alpha \arctan\left(l \frac{|q|}{\sigma}\right)\right) \\ &= \frac{2\pi Al^{-\alpha} t}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} \left[ 1 - \left( \left(\frac{ql}{\sigma}\right)^2 + 1 \right)^{\alpha/2} \cos\left(\alpha \arctan\left(\frac{ql}{\sigma}\right)\right) \right]. \end{aligned} \quad (11.9)$$



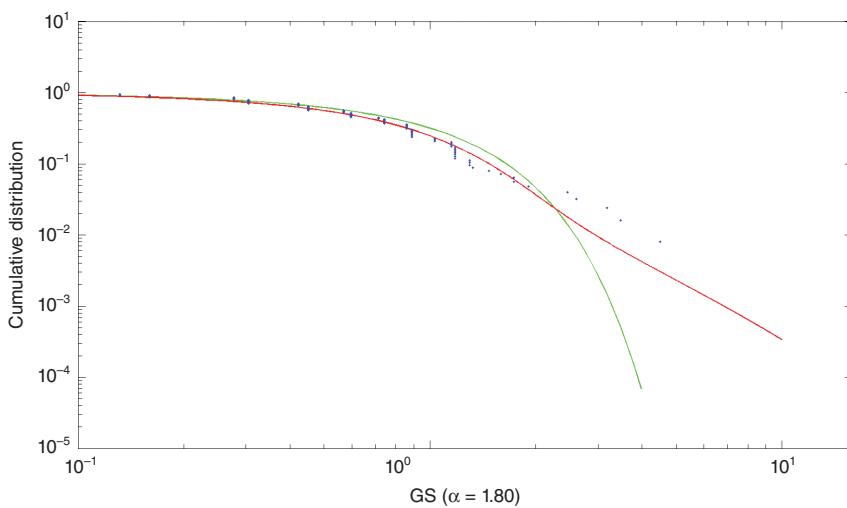
**FIGURE 11.9** Earthquake data fitting by TLF, corresponding to 11.3.

The recorded magnitude data is standardized:

$$GS = \left| \frac{M - E(M)}{\sigma(M)} \right|, \quad (11.10)$$



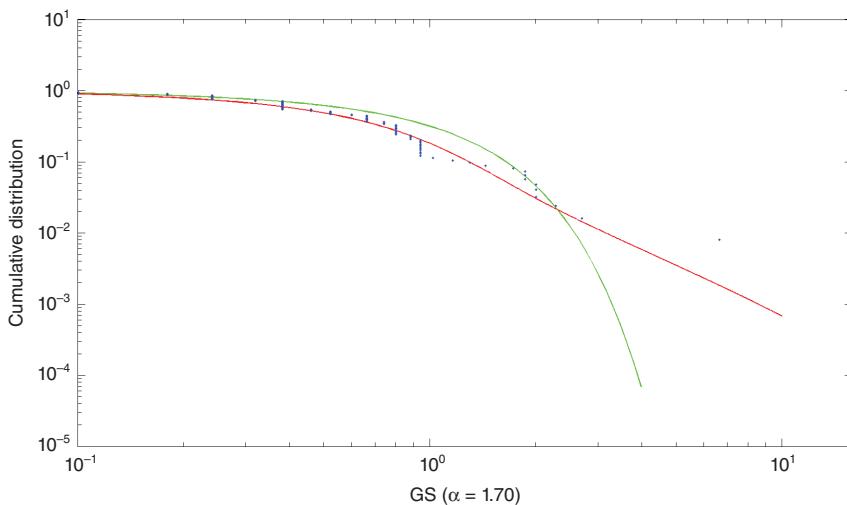
**FIGURE 11.10** Earthquake data fitting by TLF, corresponding to 11.4.



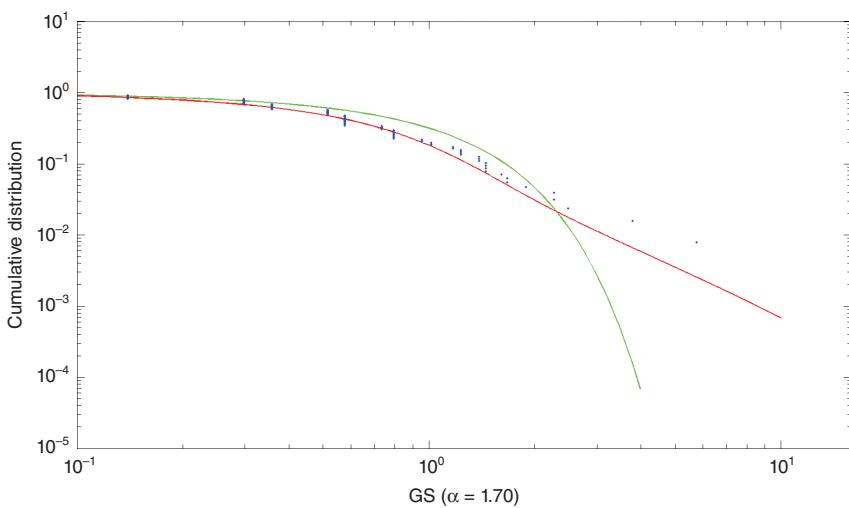
**FIGURE 11.11** Earthquake data fitting by TLF, corresponding to 11.5.

where  $M$  is magnitude of an earthquake, and  $E, \sigma$  are expected value and deviance for observed earthquake magnitude  $M$ .

In Figures 11.9, 11.10, 11.11, 11.12, and 11.13, we fit the observed standardized magnitude (11.10) with the cumulative distribution function corresponding to the characteristic function of the TLF given by (11.9).

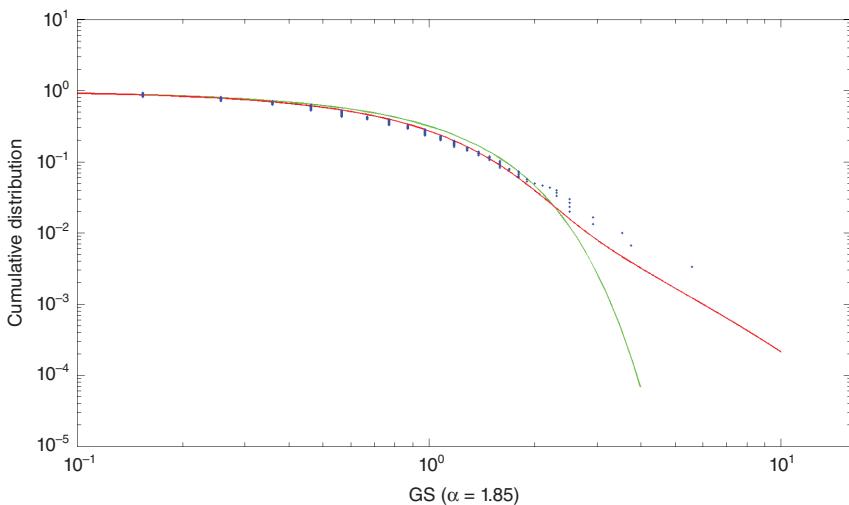


**FIGURE 11.12** Earthquake data fitting by TLF, corresponding to 11.6.



**FIGURE 11.13** Earthquake data fitting by TLF, corresponding to 11.7.

Since the empirical distribution is a *heavy tailed* distribution, it is natural to expect the standardized Lévy flight model (red line in the images) to fit better than a regular normal distribution (green line in the figures). However, recall that the TLF has exponentially decreasing tails (the variance is finite) unlike the regular Lévy distribution. Thus, we identify



**FIGURE 11.14** Earthquake data fitting by TLF, corresponding to 11.8.

**TABLE 11.2 Estimation from the Lévy model.**

Graph	$t_{\text{est}}$	Earthquake date
11.3	7,064	7,054
11.4	8,249	8,245
11.5	8,399	8,291
11.6	10,777	10,898
11.7	10,438	11,518
11.8	12,249	12,054

the outlying observations as the points at which the TLF fails to provide a good fit. The estimated time of the *crash* is reported as the date corresponding to the point where data deviates first from the fitted TLF curve.

We present the estimated crash time in Table 11.2. In the following figures, the *blue* dots represent the data points, the *red* line is the Lévy CDF, and the *green* line is the gaussian fit.

## 11.4 Application to the high-frequency market data

### 11.4.1 METHODOLOGY

In this section, we study high-frequency data corresponding to the collapse of the Bear Stearns in March 2008. The data used consists of the week (five trading days) March 10–14, 2008, before the merging announcement over the weekend as well as the two following trading days March 17 and 18. On Friday, March 14, 2008, at about 9:14 a.m., JP Morgan Chase and the Federal Reserve Bank of New York announced an emergency loan to Bear Stearns (of about 29 billion, terms undisclosed) to prevent the firm from becoming insolvent. This bailout was declared to prevent the very likely crash of the market as a result of the fall of one of the biggest investment banks at the time. This measure proved to be insufficient to keep the firm alive and 2 days later on Sunday March 16, 2008, Bear Stearns signed a merger agreement with JP Morgan Chase essentially selling the company for \$2 a share (price revised on March 24 to \$10/share). The same stock traded at \$172 in January 2007 and \$93 a share in February 2007. Today, this collapse is viewed as the first sign of the risk management meltdown

of investment bank industry in September 2008 and the subsequent global financial crisis and recession.

If we take a closer look at (11.4) and do the following transformation  $t \rightarrow t_c - t$ , where  $t_c$  is a crash date.

$$f^F(t_c - t) = \beta e^{\alpha F(\log_a(t_c - t))} \quad (11.11)$$

we can observe that our function has log-periodic structure similar to the (11.3). At this point, we do not know  $t_c$ . The distance between two periods in (11.4) will be:

$$\begin{aligned} \log_a(t_N) - \log_a(t_{N+1}) &= 1 \\ \ln(t_N) - \ln(t_{N+1}) &= \ln(a) \\ t_{N+1} &= t_N / \ln(a) \end{aligned}$$

From this, we can see that for fixed  $t_0$ , then  $N$ th period will occur at time

$$\ln t_N = N \cdot (-\ln(a)) + \ln(t_0) \quad (11.12)$$

This is exactly what we got in (11.2). If we assume that the periods of (11.11) match market periods, we can do least square fitting and estimate coefficient  $a$ . Moreover, we can estimate crash date  $t_c$  by using same technique as in the case of (11.2). This uses data only prior to the market crash. By another least square fitting, we can estimate coefficients  $\alpha, \beta$ . With estimated coefficients  $\alpha, \beta, a$  we can do another least square fitting for (11.5) with argument  $t_c - t$  for the best fit of the data.

### 11.4.2 RESULTS

In this part, we present data fitted with equation (11.5) with argument  $t_c - t$ .

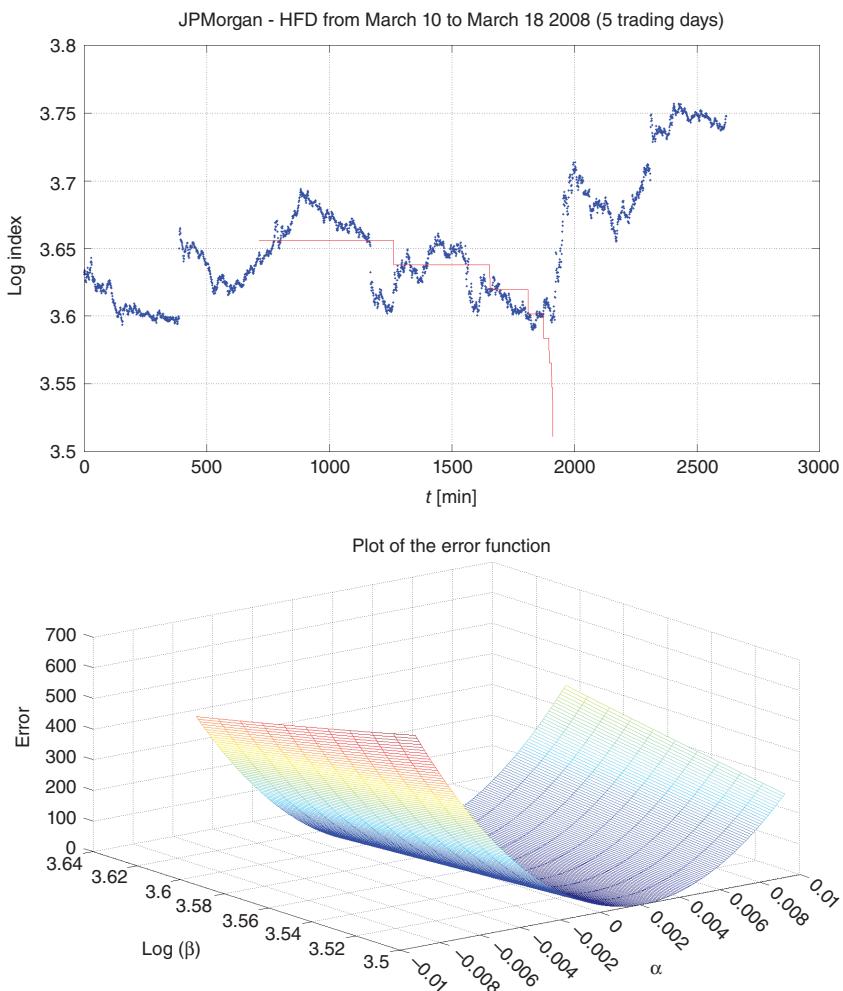
$$f_x(t_c - t) = \beta e^{\alpha(F(\log_a(t_c - t) - x) + x)}$$

We will show examples of four indices that experienced biggest change in price during March 10 and 18.

For parameters  $\alpha, \beta$ , and its least square fitting, we constructed the following error function:

$$e(\alpha, \beta, \vec{t}) = \sum_{i=1}^k (\ln(p(t_i)) - \ln(f^F(t_i)))^2 \quad (11.13)$$

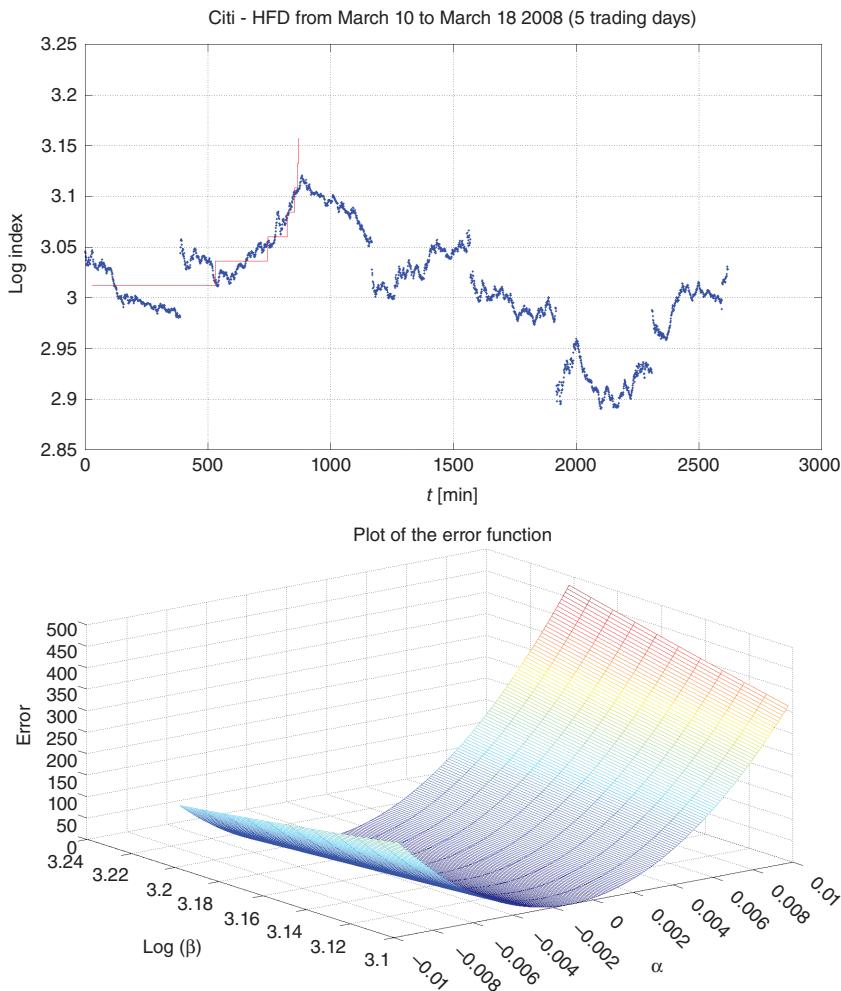
where  $\vec{t} = (t_1, \dots, t_k)$  is the time where we want  $f^F$  to touch the plot of index price  $p(t)$ . We get coefficients  $\alpha, \beta$  by minimizing  $e$ .



**FIGURE 11.15 JP Morgan Chase—The solid line represents the best fit with (11.5).**

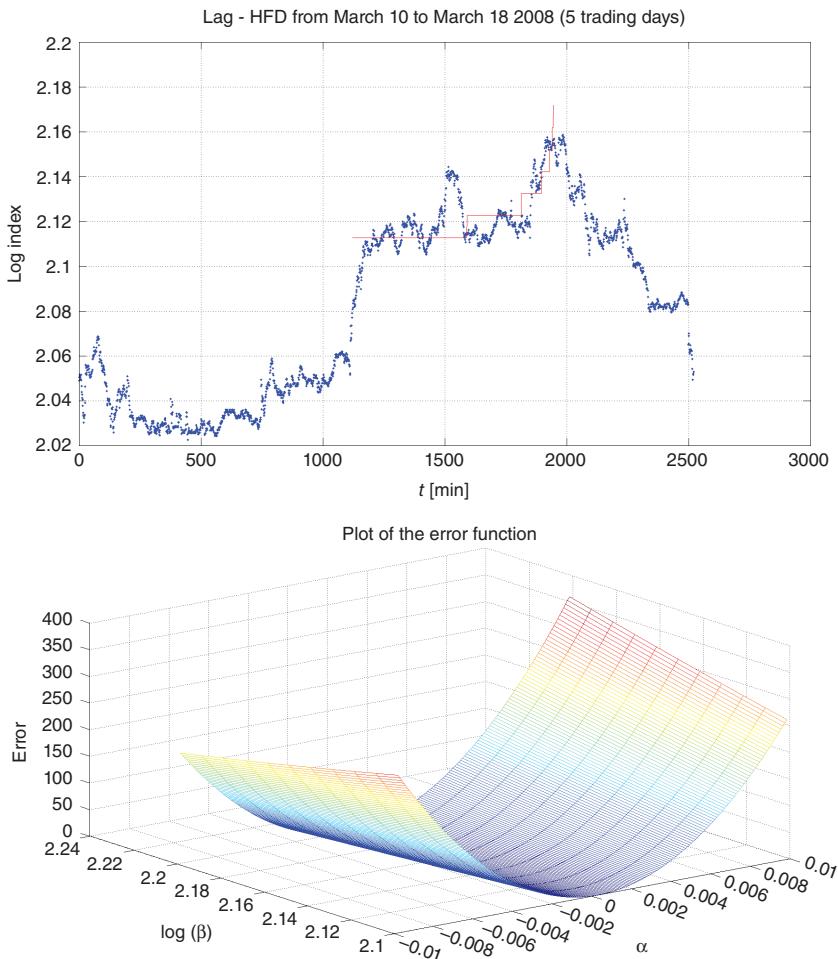
## 11.5 Brief program code description

All programs for Sections 11.2.1 and 11.4 were written in MATLAB. Let us introduce few crucial steps to clarify fitting process and whole procedure. In the first step, we need to take “measured” data and compute periods. In geophysics scenario, this would correspond to the times between two earthquakes (preceding some major even), in market scenario time between two minima or maxima of the stock price. Since periods in



**FIGURE 11.16** Citi—The solid line represents the best fit with (11.5).

(11.5) and (11.2) are invariant under transformation  $t \rightarrow (t_c - t)$  (except its number). If we denote measured periods as  $\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_n$  and since (11.7), we can estimate coefficients  $C_1, C_2$  by least square fitting, that is, minimizing equation (11.6). The critical time, cumulative point, can be found using (11.8). The connection to the market scenario is made through (11.12) and thus estimation of  $t_c$  will follow the same path. For high-frequency scenario, we now already know  $\alpha$  and  $t_c$ , and we need to estimate parameters  $\alpha, \beta$ . This can be done by minimizing error function (11.13). You can see that this is done only for “points of interest”; that is, we will minimize

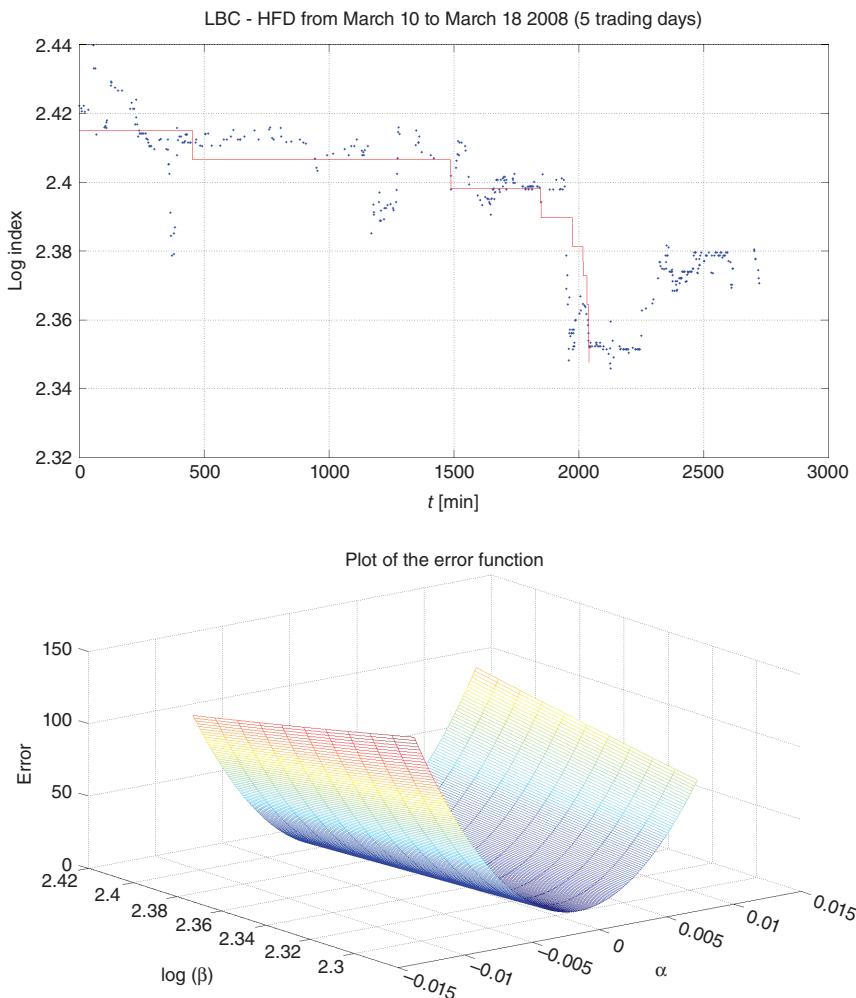


**FIGURE 11.17** IAG—The solid line represents the best fit with (11.5).

in such a way that final function will touch the graph for market minima. At last, we will estimate parameter  $x$  in (11.5), using all available data, again using least square fitting for  $x$ . We will minimize the error function  $e(x) = \sum_t (p(t) - f_x(t_c - t))^2$  with parameters  $t_c, a, \alpha, \beta$  obtained previously.

## 11.6 Conclusion

In Section 11.2.1 and Figures 11.2, 11.3, 11.4, 11.5, 11.6, 11.7, and 11.8, we use the scale invariant equation (11.2) to estimate the crash date of a



**FIGURE 11.18 LBC—The solid line represents the best fit with (11.5).**

following major event. The least square fitting methodology helps estimating the crash date  $t_c$ . We note that the data used when fitting include all events except the major crash. As such, the predicted time is typically but not always preceding the major event. This methodology could be used in real time by looking at the minor earthquakes since a major event and using the time intervals to make better and better predictions as new data (new minor events) are accumulating.

The second method takes all earthquake data surrounding a major event. By picking the date of the first outlying observation, we are able

**TABLE 11.3 Comparison.**

Graph	$t_c$	$t_{\text{est}}$	Earthquake date
11.3	7,019	7,064	7,054
11.4	8,201	8,249	8,245
11.5	8,298	8,399	8,291
11.6	10,896	10,777	10,898
11.7	11,519	10,438	11,518
11.8	12,093	12,249	12,054

to make a prediction about the event time. Clearly since we use events surrounding the major one and we pick the date of the first (smallest magnitude) of the outlying observations, sometimes the predicted date is after the major event. This should not be a problem if the methodology is run in real time. As new magnitude earthquakes are recorded, the TLF distribution is fitted and at the date when the fit is not good anymore, an earthquake warning should be issued.

We conclude by comparing the estimated *crash* dates, using the two methods with the actual crash date in Table 11.3.

Same situation holds for the data in financial market. In [9, 10], the financial data has been analyzed using the Lévy model to estimate the *crash date*. In this chapter, we used the scale invariance method to analyze the same. The closeness of the two results shows that both methods are compatible.

## 11.A Appendix

### 11.A.1 STABLE DISTRIBUTIONS

We present a brief introduction of stable distributions.

Consider the sum of  $n$  independent identically distributed (i.i.d.) random variables  $X_i$ ,

$$X(n\Delta t) = X_1 + X_2 + X_3 + \cdots + X_n.$$

Since the variables are independent, the distribution of their sum may be obtained as the  $n$ -fold convolution,

$$P[X(n\Delta t)] = P(X_1) \otimes P(X_2) \dots \otimes P(X_n).$$

The original distribution of the  $X_i$ 's is called stable if the functional form of  $P[X(n\Delta t)]$  is the same as the functional form of  $P[X(\Delta t)]$ . Specifically, for any  $n \geq 2$ , there exists a positive  $C_n$  and a  $D_n$  so that:

$$P[X(n\Delta t)] = P[C_n X + D_n]$$

where  $X$  has the same functional form of the distribution as  $X_i$  for  $i = 1, 2, \dots, n$ . If  $D_n = 0$ ,  $X$  is said to have a strictly stable distribution. It can be shown (see [13]) that

$$C_n = n^{\frac{1}{\alpha}}$$

for some parameter  $\alpha$ ,  $0 < \alpha \leq 2$ .

Stability is a very desirable property of a distribution. Many stochastic processes can be defined in such a way that their increments are independent. Specifically,

$$X(t) = X(0) + \left( X\left(\frac{t}{n}\right) - X(0) \right) + \cdots + \left( X(t) - X\left(\frac{n-1}{n}t\right) \right),$$

and if the increments are all distributed according to a stable distribution, then the process  $X(t)$  itself will have a stable distribution, regardless of the value of  $n$ .

## 11.A.2 CHARACTERIZATION OF STABLE DISTRIBUTIONS

Lévy [14] and Khintchine [15] found the most general form of the stable distributions. The general representation is through the characteristic function  $\varphi(q)$  associated with the distribution. The most common parametrization is:

$$\ln(\varphi(q)) = \begin{cases} i\mu q - \gamma^\alpha |q|^\alpha \left[ 1 - i\beta \frac{q}{|q|} \tan\left(\frac{\pi\alpha}{2}\right) \right] & \text{if } \alpha \in (0, 2] \setminus \{1\} \\ i\mu q - \gamma |q| \left[ 1 + i\beta \frac{q}{|q|} \frac{2}{\pi} \ln |q| \right] & \text{if } \alpha = 1 \end{cases}, \quad (11.14)$$

where  $\alpha$  is called the index of stability (tail index, tail exponent, or characteristic exponent),  $\gamma$  is a positive scale factor,  $\mu$  is a location parameter, and  $\beta \in [-1, 1]$  is a skewness (asymmetry) parameter.

It is very easy to understand why this form of characteristic function determines the stable distributions. The characteristic function of a convolution of independent random variables is simply the product of the

respective characteristic functions and, thus owing to the special exponential form in (11.14) all stable distributions, are closed under convolutions for a fixed value of  $\alpha$  and time increments of same size. All such distributions are heavy tailed (leptokurtic) and in fact have no second moments for any  $\alpha < 2$ . When  $\alpha < 1$  the distribution does not even have the first moment. Please note that neither third nor fourth moment exists for these distributions so the usual measures of skewness and kurtosis are undefined.

The characteristic function always exists and thus the definition given above is proper; however, finding the associated density function is not always possible (it may not exist). Typically, one of the inversion formulae is used, but the analytical form of the stable distributions is known only for a few values of  $\alpha$  and  $\beta$ :

- When  $\alpha = 2$ , one obtains the Gaussian (normal) distribution with mean  $\mu$  and variance  $2\gamma^2$  (the skewness parameter  $\beta$  has no effect).

$$f(x) = \frac{1}{\sqrt{4\pi}\gamma} e^{-\frac{(x-\mu)^2}{4\gamma^2}}.$$

- When  $\alpha = \frac{1}{2}$ ,  $\beta = 1$ , the distribution is called the Lévy–Smirnov distribution.

$$f(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{e^{-\gamma/2(x-\mu)}}{(x-\mu)^{\frac{3}{2}}}, \text{ if } x \geq \mu$$

- When  $\alpha = 1$  and  $\beta = 0$ , we obtain the Cauchy (Lorentz) distribution

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - \mu)^2}.$$

## References

- A. Johansen, O. Ledoit, D. Sornette, Crashes as critical points, *International Journal of Theoretical Applied Finance*, January 2000, Volume 3, No. 1, pp. 219–255.
- M.C. Mariani, Y. Liu, A new analysis of the effect of the Asian crisis of 1997 on emergent markets, *Physica A: Statistical Mechanics and Its Applications*, July 1, 2007, Volume 380, pp. 307–316.
- M.C. Mariani, P. Bezdek, L. Serpa, I. Florescu, Ising type models applied to geophysics and high frequency market data, *Physica A*, November 1, 2011, Volume 390, Issues 23–24, pp. 4396–4402.

4. R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, 1999.
5. A. Johansen, D. Sornette, Large financial crashes, *Physica A*, 1997, Volume 245 (N3-4), pp. 411–422.
6. M.C. Mariani, Y. Liu, A new analysis of intermittence, scale invariance and characteristic scales applied to the behavior of financial indices near a crash, *Physica A*, July 15, 2006, Volume 367, pp. 345–352.
7. M. Ferraro, N. Furman, Y. Liu, M.C. Mariani, D. Rial, Analysis of intermittence, scale invariance and characteristic scales in the behavior of major indices near a crash, *Physica A*, January 1, 2006, Volume 359, pp. 576–588.
8. The USGS Earthquake Magnitude Working Group, USGS Earthquake Magnitude Policy, *USGS*. March 29, 2010. Web November 26, 2010. [http://earthquake.usgs.gov/aboutus/docs/020204mag\\_policy.php](http://earthquake.usgs.gov/aboutus/docs/020204mag_policy.php).
9. M.C. Mariani, Y. Liu, Normalized truncated Levy walks applied to the study of financial indices, *Physica A*, 2007, Volume 377, pp. 590–598
10. E. Barany, M.P. Beccar Varela, I. Florescu, I. SenGupta, Detecting market crashes by analysing long-memory effects using high-frequency data, *Quantitative Finance*, 2012, Volume 12, Issue 4, pp. 623–634.
11. R.N. Mantegna, H.E. Stanley, *Physical Review Letters*, 1994, Volume 73, pp. 2946.
12. I. Koponen, *Physical Review*, 1995, Volume E 52, pp. 1197.
13. G. Samorodnitsky, M.S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York, 1994.
14. P. Lévy, *Calcul des probabilités*, Gauthier-Villars, Paris, 1925.
15. A. Ya. Khintchine, P. Lévy, Sur les lois stables, *Comptes Rendus de l'Academie des Sciences Paris*, 1936, Volume 202, pp. 374.
16. F. Black, M. Scholes, The pricing of options and corporate liabilities, *Journal of Political Economics*, 1973, Volume 81, pp. 637–659, 1973.
17. D. Stauffer, Social applications of two-dimensional Ising models, *American Journal of Physics*, April 2008, Volume 76, Issue 4, pp. 470.

## Chapter Twelve

# Analysis of Generic Diversity in the Fossil Record, Earthquake Series, and High-Frequency Financial Data

M. P. Beccar Varela<sup>1</sup>, F. Biney<sup>1</sup>, Maria C. Mariani<sup>1</sup>,  
I. SenGupta<sup>2</sup>, M. Shpak<sup>3</sup>, and P. Bezdek<sup>4</sup>

<sup>1</sup>*Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX, USA*

<sup>2</sup>*Department of Mathematics, North Dakota State University, Fargo, ND, USA*

<sup>3</sup>*NeuroTexas Institute, St. David's Medical Center, Austin, TX, USA*

<sup>4</sup>*Department of Mathematics, The University of Utah, Salt Lake City, UT, USA*

### 12.1 Introduction

In order to develop an understanding of the patterns and processes in the history of plant and animal life on Earth, paleontologists have attempted to quantify and analyze the diversity of taxonomic diversity through geological time (see Sepkoski, 1977, 1979 [1], and the references therein). Changes in biological diversity are driven by the origination of new taxa and by the extinction of previously extant taxa. In describing extinction dynamics, several studies have noted a qualitative difference between “background” (low mean and variance) and mass extinction (high mean and variance)

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

events. Similarly, one can speak of background and mass origination events, where the latter would correspond to large-scale adaptive radiations characterized by the rapid origin of large numbers of species and higher taxa, while the former refers to origination rates during more ecologically “normal” periods.

In this chapter, we analyze the paleontological database compiled by Sepkoski [2] on the origination and extinction of fossil genera during the Phanerozoic (a time interval that spans from shortly before the Cambrian period, approximately 540 million years ago to the present). The principal goals of this analysis are to (a) determine what class of probability distributions best characterizes the magnitude of extinction events, particularly when dealing with extreme values induced by mass extinctions, and (b) determine what class of distributions provides the most accurate representation of the overall distribution of generic diversity in geological time.

We consider the following variables when describing the temporal dynamics of generic diversity in the fossil record: following [3], where *age* is the time before present (in epochs and stages measured in millions of years, Ma), *diversity* is the number of individuals at time *age*, *orig* is the number of genera born at time *age*, and *ext* is the number of genera that became extinct at time *age*.

This study endeavors to characterize both the distribution of extinction magnitudes and the overall distribution of generic diversity.

We begin by presenting some theoretical results on convolutions of exponential waiting times, as this simple model offers a straightforward starting point for analyzing origination and extinction data. The resulting distributions will be used for comparison with the data sets. We also analyze high-frequency financial data by using Lévy models.

With the availability of high-frequency data (HFD) for financial market analysis, there has been an increase in the studies dealing with the persistence of shocks in both the mean and variance of financial instruments return. Several studies report evidence of persistence (long-memory) behavior in squared returns or empirical volatilities; see Breidt et al. [4], Robinson [5], Shephard [6], Lobato and Savin [7], and Baillie [8]. Similar features have been observed in data from other fields such as physics and geophysics. In physics, the presence of strong autocorrelation in the squares of differences in velocity of the mean wind direction has been explored by Barndorff-Nielsen and Shephard [9].

The pioneering work of Box et al. [10] in the area of autoregressive moving average models paved the way for related work in the area of

volatility modeling with the introduction of ARCH and then GARCH models by Engle [11] and Bollerslev [12], respectively. In terms of the statistical framework, these models provide motion dynamics for the dependency in the conditional time variation of the distributional parameters of the mean and variance, in an attempt to capture such phenomena as autocorrelation in returns and squared returns. Extensions to these models have included more sophisticated dynamics such as the threshold model (TGARCH) [13] to capture the asymmetry in the news impact, the NGARCH model [14], the EGARCH models [15], the stochastic volatility models [16], the FIGARCH and FIEGARCH models [17], and the long-memory generalized autoregressive conditionally heteroskedastic (LMGARCH) models [5, 18] as well as distributions other than the normal to account for the skewness and excess kurtosis observed in practice such as Student's *t test*, Generalized error, Generalized Hyperbolic, the Normal inverse Gaussian distribution, and Johnson's SU distribution.

In the literature, the most popular GARCH model is the GARCH(1,1), where the persistence parameter is less than 1 to ensure covariance stationarity. It turns out that on modeling using GARCH, most often the persistence parameter is approximately 1 but yet the model does not adequately capture the persistence in volatility. This fact motivated the introduction of the integrated GARCH (IGARCH) model where Bollerslev and Engel [19] allowed for unit persistence in the GARCH model; that is, the persistence was set to 1. The IGARCH model has some structural complication in the sense that its unconditional variance does not exist. Baillie et al. [8] extended the IGARCH to the fractional IGARCH (FIGARCH) by allowing for high persistence (long memory) directly in the conditional variance while avoiding the complications of IGARCH; that is, they allowed the integration coefficient to vary between [0,1].

Our main interest is to investigate the underlying volatility process in earthquake series, explosive series, high-frequency financial data, and financial indices and examines the applicability of a range of GARCH specifications for modeling volatility of these series to identify similarities and differences in the volatility structure.

## 12.2 Statistical preliminaries and results

In this section, we present some theoretical results, which will be used later in this chapter for the analysis of the data.

**Theorem 12.1.** If  $X_i$  are independent and identically distributed exponential random variables with rate parameter  $\lambda$

$$X_i \sim \text{Exponential}(1/\lambda)$$

then

$$Y = \sum_{i=1}^n X_i \sim \text{Gamma}(n, 1/\lambda).$$

Sum of  $X_i$  has Gamma distribution with shape parameter  $n$  and scale parameter  $1/\lambda$ .

As a result, if we assume that generic extinction time is being exponentially distributed with mean  $1/\lambda$ , then by Theorem 12.1, the waiting time for the extinction of an ensemble of genera will approximately follow a gamma distribution.

### 12.2.1 SUM OF EXPONENTIAL RANDOM VARIABLES WITH DIFFERENT PARAMETERS

Suppose  $X$  and  $Y$  follow exponential distributions with parameters  $\beta_1, \beta_2$ . Let  $X$  has probability density function  $f_X(x) = \frac{e^{-x/\beta_1}}{\beta_1}$  and  $Y$  has p.d.f.  $f_Y(y) = \frac{e^{-y/\beta_2}}{\beta_2}$ , where  $X$  and  $Y$  are independent. Applying the change of variables  $u = x + y$  and  $v = x$ , we obtain the Jacobian of this transformation is 1. Thus, the joint pdf in new variables is given by

$$f(u, v) = \frac{1}{\beta_1 \beta_2} \exp\left(-v\left(\frac{1}{\beta_1} - \frac{1}{\beta_2}\right)\right) \exp\left(-\frac{u}{\beta_2}\right).$$

Therefore, when  $\beta_1 \neq \beta_2$ ,

$$f_U(u) = \int_0^u f(u, v) dv = \frac{1}{\beta_1 - \beta_2} \left[ e^{-u/\beta_1} - e^{-u/\beta_2} \right].$$

When  $\beta_1 = \beta_2$ ,

$$f_U(u) = \frac{ue^{-u/\beta_1}}{\beta_1^2} = \Gamma(2, \beta_1).$$

So  $u \sim \Gamma(2, \beta_1)$ . Similarly with random variables  $X, Y, Z$  with probability density functions  $f_X(x) = \frac{e^{-x/\beta_1}}{\beta_1}$ ,  $f_Y(y) = \frac{e^{-y/\beta_2}}{\beta_2}$ , and  $f_Z(z) = \frac{e^{-z/\beta_3}}{\beta_3}$ , then using a change of variable

$$u = x + y + z, \quad v = x + y, \quad w = x,$$

we arrive to

$$f(u, v) = \frac{e^{-u/\beta_3}}{\beta_3(\beta_1 - \beta_2)} [\exp(-v(1/\beta_1 - 1/\beta_3)) - \exp(-v(1/\beta_2 - 1/\beta_3))],$$

so that

$$\begin{aligned} \int_0^u f(u, v) dv &= \frac{\beta_1}{\beta_1 - \beta_2} \left[ \frac{e^{-u/\beta_1} - e^{-u/\beta_3}}{\beta_1 - \beta_3} \right] \\ &\quad - \frac{\beta_2}{\beta_1 - \beta_2} \left[ \frac{e^{-u/\beta_2} - e^{-u/\beta_3}}{\beta_2 - \beta_3} \right] = f_U(u), \end{aligned}$$

and if  $\beta_1 = \beta_2 = \beta_3$ ,  $f(u, v) = \frac{1}{\beta^3} e^{-u/\beta} v$ , then

$$f_U(u) = \int_0^u f(u, v) dv = \frac{1}{\beta^3} e^{-u/\beta} \frac{u^2}{2} = \Gamma(3, \beta).$$

Generally, if we have  $n$  random variables  $X_i$  with probability density functions  $f_{X_i}(x)$ , then the probability density function of  $X = \sum_{i=1}^n X_i$  will have the probability density function given by a convolution of probability density functions [20].

$$f_X(x) = (f_{X_1} * (f_{X_2} * (f_{X_3} * \dots)))[x]. \quad (12.1)$$

In the case of  $n$  independent variables following p.d.f.  $f(x) = \frac{e^{-x/\beta}}{\beta}$  with different coefficient  $\beta$ , it is more convenient to work with the *characteristic function*, which is just a Fourier transform of a probability density function. Convolution theorem states that Fourier transform of convolution of two functions is the product of their Fourier transform.

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\}$$

thus if we use this theorem on (12.1), we get

$$\mathcal{F}\{f_X\} = \mathcal{F}\{(f_{X_1} * (f_{X_2} * (f_{X_3} * \dots)))\} = \prod_{i=1}^n \mathcal{F}\{f_{X_i}\}$$

but  $\mathcal{F}\{f_{X_i}\}$  is just characteristic function of  $X_i$  denoted by  $\varphi_{X_i}$ . We showed that

$$\varphi_X = \prod_{i=1}^n \varphi_{X_i}. \quad (12.2)$$

In other words, characteristic function of sum of independent random variables is product of their characteristic functions.

### 12.2.1.1 General version for $n$ random variables

In this subsection, we assume the following parametrization, using scale parameter  $\beta$ , of the exponential random variable probability density function.

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, x \geq 0 \quad (12.3)$$

The characteristic function for this random variable is [20]

$$\varphi_X(t) = (1 - it\beta)^{-1}.$$

We also know, using (12.2), that sum of independent identically distributed exponential random variables will have characteristic function

$$\varphi_{\sum X_k}(t) = \prod_{k=1}^n \varphi_{X_k}(t) = (1 - it\beta)^{-n} \quad (12.4)$$

and this defines the gamma distribution by Theorem 12.1.

Similarly, the sum of independent exponential random variables  $X_1, \dots, X_n$  with scale parameters  $\beta_1, \dots, \beta_n$  will have the characteristic function

$$\varphi_{\sum X_k}(t) = \prod_{k=1}^n \varphi_{X_k}(t) = \prod_{k=1}^n (1 - it\beta_i)^{-1} \quad (12.5)$$

**Definition 12.1.** (Hypoexponential distribution) [21, Ch. 7.6.3.] We say that the random variable  $X$  is hypoexponentially distributed if its characteristic function is of form (12.5). Sometimes it is also called a generalized Erlang distribution.

We will show that the sum of exponential random variables with different parameters  $\beta_k$  (12.5), where  $\beta_k$  differs by a small variation from some  $\bar{\beta}$ , can be reasonably approximated by the gamma distribution (12.4).

We assume the following scale parameters  $\beta_k$ :

$$\begin{aligned} \beta_k &= \bar{\beta} + \Delta\beta_k, k \in \{1, \dots, n\} \\ \bar{\beta} &= \frac{1}{n} \sum_{j=1}^n \beta_j \\ 0 &= \sum_{k=0}^n \Delta\beta_k \end{aligned} \quad (12.6)$$

where  $\Delta\beta_k$  is a small perturbation from the mean value  $\bar{\beta}$ . If we substitute  $\beta_k$  into (12.3), we get

$$\varphi_{X_k} = (1 - it\bar{\beta} - it\Delta\beta_k)^{-1}. \quad (12.7)$$

This function is holomorphic in  $\Delta\beta_k$  for small values of  $\Delta\beta_k$ ; thus, it has power series expansion in  $\Delta\beta_k$  at zero.

$$\varphi_{X_k} = (1 - it\bar{\beta})^{-1} + it(1 - it\bar{\beta})^{-2}\Delta\beta_k + \mathcal{O}(\Delta\beta_k^2) \quad (12.8)$$

Using this result and substituting (12.8) into (12.5), we get

$$\begin{aligned} \varphi_{\sum X_k}(t) &= \left( (1 - it\bar{\beta})^{-1} + it(1 - it\bar{\beta})^{-2}\Delta\beta_1 + \mathcal{O}(\Delta\beta_1^2) \right) \cdot \\ &\quad \left( (1 - it\bar{\beta})^{-1} + it(1 - it\bar{\beta})^{-2}\Delta\beta_2 + \mathcal{O}(\Delta\beta_2^2) \right) \cdot \\ &\quad \left( (1 - it\bar{\beta})^{-1} + it(1 - it\bar{\beta})^{-2}\Delta\beta_3 + \mathcal{O}(\Delta\beta_3^2) \right) \cdot \\ &\quad \left( (1 - it\bar{\beta})^{-1} + it(1 - it\bar{\beta})^{-2}\Delta\beta_4 + \mathcal{O}(\Delta\beta_4^2) \right) \cdot \\ &\quad \vdots \\ &\quad \left( (1 - it\bar{\beta})^{-1} + it(1 - it\bar{\beta})^{-2}\Delta\beta_n + \mathcal{O}(\Delta\beta_n^2) \right) \end{aligned}$$

multiplying this expression gives us

$$\varphi_{\sum X_k}(t) = (1 - it\bar{\beta})^{-n} + it(1 - it\bar{\beta})^{-(n+1)} \sum_{k=1}^n \Delta\beta_k + \mathcal{O}(\Delta\beta_k^2) \quad (12.9)$$

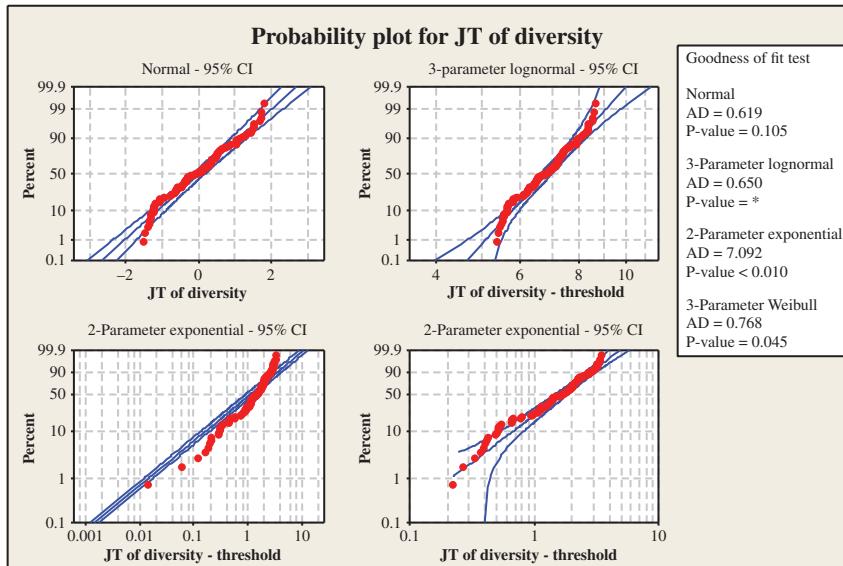
and from  $\sum_{k=0}^n \Delta\beta_k = 0$  from (12.6), we finally get

$$\varphi_{\sum X_k}(t) = (1 - it\bar{\beta})^{-n} + \mathcal{O}(\Delta\beta_k^2). \quad (12.10)$$

Therefore, the characteristic function (12.10) differs from that of the gamma distribution (12.4) by a term depending on a square of perturbation  $\Delta\beta_k$ . This justifies our fit using the gamma distribution with a scale parameter equal to the mean value  $\bar{\beta}$ .

## 12.3 Statistical and numerical analysis

The first analysis involves constructing a frequency distribution of the number of genera over all geological stages. At first, we transform the



**FIGURE 12.1** Plot of the frequency of the number of genera compared with different distribution (with  $p$ -values computed from AD test indicating goodness of fit).

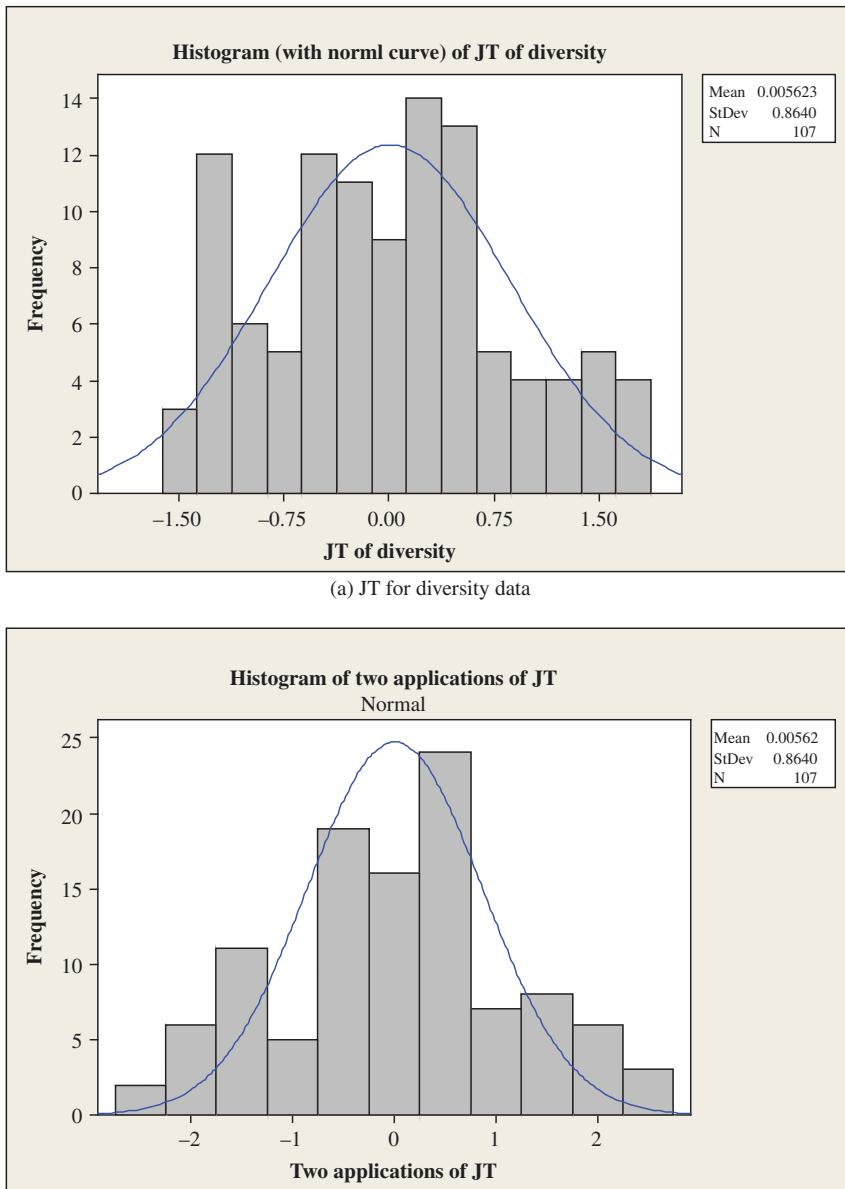
diversity data set, using the Johnson transformation (JT) function. The transformation is applied because the raw data is nonnormal, principally as a consequence of the net increase in the number of taxa (both actual and as a sampling artifact of the “pull of the recent”) over recent intervals. The transformed function becomes

$$f(x) = -0.292746 + 0.616307 \sinh^{-1}((x - 1080.08)/297.129)$$

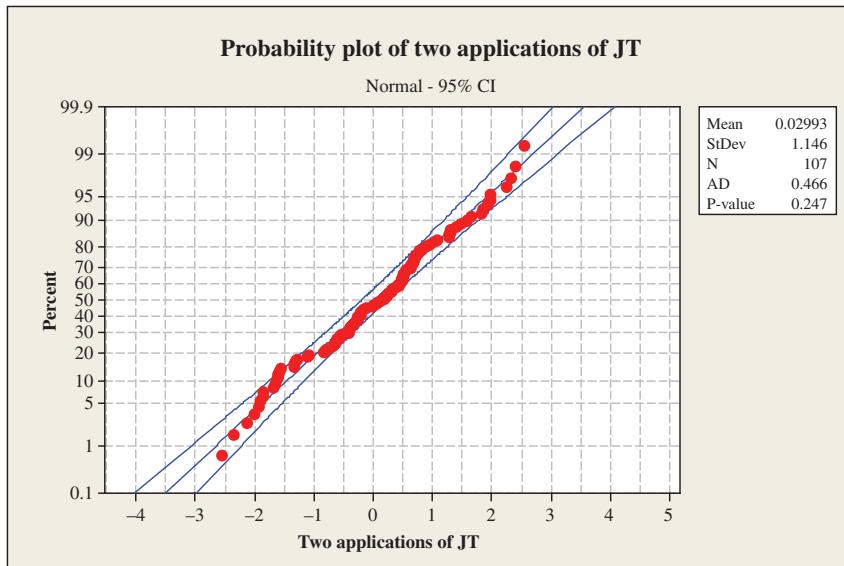
where  $x$  represents the number of genera.

Fits of the generic diversity data against different classes of distributions (normal, lognormal, two-parameter exponential, and Weibull) are shown in Figure 12.1 given later. The normal distribution gives the best fit, corresponding to the highest  $p$ -value under an Anderson–Darling (AD) test. In Figure 12.2, a histogram for the JT diversity data is given. It is clear from Figure 12.3 that a JT for the diversity data gets provides a closer to the normal distribution, particularly following two applications of the transformation.

- $p$ -value for the best fit: 0.247484
- Z for best fit: 0.84



**FIGURE 12.2** Histogram of (a) JT for diversity data and (b) two applications of JT for diversity data.



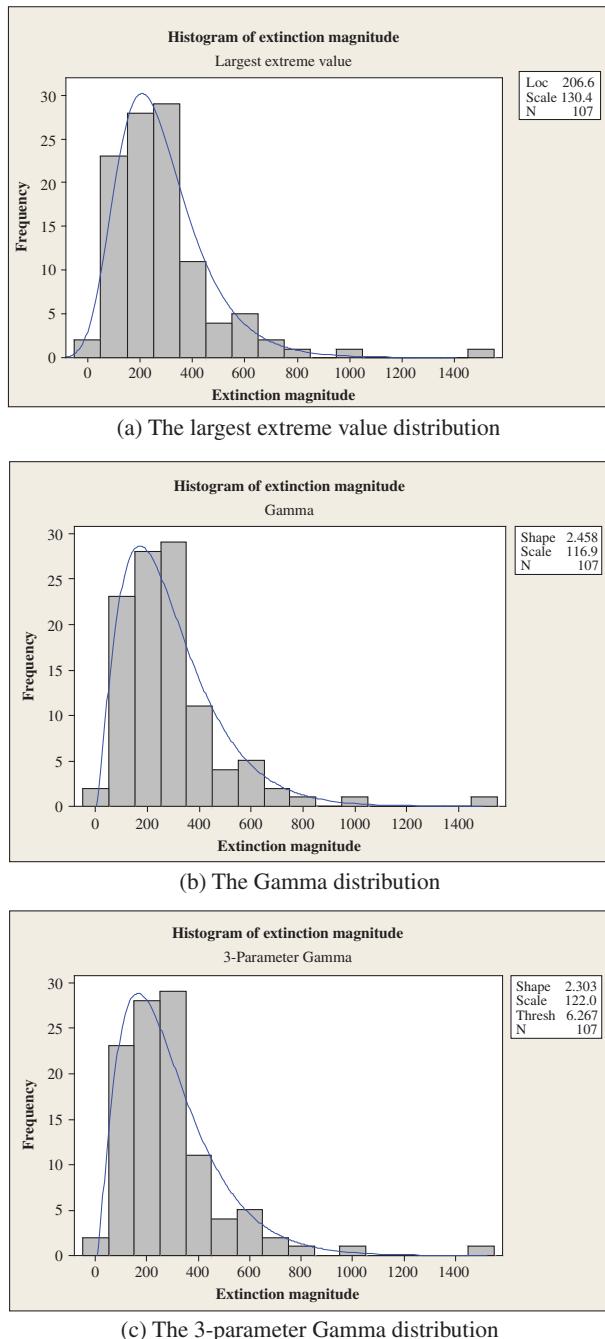
**FIGURE 12.3 Plot of two applications of Johnson's transformation to the diversity data.**

- Best transformation type: SB
- Transformation function equals  $0.4440.95 + 1.13368 \ln \left( \frac{X+1.77033}{2.39370-X} \right)$

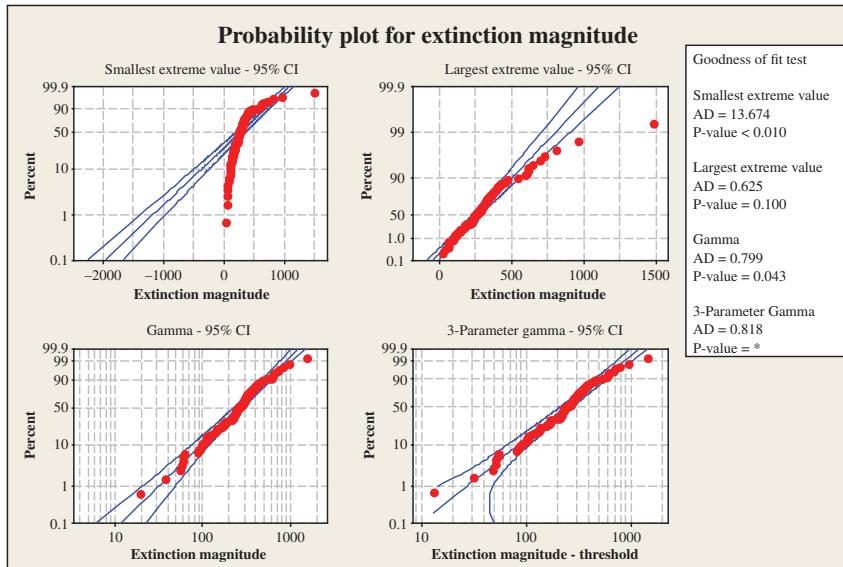
We also fitted distributions of the magnitude of extinction events to the different classes of hypothesized distributions; as shown in Figure 12.4, this data is plotted with different distributions. In view of all these, the best fit, defined again as having the highest *p*-value under the AD test, is provided by the maximum extreme value distribution. The second best fit is to the Gamma distribution, as we would expect from the theory presented in the previous sections. Figures 12.5, 12.6, 12.7 and 12.10 provide other statistical studies for the extinction magnitude data and the diversity data.

## 12.4 Analysis with Lévy distribution

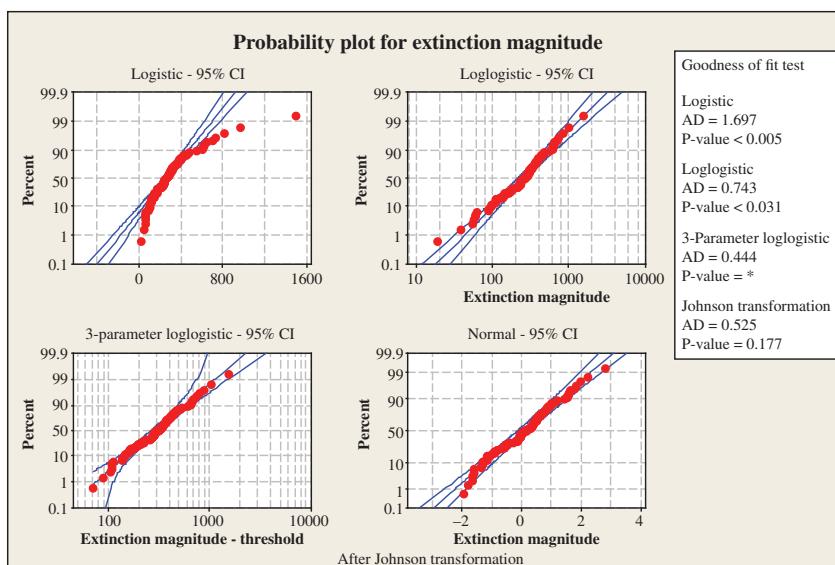
In this section, we present an analysis of the total number of genera through geological time, using truncated Lévy distributions. It is proposed that this class of distributions may optimally characterize diversity dynamics, which are driven by low background origination and extinction events punctuated by rare events (mass extinctions or adaptive radiations) of much larger magnitude.



**FIGURE 12.4 Histogram of extinction magnitude fitting with (a) the largest extreme value distribution; (b) the Gamma distribution; and (c) the 3-parameter Gamma distribution.**



**FIGURE 12.5** Probability plot of extinction magnitude data.



**FIGURE 12.6** Probability plot of extinction magnitude data.

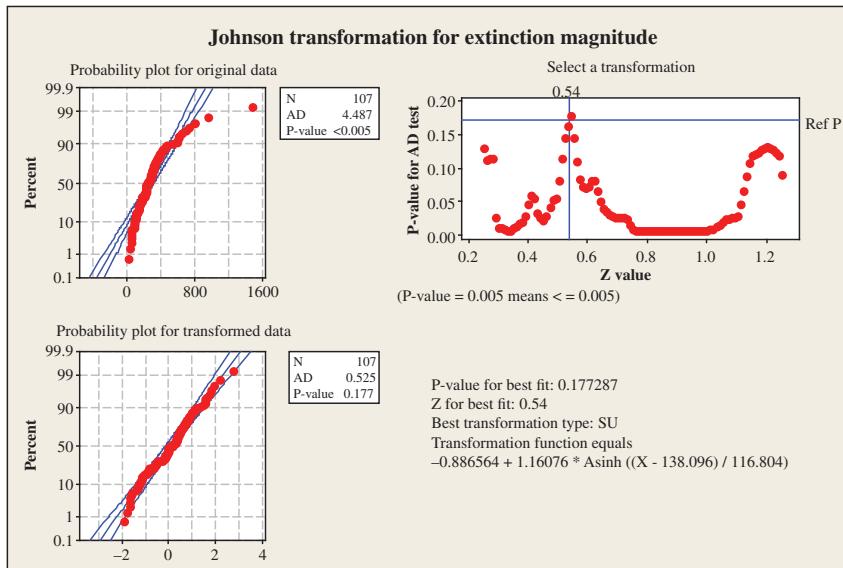


FIGURE 12.7 JT for extinction magnitude data.

### 12.4.1 CHARACTERIZATION OF STABLE DISTRIBUTIONS

Lévy [22] and Khintchine [23] found the most general form of the stable distributions. The general representation is through the characteristic function  $\varphi(q)$  associated with the distribution. The most common parametrization is:

$$\ln(\varphi(q)) = \begin{cases} i\mu q - \gamma^\alpha |q|^\alpha \left[ 1 - i\beta \frac{q}{|q|} \tan\left(\frac{\pi\alpha}{2}\right) \right] & \text{if } \alpha \in (0, 2] \setminus \{1\} \\ i\mu q - \gamma |q| \left[ 1 + i\beta \frac{q}{|q|} \frac{2}{\pi} \ln |q| \right] & \text{if } \alpha = 1 \end{cases}, \quad (12.11)$$

where  $\alpha$  is called the index of stability (tail index, tail exponent, or characteristic exponent),  $\gamma$  is a positive scale factor,  $\mu$  is a location parameter, and  $\beta \in [-1, 1]$  is a skewness (asymmetry) parameter. For the index  $\alpha > 2$ , the inverse Fourier transform of  $\varphi$  gives a function that is not positive semidefinite and thus cannot be considered as probability density function [24].

It is easily demonstrated that the form of the characteristic function determines the stable distributions. The characteristic function of a convolution of independent random variables is simply the product of the

respective characteristic functions and thus, owing to the special exponential form above, all stable distributions are closed under convolutions for a fixed value of  $\alpha$  and time increments of the same size. All such distributions are heavy tailed (leptokurtic) and in fact have no second moments for any  $\alpha < 2$ . When  $\alpha < 1$ , the distribution does not even have the first moment. It is remarked that neither third nor fourth moment exists for these distributions so the usual measures of skewness and kurtosis are undefined.

The characteristic function always exists and thus the definition given in (12.11) always holds; however, finding the associated density function is not always possible (i.e., it may not exist). Typically, one of the inversion formulae is used but the analytical form of the stable distributions is known only for a few values of  $\alpha$  and  $\beta$ :

1. When  $\alpha = 2$ , one obtains the Gaussian (normal) distribution with mean  $\mu$  and variance  $2\gamma^2$  (the skewness parameter  $\beta$  has no effect).

$$f(x) = \frac{1}{\sqrt{4\pi}\gamma} e^{-\frac{(x-\mu)^2}{4\gamma^2}}.$$

2. When  $\alpha = \frac{1}{2}$ ,  $\beta = 1$ , the distribution is called the Lévy–Smirnov distribution.

$$f(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{e^{-\gamma/2(x-\mu)}}{(x-\mu)^{\frac{3}{2}}}, \text{ if } x \geq \mu$$

3. When  $\alpha = 1$  and  $\beta = 0$ , we obtain the Cauchy (Lorentz) distribution,

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-\mu)^2}.$$

### 12.4.2 TRUNCATED LÉVY FLIGHT (TLF) DISTRIBUTION

The leptokurtic property of the stable distributions for  $\alpha < 2$  has wide applicability to model processes where extreme events have a greater probability of occurring in comparison to normal or exponential distributions. Choosing to use Lévy–Khintchine-type distributions is seemingly straightforward; however, the infinite variance of these distributions for any  $\alpha < 2$  becomes a source of complication when working with real data. In order to avoid this problem, Mantegna and Stanley [25] consider a Lévy-type distribution truncated at some parameter  $l$ , which gives a distribution with

finite variance. This distribution was named the Truncated Lévy flight (henceforth *TLF*):

$$T(x) = cP(x)\mathbf{1}_{(-l,l)}(x),$$

where  $P(x)$  denotes any symmetric Lévy distribution obtained from its characteristic function in (12.11) when  $\beta = 0$ ,  $\mathbf{1}_A(x)$  is the indicator function of the set  $A$  and  $c$  is normalizing constant.

Clearly, as  $l \rightarrow \infty$ , one obtains the regular Lévy distribution characterized by  $P(\cdot)$  (a stable distribution). However, the *TLF* distribution itself is not stable for any finite truncation level  $l$ . Instead, this distribution has finite variance; thus, independent variables drawn from this distribution satisfy a regular central limit theorem. If the parameter  $l$  is large, the convergence to the limiting Gaussian may be very slow [25]. If the parameter  $l$  is small (so that the convergence is fast), the cut in its tails is very abrupt.

In order to have continuous tails, Koponen [26] considered a *TLF* in which the cut function is a decreasing exponential characterized by a separate parameter  $l$ . The characteristic function of this distribution, when  $\alpha \neq 1$ , can be expressed as:

$$\varphi(q) = \exp \left\{ c_0 - c_1 \frac{(q^2 + 1/l^2)^{\frac{\alpha}{2}}}{\cos(\pi\alpha/2)} \cos(\alpha \arctan(l|q|)) (1 + il|q|\beta \tan(q \arctan l|q|)) \right\}$$

with  $c_1$  a scale factor:

$$c_1 = \frac{2\pi \cos(\pi\alpha/2)}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} A t$$

and

$$c_0 = \frac{l^{-\alpha}}{\cos(\pi\alpha/2)} c_1 = \frac{2\pi}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} A l^{-\alpha} t.$$

In the case of symmetric distributions  $\beta = 0$ , the variance can be calculated using the characteristic function:

$$\sigma^2(t) = - \left. \frac{\partial^2 \varphi(q)}{\partial q^2} \right|_{q=0} = t \frac{2A\pi(1-\alpha)}{\Gamma(\alpha) \sin(\pi\alpha)} l^{2-\alpha}$$

All the following discussion is in this symmetric case ( $\beta = 0$ ).

If we use time steps  $\Delta t$ , and  $T = N\Delta t$ , following the discussion in the previous section, at the end of each interval, we must calculate the sum

of  $N$  stochastic variables that are independent and identically distributed. Therefore, the characteristic function of the sum is:

$$\varphi(q, N) = \varphi(q)^N = \exp \left\{ c_0 N - c_1 \frac{N(q^2 + 1/l^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos(\alpha \arctan(l|q|)) \right\}.$$

The model can be improved by standardizing it. If the variance is given by:

$$\sigma^2 = - \left. \frac{\partial^2 \varphi(q)}{\partial q^2} \right|_{q=0}$$

we have that

$$-\left. \frac{\partial^2 \varphi(q/\sigma)}{\partial q^2} \right|_{q=0} = -\frac{1}{\sigma^2} \left. \frac{\partial^2 \varphi(q)}{\partial q^2} \right|_{q=0} = 1$$

Therefore, a standardized model is:

$$\begin{aligned} \ln \varphi_S(q) &= \ln \varphi \left( \frac{q}{\sigma} \right) = c_0 - c_1 \frac{((q/\sigma)^2 + 1/l^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos \left( \alpha \arctan \left( l \frac{|q|}{\sigma} \right) \right) \\ &= \frac{2\pi Al^{-\alpha} t}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} \left[ 1 - \left( \left( \frac{ql}{\sigma} \right)^2 + 1 \right)^{\alpha/2} \cos \left( \alpha \arctan \left( \frac{ql}{\sigma} \right) \right) \right]. \end{aligned} \quad (12.12)$$

Next we state a well-known theorem applicable in our case.

**Theorem 12.2.** Let  $\hat{p}(k)$  be the characteristic function (Fourier transform) of  $p(x)$ . Let  $\log \hat{p}(k) = f_a(k) + f_s(k)$ , where  $f_a(k)$  is analytic and  $f_s(k) \sim -a|k|^\alpha$  as  $k \rightarrow 0$  with  $a \in \mathbb{R}$  and finite. Then

$$p(x) \sim \frac{A}{|x|^{1+\alpha}}, \quad \text{as } x \rightarrow \infty,$$

where

$$A = \frac{a \sin(\alpha\pi/2)\Gamma(\alpha)\alpha}{\pi}.$$

Going back to the TLF distribution, for small  $q$ ,

$$\log \varphi_S(q) \sim \frac{2\pi Al^{-\alpha} t}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} \left[ 1 - \left( 1 + \frac{\alpha}{2} \left( \frac{ql}{\sigma} \right)^2 \right) \cos \left( \alpha \arctan \left( \frac{ql}{\sigma} \right) \right) \right].$$

Therefore, using the above theorem,  $p(x) \sim \frac{1}{|x|^{1+\alpha}}$ .

We can also justify this from the viewpoint of the cutoff function used in [26] given by:

$$f(x) = \begin{cases} A_- e^{-|x|/l} |x|^{-1-\alpha} & x < 0 \\ A_+ e^{-|x|/l} |x|^{-1-\alpha} & x \geq 0. \end{cases}$$

We are considering

$$p(x) = \begin{cases} f(x) & x > l \\ cL(x) & -l \leq x \leq l \\ f(x) & x < -l. \end{cases}$$

Here  $L(x)$  is the Lévy distribution.

The decay is faster than the *heavy tail* decay because of the presence of the exponential function in  $f(x)$ . However, the main advantage of using this form of density  $p(x)$  is that we can keep the *heavy tail* of the Lévy for arbitrarily large cutoff values (by adjusting the parameter  $l$ ) and then have a sharp decay (faster than *heavy tail*). In other words, we are keeping the features of the heavy tail of the Lévy distribution as well as the finite variance feature given by exponential decay and truncation.

To simulate the standardized truncated Lévy model, a Matlab module (available from the first author upon request) was developed. The parameter  $l$  is fixed at 1 and then the parameter  $A$  and the characteristic exponent  $\alpha$  are adjusted simultaneously to fit the cumulative function. On the same grid, the cumulative distribution of the simulated data is plotted for different time lags  $T$  in order to visually evaluate the goodness of fit. Time lag  $T = 1$  means that the fit is done by using consecutive data points, while for a general lag  $T = k$ , the fit was generated by using points  $k$  observations apart.

#### 12.4.2.1 Kurtosis

In general,  $n^{\text{th}}$  moment of distribution, if it is finite, can be computed from characteristic function  $\varphi(q)$  in a similar way we would compute  $n^{\text{th}}$

moment from the moment generating function. The  $n^{\text{th}}$  moment computed from characteristic function  $\varphi$  will be

$$\mu_n = \mathbb{E}[X^n] = \frac{1}{l^n} \left. \frac{\partial^n \varphi}{\partial q^n} \right|_{q=0}.$$

For all parameters of  $\alpha$  except  $\alpha = 2$  of the Lévy stable distribution, we have infinite fourth moment. For a truncated Lévy distribution, we can compute the fourth moment of (12.12). A lengthy but straightforward computation gives us

$$\mu_4 = -\frac{\alpha c_0 l^4 (\alpha - 1)(3\alpha c_0 - 5\alpha - 3\alpha^2 c_0 + \alpha^2 + 6)}{\sigma^4}$$

Since (12.12) is normalized, the kurtosis will be [20]:

$$\begin{aligned} \text{Kurt}[X] &= \gamma_2 = \mu_4 - 3 \\ &= -\frac{\alpha c_0 l^4 (\alpha - 1)(3\alpha c_0 - 5\alpha - 3\alpha^2 c_0 + \alpha^2 + 6)}{\sigma^4} - 3 \end{aligned} \quad (12.13)$$

Kurtosis is useful as a measure of the degree to which there is “excess” probability density in comparison to a normal distribution. Specifically, a leptokurtic ( $\text{Kurt}[X] > 0$ ) distribution is one where extreme values have a higher probability than predicted from a Gaussian model. The distribution of extinction magnitudes is leptokurtic, and we propose that a truncated Lévy distribution may provide a better fit than a normal.

#### 12.4.2.2 Infinite divisibility

We say that random variable  $Y$  is infinitely divisible if it can be written as sum or  $n$  independent copies of some random variable  $X$ . To see whether TLS is infinitely divisible, we introduce the following theorem.

**Theorem 12.3.** *Let  $X_i$  be independent and identically distributed random variables with characteristic function  $\varphi_X$ . Then random variable  $Y$  given by*

$$Y = \sum_{i=1}^n X_i$$

*has characteristic function  $\varphi_Y$  given by*

$$\varphi_Y = \prod_{i=1}^n \varphi_X = (\varphi_X)^n.$$

From the form of (12.12), we propose a characteristic function showing that TLF is infinitely divisible

$$\log \varphi_{S,N} = \frac{2\pi Al^{-\alpha}t}{N\alpha\Gamma(\alpha)\sin(\pi\alpha)} \left[ 1 - \left( \left( \frac{ql}{\sigma} \right)^2 + 1 \right)^{\alpha/2} \cos \left( \alpha \arctan \left( \frac{ql}{\sigma} \right) \right) \right].$$

Similarly, we can see that the random variable given by the equation above is again TLF. For sum of random variables given by  $\varphi_{S,N}$  and using above theorem, we get:

$$\varphi_S = (\varphi_{S,N})^N$$

This means that TLF is infinitely divisible, giving us (12.12).

### 12.4.3 DATA ANALYSIS WITH TLF DISTRIBUTION

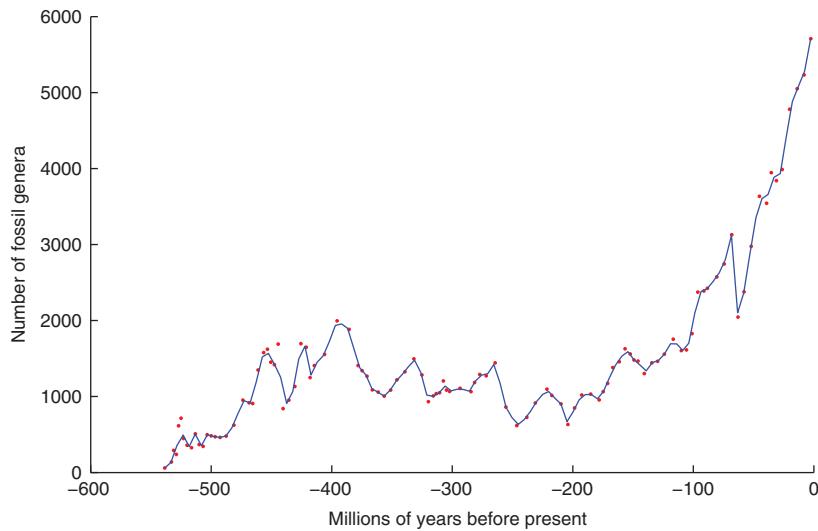
In this subsection, we present an analysis of temporal change in diversity from [1]. Lévy stable distribution assumes data that are equally spaced in time, that is not the case in the data set [2]. To adjust for heterogeneous time intervals (a necessary consequence of differences in the length of geological epochs and stages at which generic diversity was estimated), we need to make the following assumption.

**Assumption 12.1.** *Diversity can be linearly interpolated with only a small error.*

This assumption is not unrealistic, as in Figure 12.8, one can see the interpolation of diversity (blue line) compared with the original data (red dots). It can be seen throughout that the interpolation line does not deviate from the data and thus Assumption 12.1 is supported by the data.

Now we can proceed to analyze diversity time series  $D(t_i)$ , which is defined for any time  $t_i \in \{t_0 = t_{\min}, t_1 = t_0 + \Delta t, t_2 = t_1 + 2\Delta t, \dots, t_n = t_{\max} = t_0 + n\Delta t\}$  due to interpolation. Diversity is assumed to behave exponentially, like any other population dynamics model. We will analyze distribution of logarithm of changes in diversity  $D$  for specific time lag  $T$

$$G_T(i) = \log \frac{D(i+T)}{D(i)}, \quad (12.14)$$



**FIGURE 12.8** Diversity (red dots) and its linear interpolation (blue line).

where  $D(i)$  is diversity at time  $t_i$  and  $D(i + T)$  is diversity at time  $t_{i+T}$ , where  $T$  is time lag. For our purposes, we will standardize (12.14) to

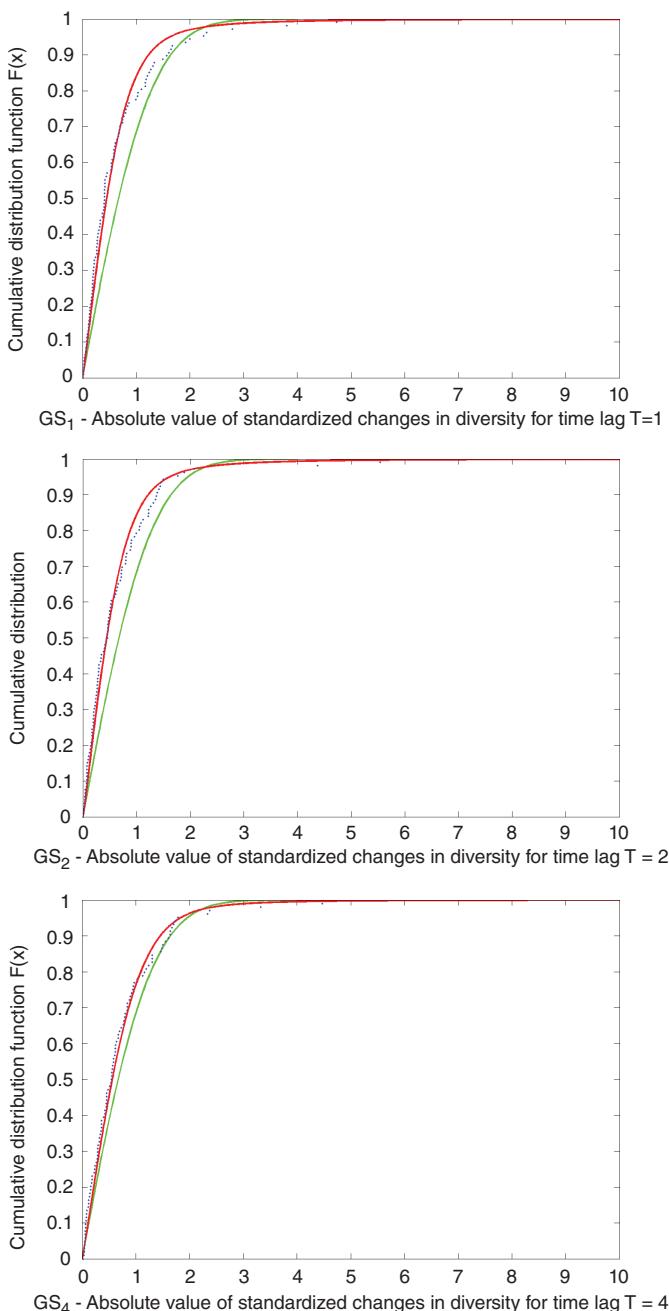
$$GS_T = \left| \frac{G_T - E(G_T)}{\sigma(G_T)} \right| \quad (12.15)$$

where term  $\frac{G_T - E(G_T)}{\sigma(G_T)}$  will have mean zero and variance equal to 1.

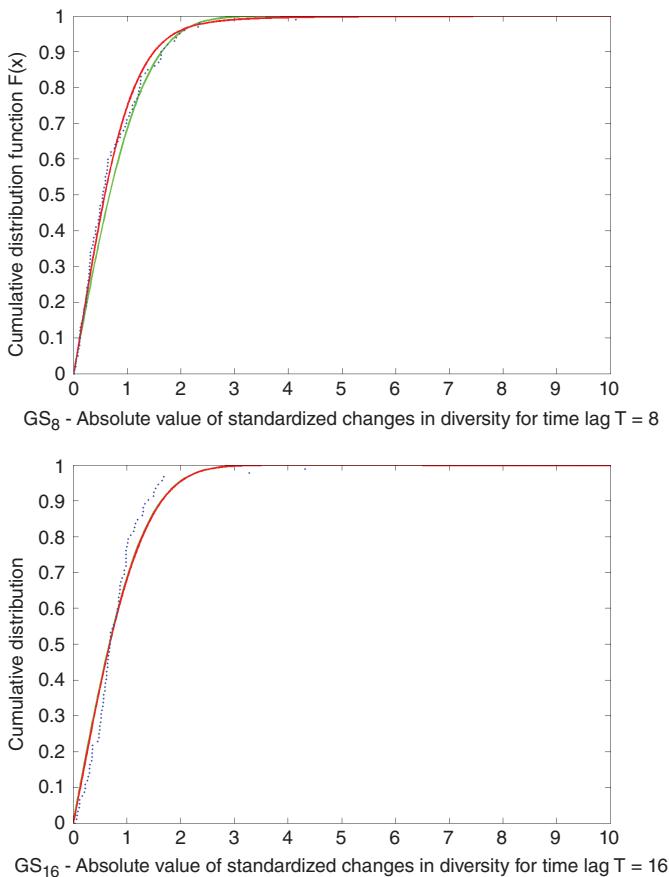
#### 12.4.4 SUM OF LÉVY RANDOM VARIABLES WITH DIFFERENT PARAMETERS

We present a result for the Lévy distribution that is analogous to the one found in the previous sections for convolutions of exponential distributions. Let us assume random variables  $X_1, \dots, X_n$  with characteristic function (12.12) and coefficients  $\beta = 0, \mu = 0, \gamma_k = \gamma$ . The characteristic function for  $X_k$  is

$$\varphi_{X_k}(q) = e^{-|\gamma q|^{\alpha_k}}. \quad (12.16)$$



**FIGURE 12.9** Data fitted with distribution (12.12), empirical distribution for  $GS_T$ , and standard normal distribution. Time lag  $T = 1, 2, 4, 8$ , and 16.

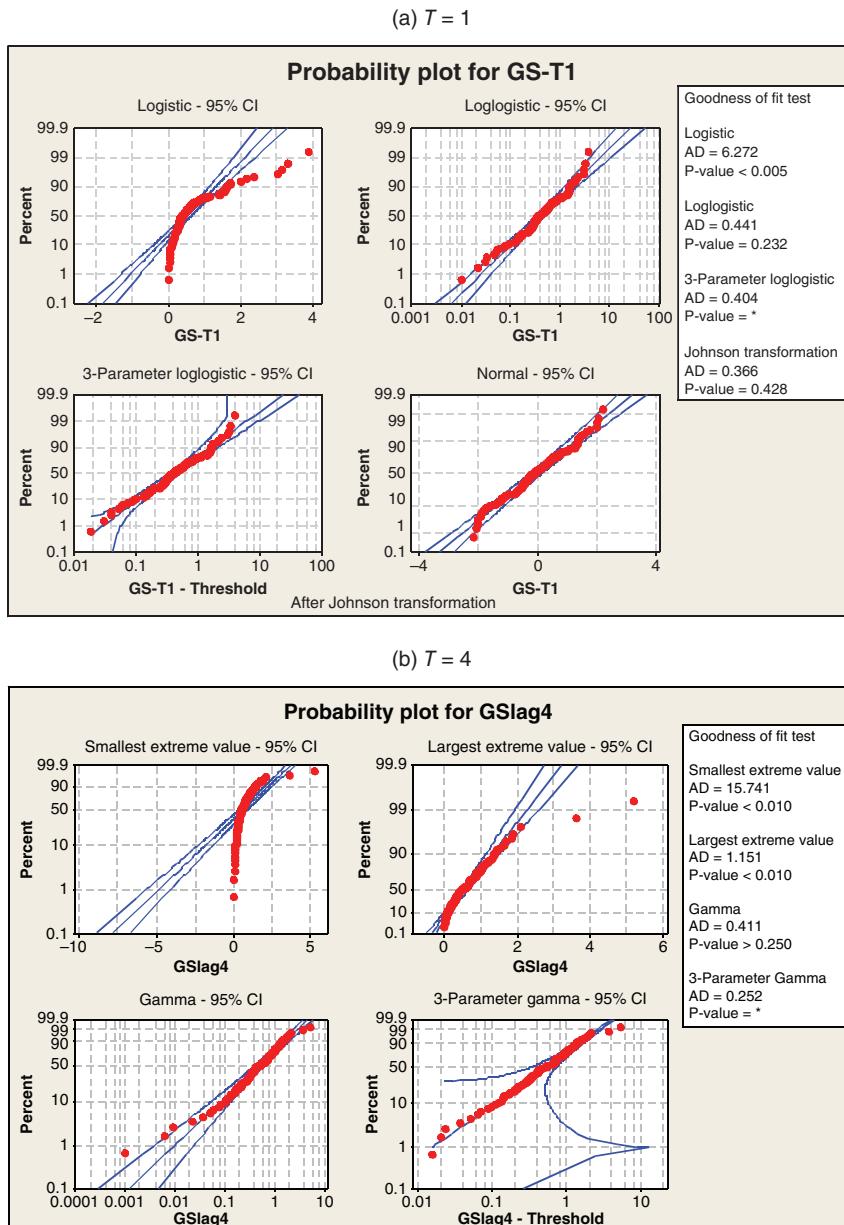
FIGURE 12.9 (*Continued*)

We will use a transformation of parameters  $\alpha_k$

$$\begin{aligned}\alpha_k &= \bar{\alpha} + \Delta\alpha_k, k \in \{1, \dots, n\} \\ \bar{\alpha} &= \frac{1}{n} \sum_{j=1}^n \alpha_j \\ 0 &= \sum_{j=0}^n \Delta\alpha_k\end{aligned}\tag{12.17}$$

where  $\Delta\alpha_k$  is a small perturbation from the mean value  $\bar{\alpha}$ . If we substitute  $\alpha_k$  into (12.16), we get

$$\varphi_{X_k} = e^{-|\gamma q|^{\bar{\alpha} + \Delta\alpha_k}}.\tag{12.18}$$



**FIGURE 12.10** Distribution of log change in diversity for time lag  $T$  plotted against various distributions. (a)  $T=1$ ; and (b)  $T=4$ .

Function (12.18) is holomorphic (it is also entire) in  $\Delta\alpha_i$  and has a power series expansion in terms of  $\Delta\alpha_i$  at zero, that is,

$$\varphi_{X_k} = e^{-|\gamma q|^{\bar{\alpha}}} + e^{-|\gamma q|^{\bar{\alpha}}}(-|\gamma q|^{\bar{\alpha}}) \ln(|\gamma q|) \cdot \Delta\alpha_k + \mathcal{O}(\Delta\alpha_k^2) \quad (12.19)$$

Again, the sum of random variables  $X_k$  will have the characteristic function given by a product of characteristic functions

$$\varphi_{\sum X_k} = \prod_{k=1}^n \varphi_{X_k} = \prod_{k=1}^n e^{-|\gamma q|^{\bar{\alpha}} + \Delta\alpha_k}$$

Substituting (12.19) into the equation above, we get

$$\begin{aligned} \varphi_{\sum X_k} = & \left( e^{-|\gamma q|^{\bar{\alpha}}} + e^{-|\gamma q|^{\bar{\alpha}}}(-|\gamma q|^{\bar{\alpha}}) \ln(|\gamma q|) \cdot \Delta\alpha_1 + \mathcal{O}(\Delta\alpha_1^2) \right) \cdot \\ & \left( e^{-|\gamma q|^{\bar{\alpha}}} + e^{-|\gamma q|^{\bar{\alpha}}}(-|\gamma q|^{\bar{\alpha}}) \ln(|\gamma q|) \cdot \Delta\alpha_2 + \mathcal{O}(\Delta\alpha_2^2) \right) \cdot \\ & \left( e^{-|\gamma q|^{\bar{\alpha}}} + e^{-|\gamma q|^{\bar{\alpha}}}(-|\gamma q|^{\bar{\alpha}}) \ln(|\gamma q|) \cdot \Delta\alpha_3 + \mathcal{O}(\Delta\alpha_3^2) \right) \cdot \\ & \left( e^{-|\gamma q|^{\bar{\alpha}}} + e^{-|\gamma q|^{\bar{\alpha}}}(-|\gamma q|^{\bar{\alpha}}) \ln(|\gamma q|) \cdot \Delta\alpha_4 + \mathcal{O}(\Delta\alpha_4^2) \right) \cdot \\ & \vdots \\ & \left( e^{-|\gamma q|^{\bar{\alpha}}} + e^{-|\gamma q|^{\bar{\alpha}}}(-|\gamma q|^{\bar{\alpha}}) \ln(|\gamma q|) \cdot \Delta\alpha_n + \mathcal{O}(\Delta\alpha_n^2) \right) \end{aligned}$$

multiplying this expression gives us

$$\varphi_{\sum X_k} = e^{-n|\gamma q|^{\bar{\alpha}}} + e^{-n|\gamma q|^{\bar{\alpha}}}(-|\gamma q|^{\bar{\alpha}}) \ln(|\gamma q|) \cdot \sum_{k=1}^n \Delta\alpha_k + \mathcal{O}(\Delta\alpha_k^2) \quad (12.20)$$

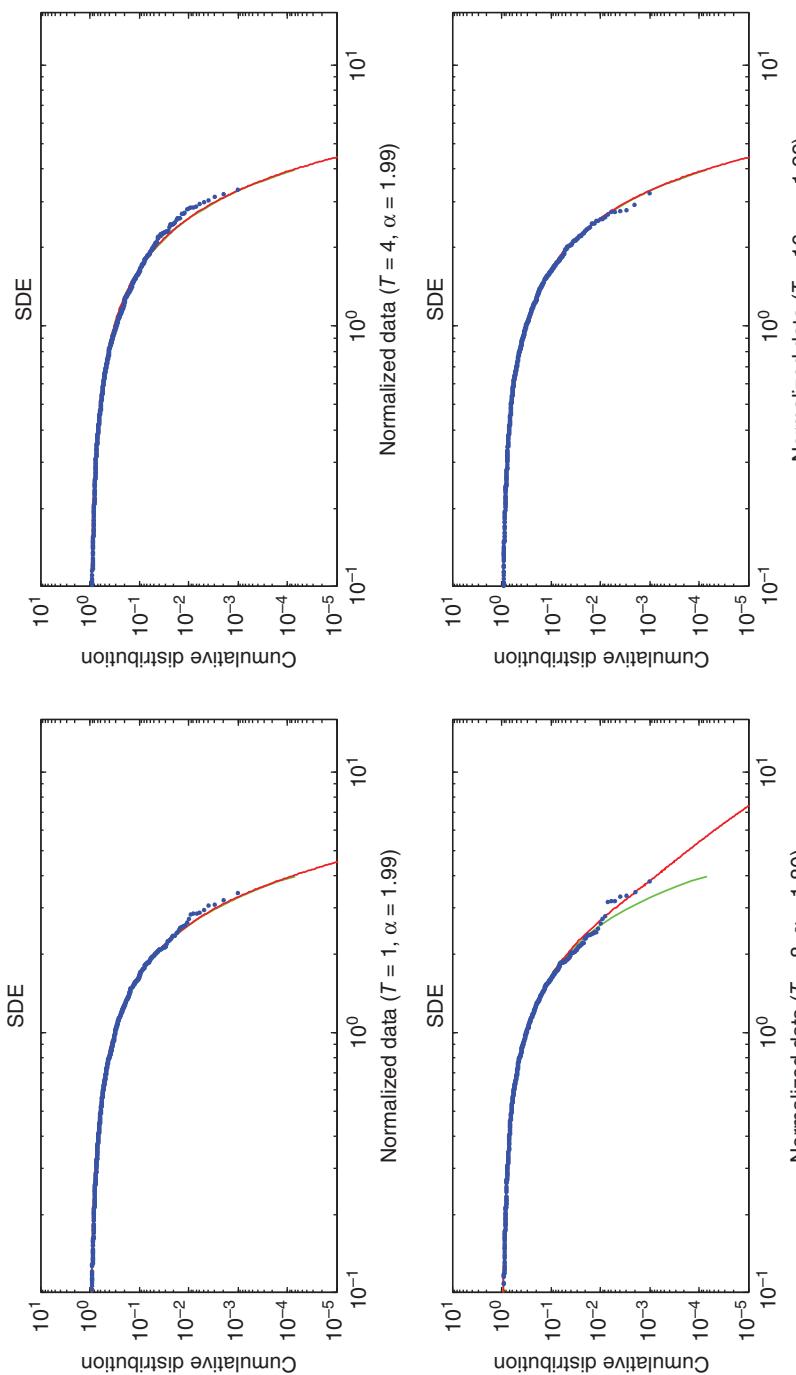
and using that  $\sum_{j=0}^n \Delta\alpha_k = 0$  from (12.17), we finally get

$$\varphi_{\sum X_k} = e^{-n|\gamma q|^{\bar{\alpha}}} + \mathcal{O}(\Delta\alpha_k^2) \quad (12.21)$$

which is *almost* a characteristic function for Lévy stable distribution with a scale factor  $N\gamma$ .

## 12.5 Analysis of the Stock Indices, high-frequency (tick) data, and explosive series

We now present applications to the three different sets of data: simulated data from a population model (Figure 12.11), the stock prices comprising the Dow Jones Industrial Average (DJIA) Index, along with the index itself



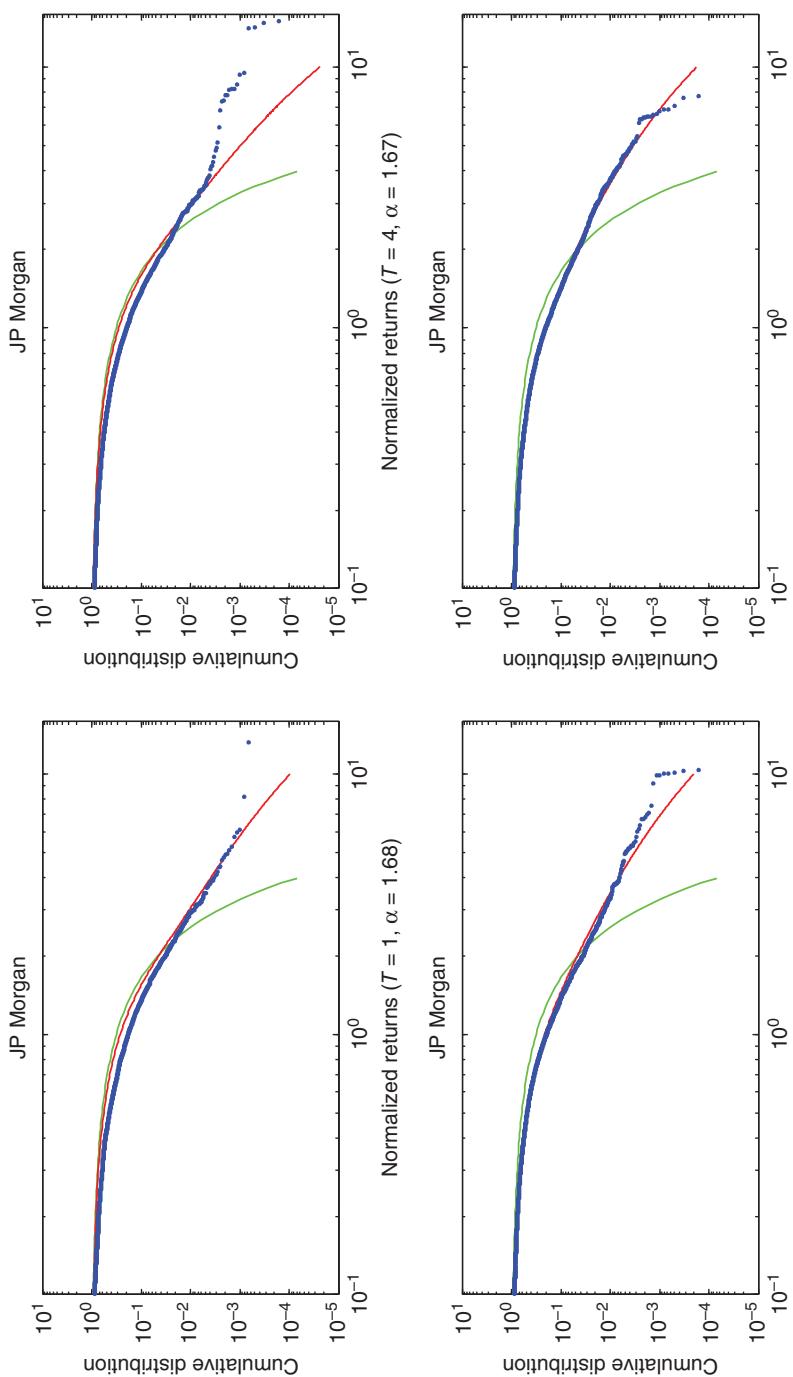
**FIGURE 12.11** The figure shows the estimates of the Lévy Flight parameter for the solution of the SDE.

(Figures 12.12, 12.13, and 12.14), and to the study of HFD from several influential companies (Figures 12.15, 12.16, 12.17, 12.18, and 12.19).

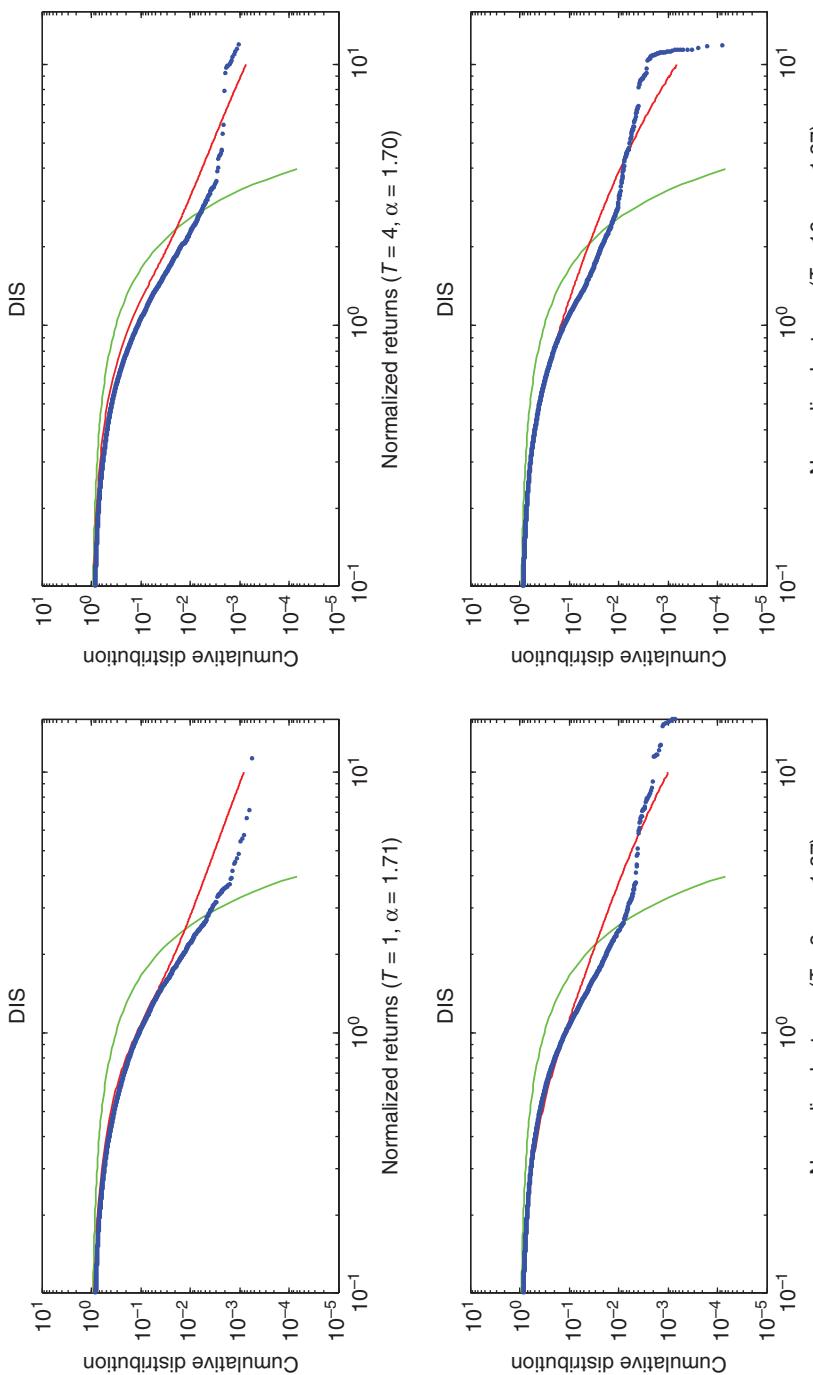
The analyzed stochastic variable is the return  $r_t$ , defined as the difference of the logarithm of two consecutive stock (or index) prices. In this case, we plot on the same grid the cumulative distribution of the observed returns for different time lags  $T$  in order to visualize how good the fitting is. Time lag  $T = 1$  means that the returns are calculated by using two consecutive observations; for a general  $T$ , the returns are calculated by using  $r_t = \log(X_t/X_{t-T})$ . Now,  $X_t = I_t$ , where  $I_t$  denotes the stock (or index) price at time  $t$ , and  $T$  is the difference (in labor days) between two values of the stock or index. We study the behavior of stock prices comprising the DJIA Index, along with the index itself. The values are from 1985 to 2010. We finally analyzed high-frequency (minute) data from 2008. In this case,  $T$  is the difference in minutes between two values of the stock.

We conclude that the Lévy flights are appropriate for modeling the three different sets of data. We recall that a value close to 2.0 indicates Gaussian behavior of the stochastic variable.

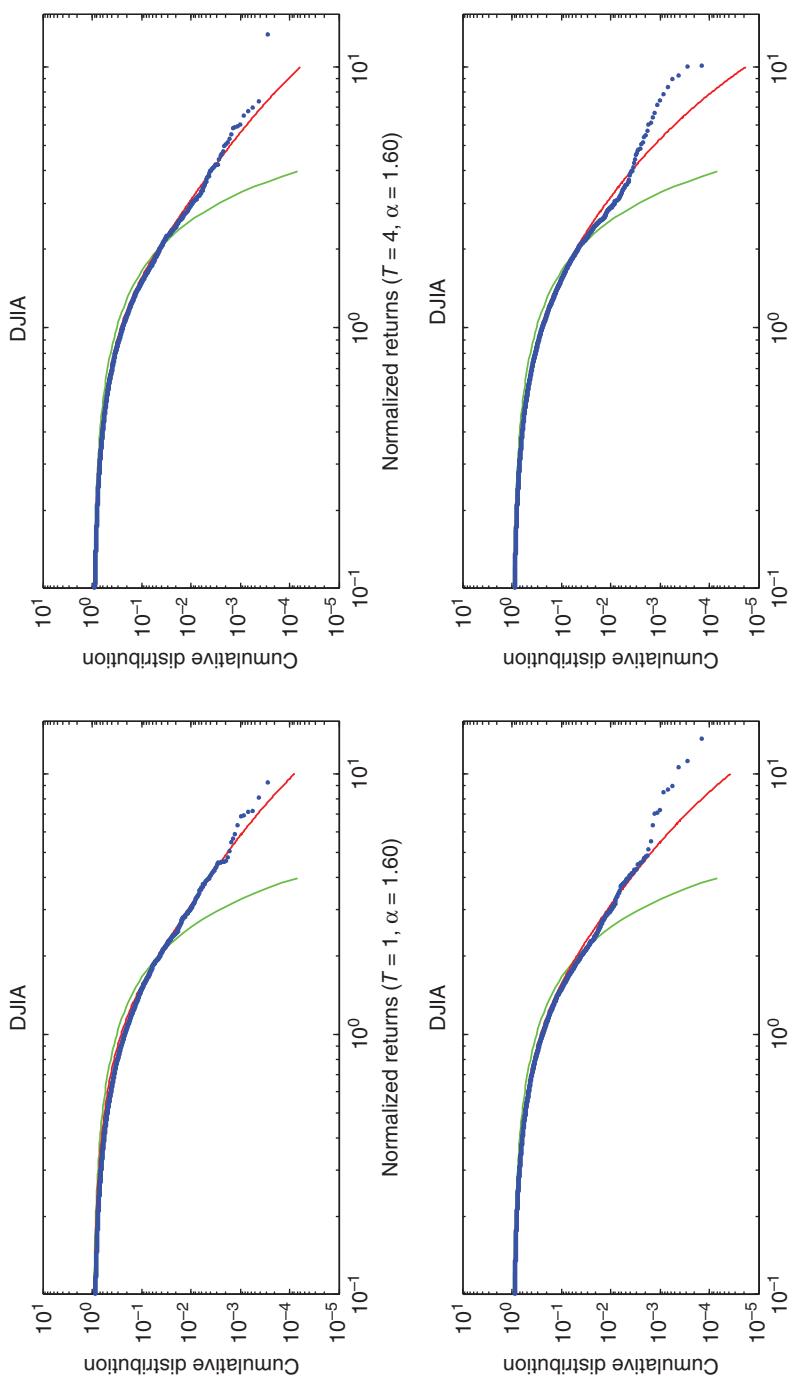
Finally, volatility analysis was done on S&P500 index, Bank of America Corp (BAC), JP Morgan Chase (JPM), International Business Machines Corp (IBM), and Wal-Mart Stores (WMT) HFD, earthquake series (EQ2), and explosives series (EXP). The results are shown in Tables 12.2, 12.3, 12.4, 12.5, 12.6, 12.7, and 12.8 and Figures 12.20, 12.21, 12.22, 12.23, 12.24, 12.25, 12.26, 12.27, 12.28, and 12.29. The studied HFD corresponding to the collapse of the Bear Stearns in March 2008 consists of the week (five trading days) March 10–14, 2008, before the merging announcement over the weekend as well as the two following trading days March 17 and 18. On Friday, March 14, 2008, at about 9:14 a.m., JPM and the Federal Reserves Bank of New York announced an emergency loan to Bear Stearns (of about 29 billion, terms undisclosed) to prevent the firm from becoming insolvent. This bailout was declared to prevent the very likely crash of the market as a result of the fall of one of the biggest investment banks at the time. This measure proved to be insufficient to keep the firm alive and 2 days later, on Sunday, March 16, 2008, Bear Stearns signed a merger agreement with JPM essentially selling the company for \$2 a share (price revised on March 24 to \$10/share). The same stock traded at \$172 in January 2007 and \$93 a share in February 2007. Today, this collapse is viewed as the first sign of the risk management meltdown of investment bank industry in September 2008 and the subsequent global financial crisis and recession.



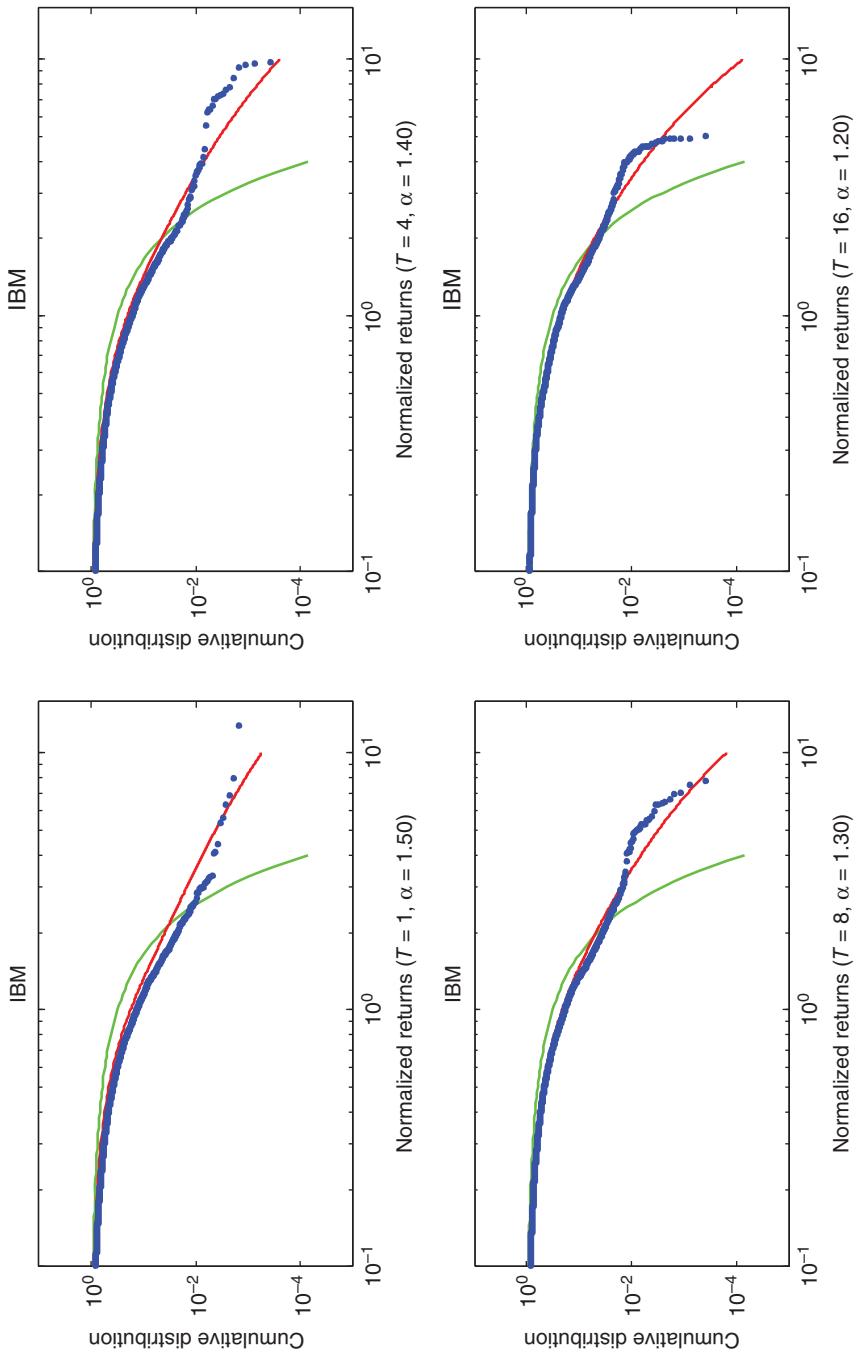
**FIGURE 12.12** The figure shows the estimates of the Lévy flight parameter for JP Morgan.



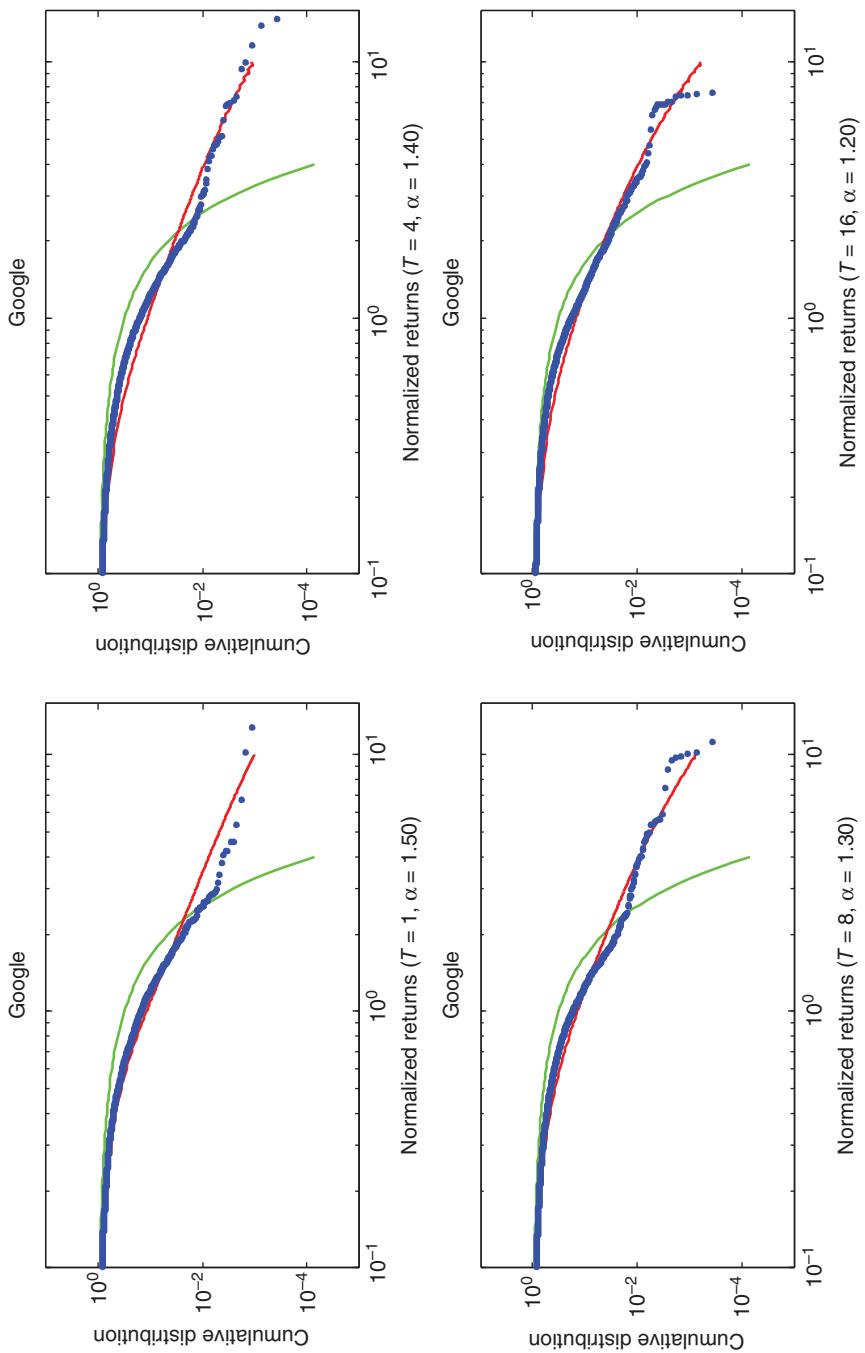
**FIGURE 12.13** The figure shows the estimates of the Lévy Flight parameter for The Walt Disney Company.



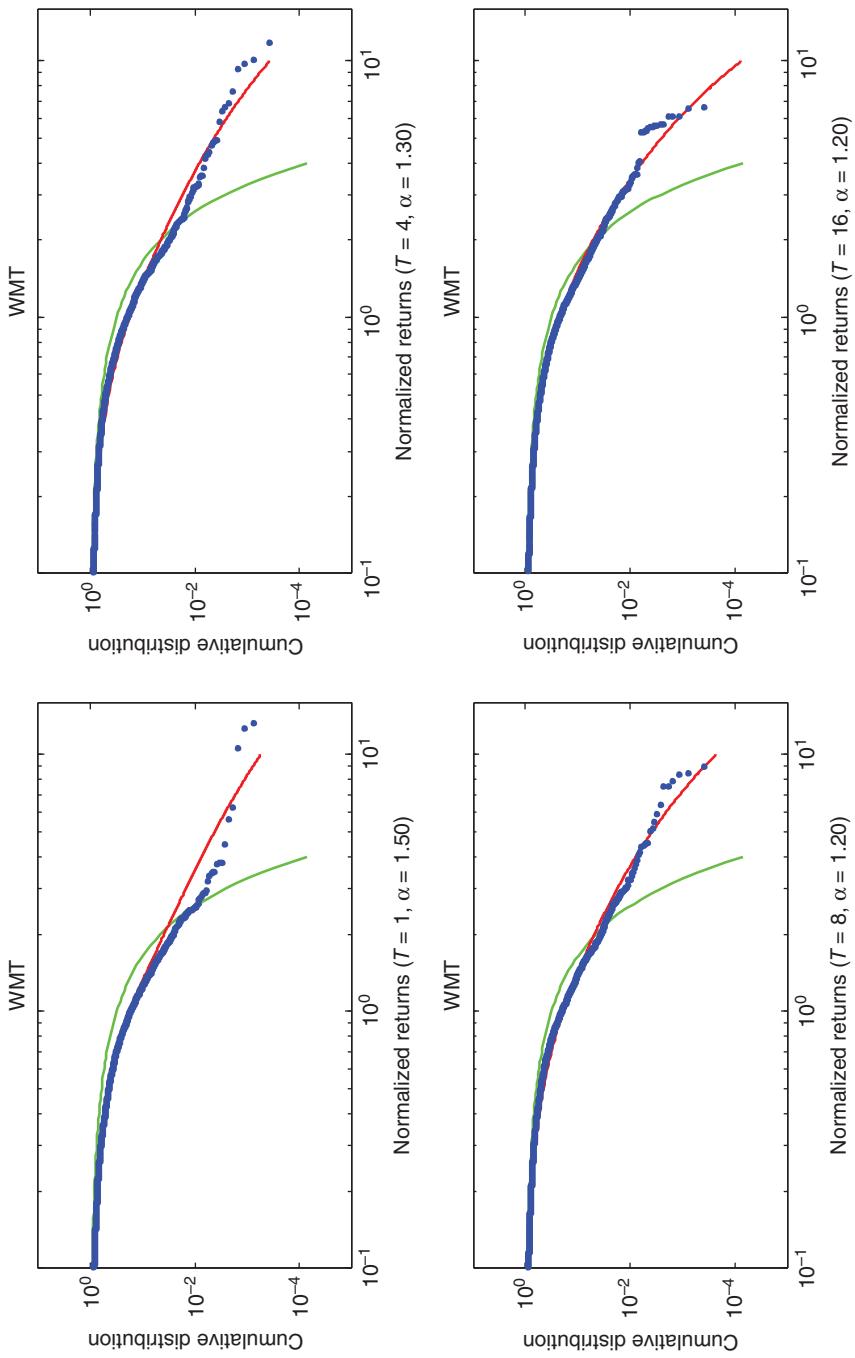
**FIGURE 12.14** The figure shows the estimates of the Lévy Flight parameter for the DJIA index.



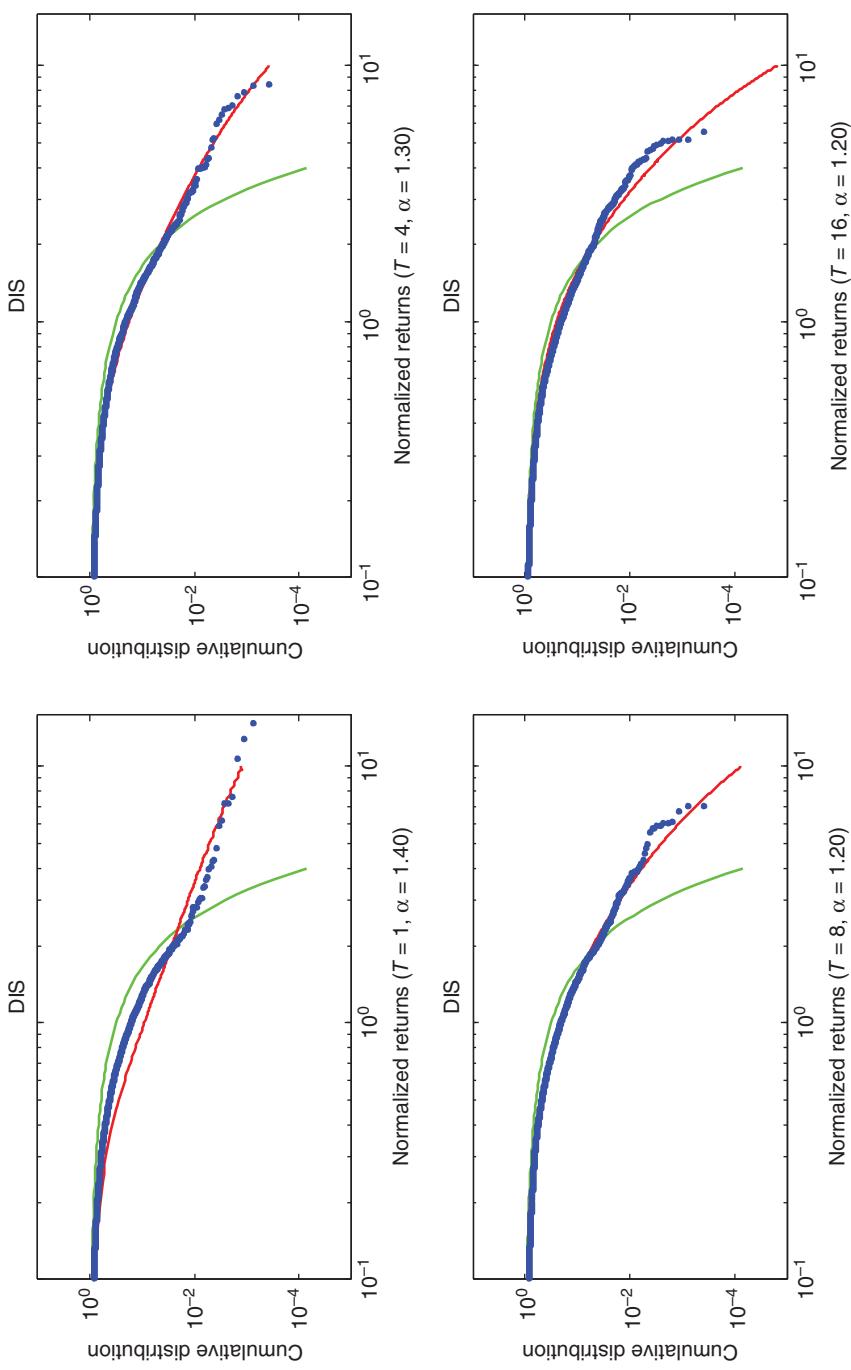
**FIGURE 12.15** The figure shows the estimates of the Lévy Flight parameter for IBM. These are high-frequency (tick) data.



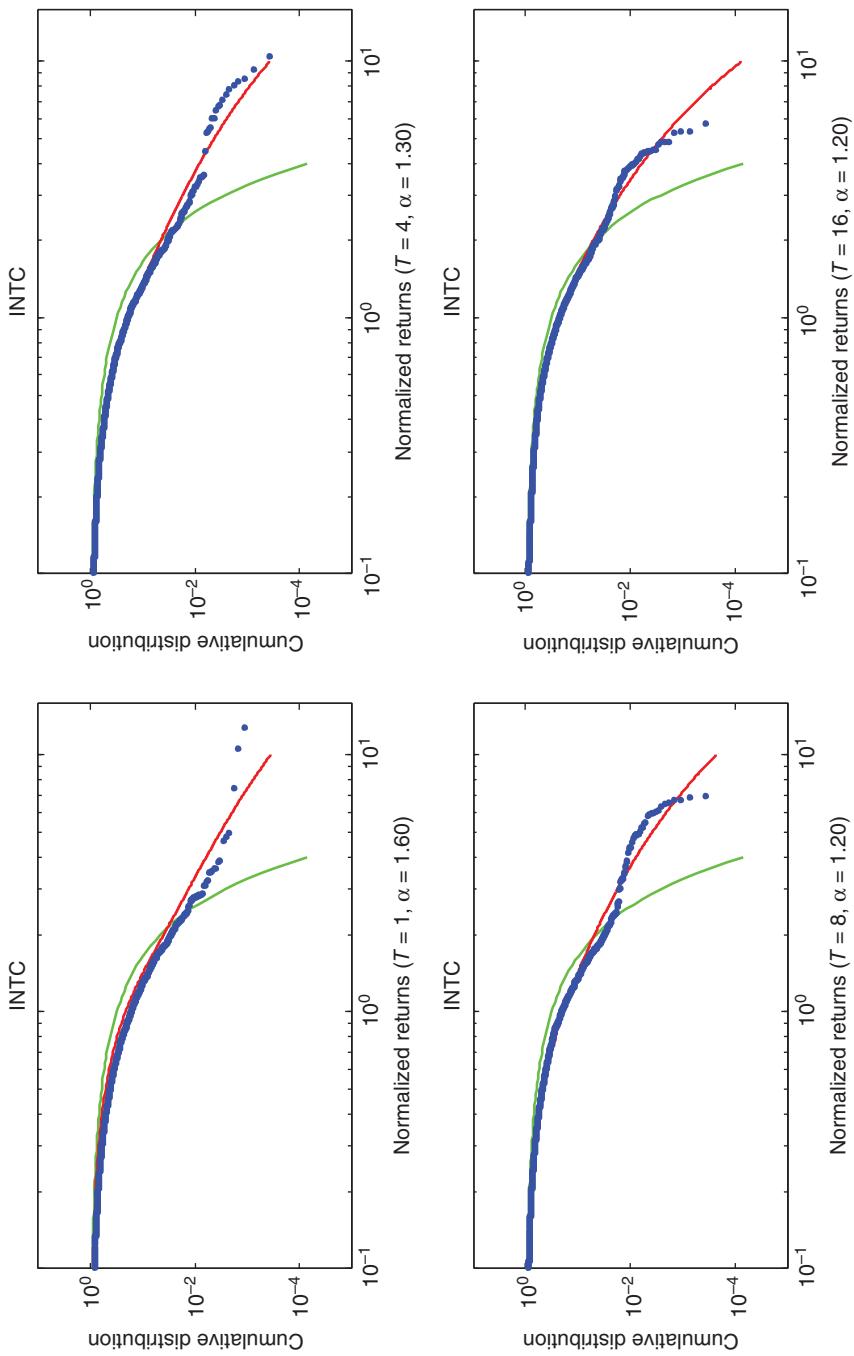
**FIGURE 12.16** The figure shows the estimates of the Lévy Flight parameter for Google. These are high-frequency (tick) data.



**FIGURE 12.17** The figure shows the estimating of the Levy Flight parameter for Walmart. These are high frequency (tick) data.



**FIGURE 12.18** The figure shows the estimates of the Lévy flight parameter for The Walt Disney Company. These are high-frequency (tick) data.



**FIGURE 12.19** The figure shows the estimates of the Lévy flight parameter for Intel Corporation. These are high-frequency (tick) data.

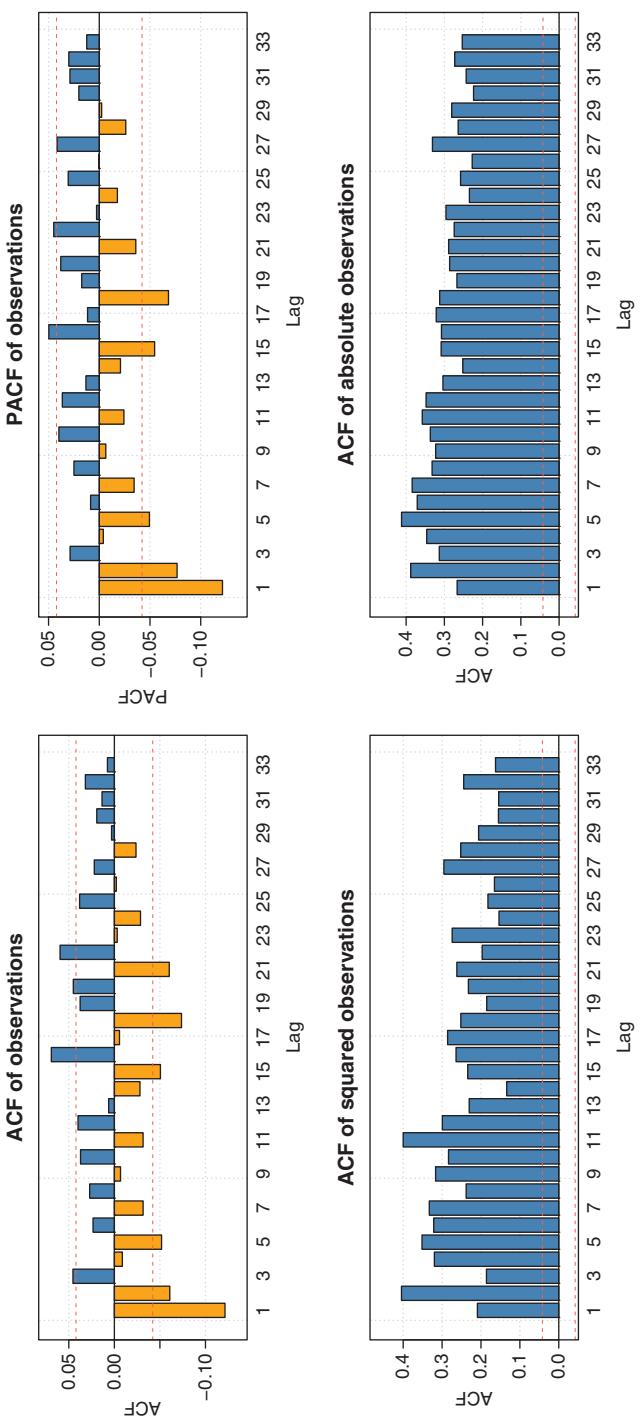
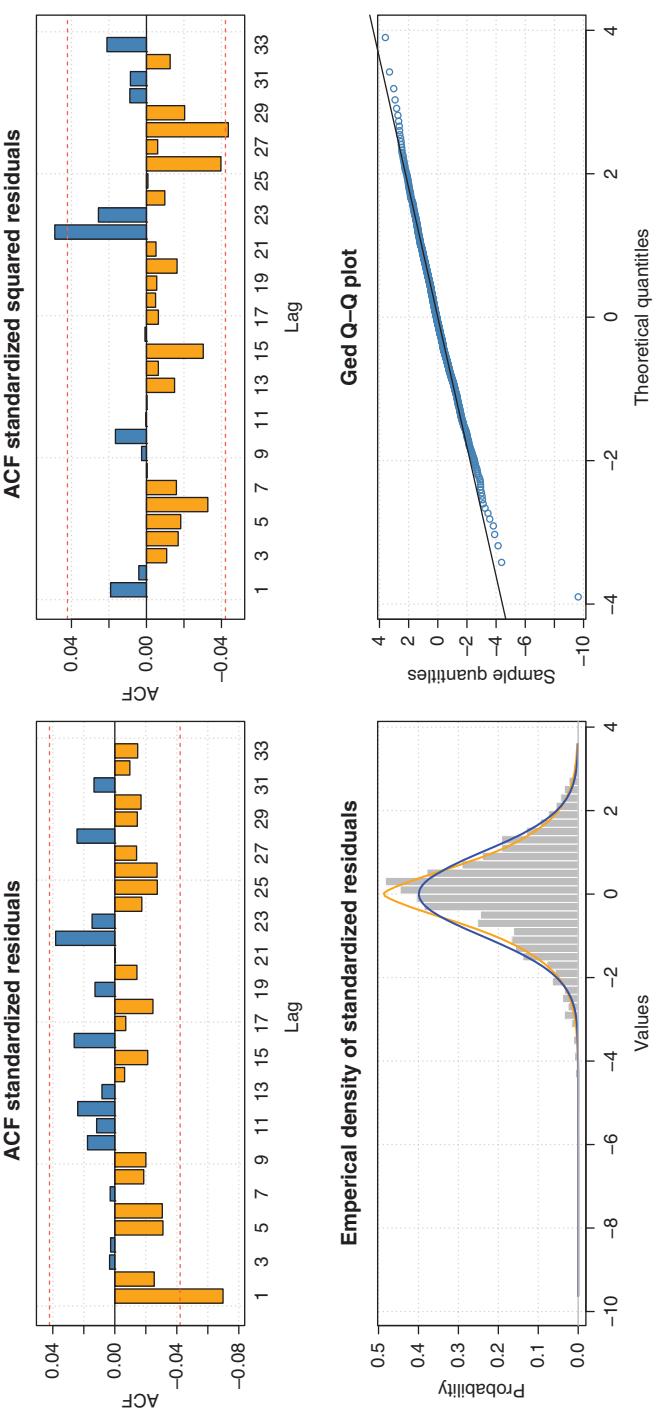


FIGURE 12.20 S&P 500 return and squared and absolute returns ACF and PACF.



**FIGURE 12.21** S&P 500 volatility model diagnostics: On top are the ACF of the standardized and standardized squared residual and on bottom are the distribution and GED Q-Q plot for the fitted model.

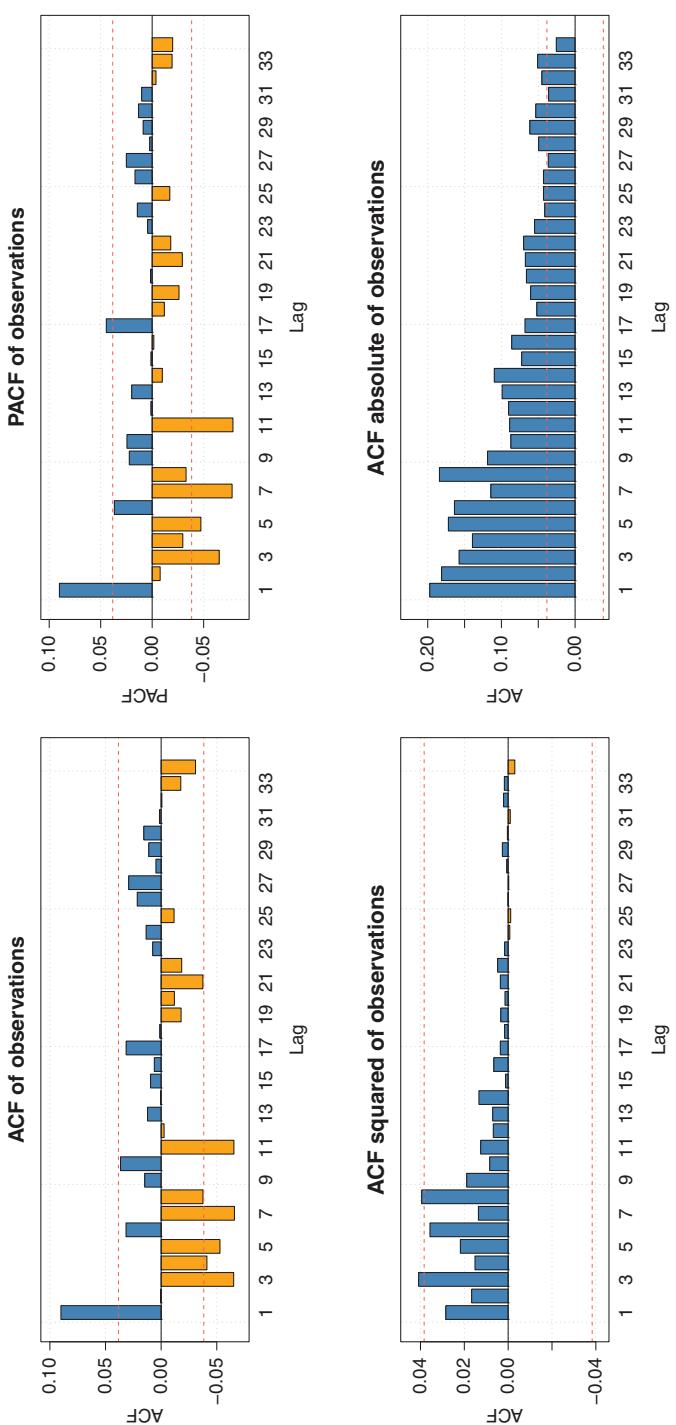
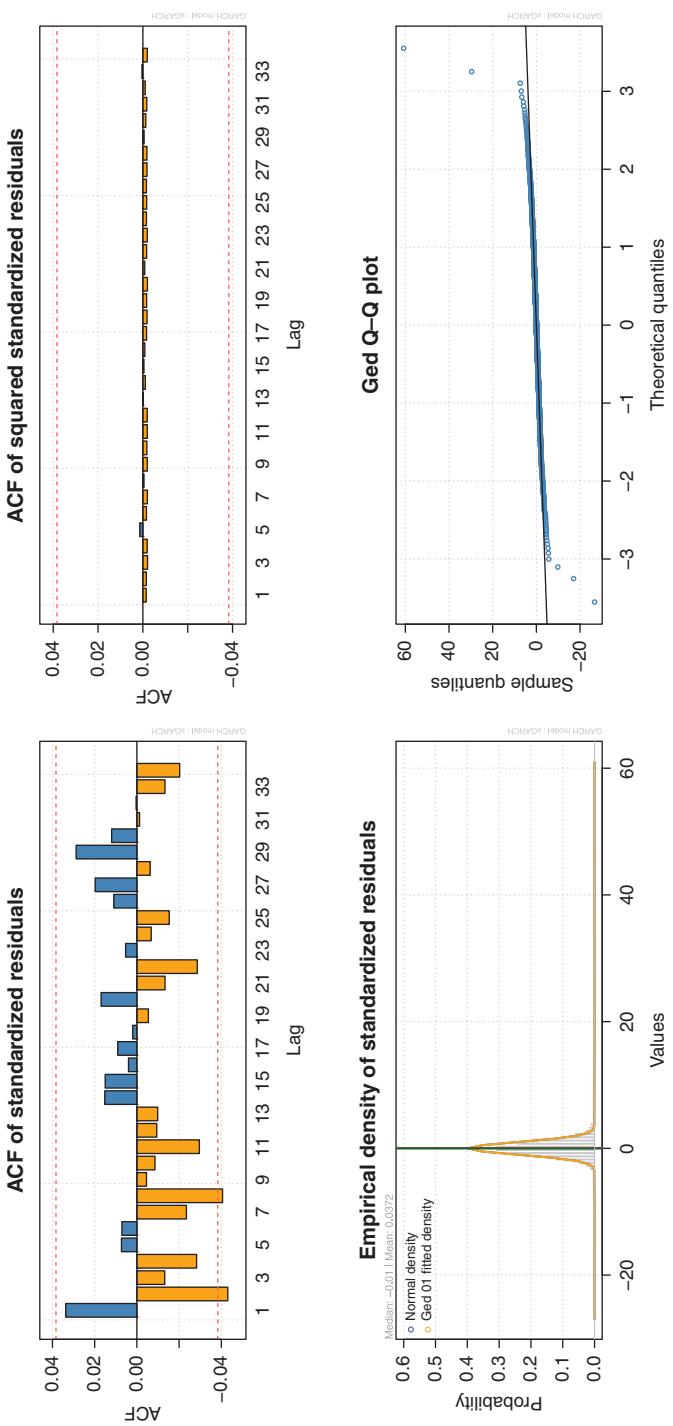


FIGURE 12.22 BAC high-frequency returns and squared and absolute returns ACF and PACF.



**FIGURE 12.23** BAC conditional volatility model diagnostics: On top are the ACF of the standardized and standardized squared residual, and on bottom are the distribution and GED Q–Q plot for the fitted model.

The behavior of stocks representing the financial institutions that were affected by the crisis such as JPM and BAC was studied. We also looked at other institutions such as IBM and WMT that should not have been much affected. All the stocks were sampled with the period  $T = 1$  min. During the days considered, the BAC and JPM return fluctuated more than 10% while that of WMT and IBM below 1%.

The data on earthquakes magnitude was obtained from the US Geological Survey (USGS) from 1 January 1, 1973, to November 9, 2010. The downloaded data contains information about the date, longitude, latitude, and the magnitude of each recorded earthquake. The location of the major earthquake and its distribution defines the studied area. The earthquake magnitude is the data used in the analysis. The sample data is from July 24, 2010, to November 9, 2010, containing 3000 observations with a maximum (7.1), which occurred on August 12, 2011.

We then continue our analyses with the study of seismic traces of a mining explosion. A seismic trace is a plot of the earth's motion over time. The data presented are measurements of the earth's vertical displacement where the recording frequency is 40 per second. The data sets are from a recording station in Scandinavia and are reported by Shumway and Stoffer [27].

## 12.6 Results and discussion

---

In this chapter, we performed different statistical analyses to evaluate the fit of temporal generic diversity and extinction data to different model distributions. From the fact that largest  $p$ -values in an AD test are associated with the distributions with the best fit, it was concluded that the best fit to the distribution of extinction magnitudes is the maximum extreme value distribution. The second best fit comes from the Gamma distribution, as one would expect from the approximation of extinction waiting times by an exponential.

In Figure 12.9, there are two solid lines representing cumulative distribution of diversity summed over time plotted against the empirical data. The  $y$ -axis represents cumulative distribution function and the  $x$ -axis is the absolute value of standardized changes in diversity. The dotted line represents the empirical distribution function of  $GS_T$  (see (12.15)); the absolute value of standardized changes in diversity. The solid green line is the

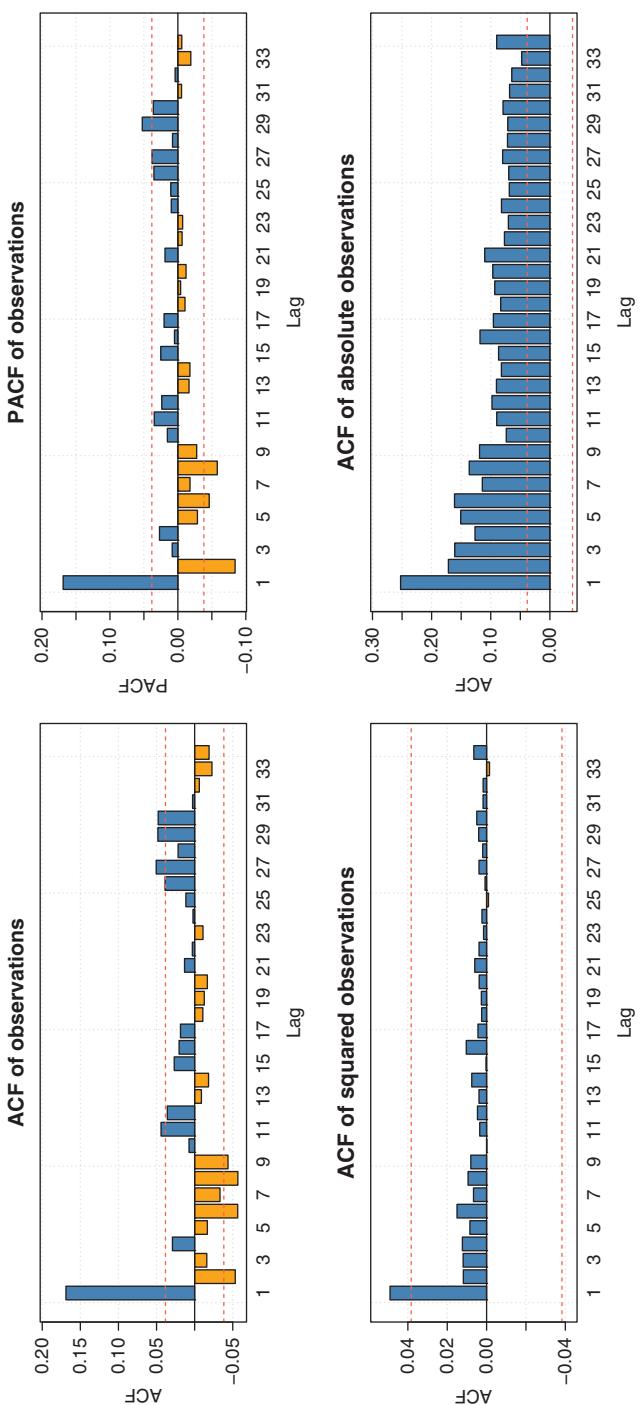
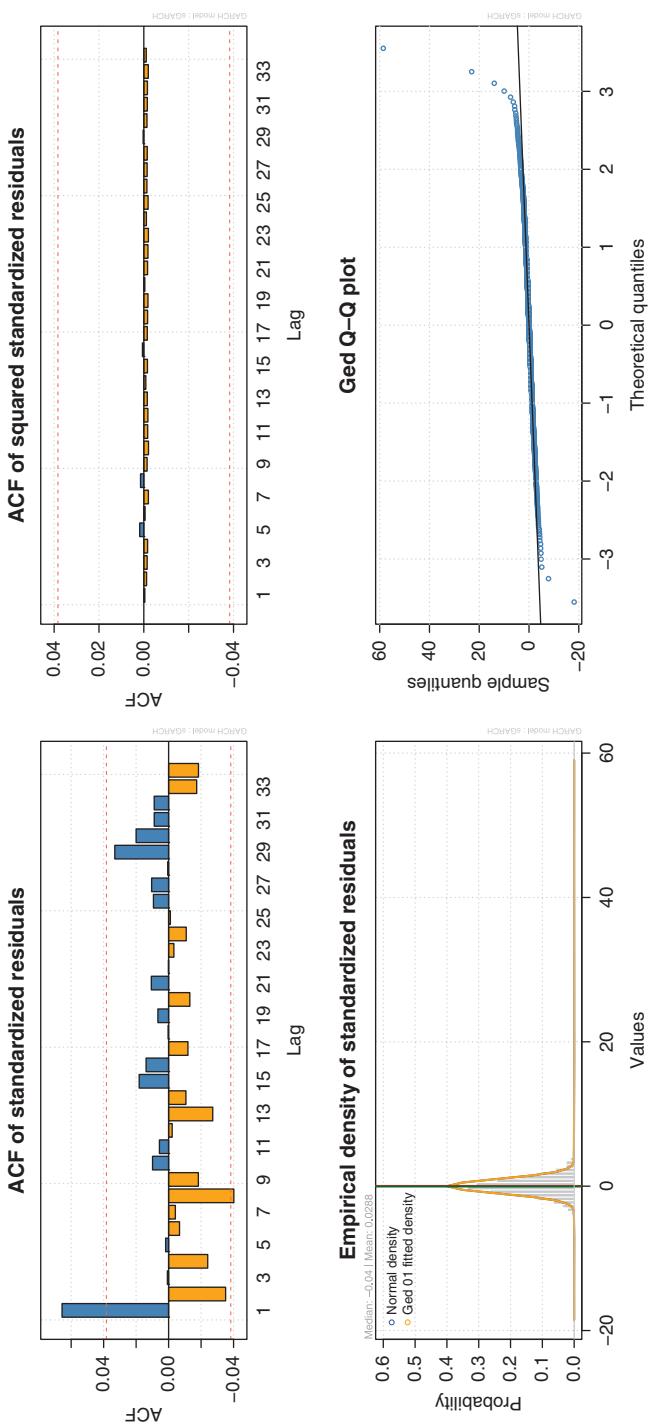


FIGURE 12.24 JPM high-frequency returns and squared and absolute returns ACF and PACF.



**FIGURE 12.25** JPM conditional volatility model diagnostics: On top are the ACF of the standardized and standardized squared residual, and on bottom are the distribution and GED Q–Q plot for the fitted model.

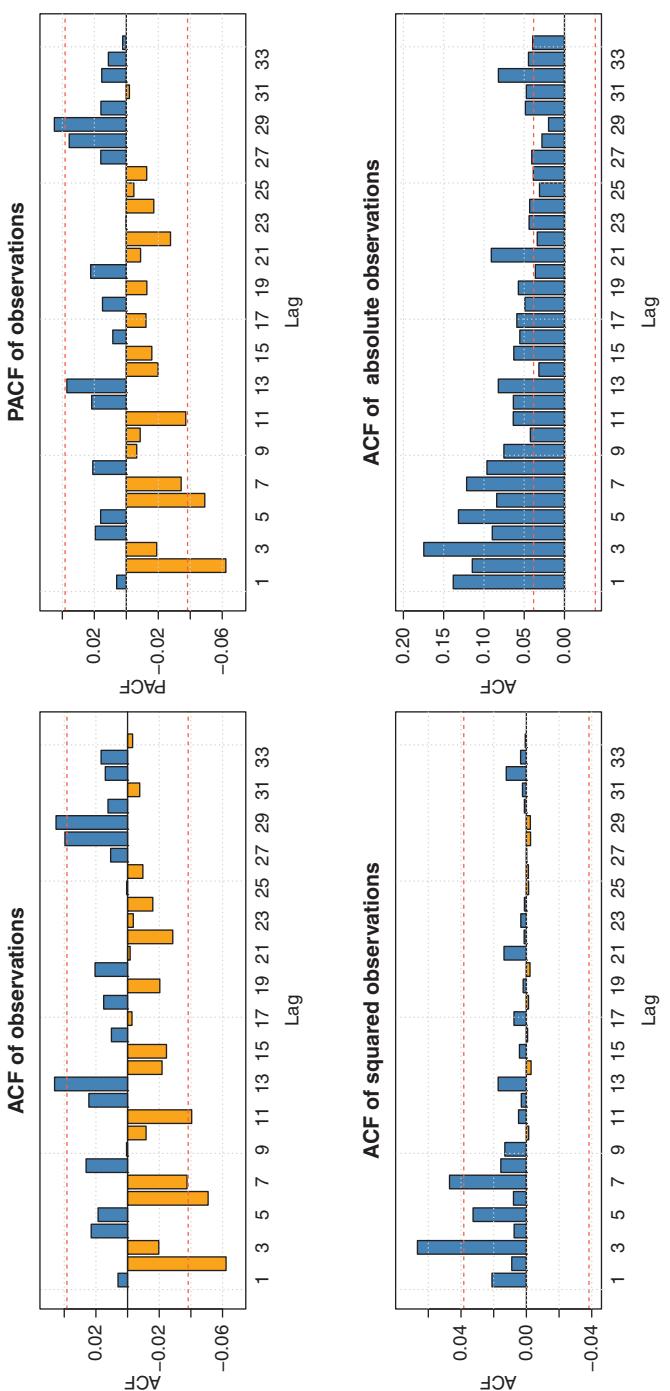
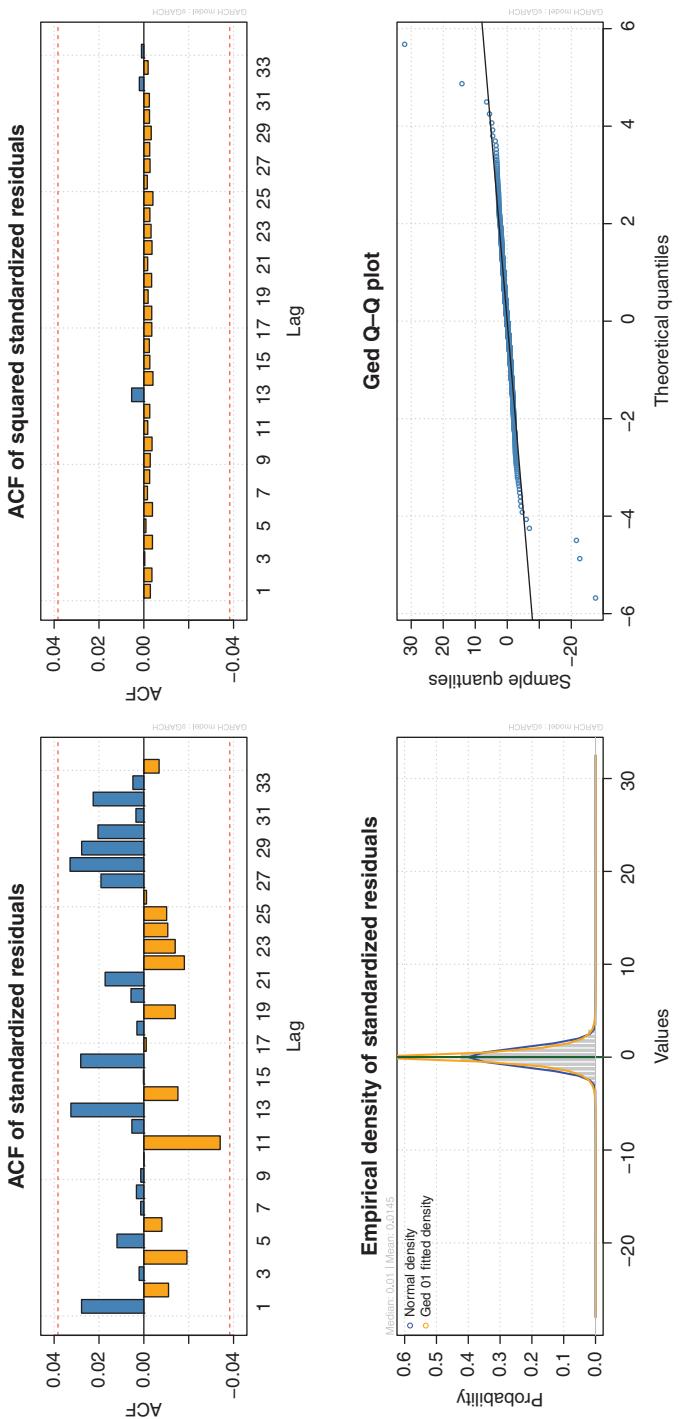


FIGURE 12.26 IBM high-frequency returns and squared and absolute returns ACF and PACF.



**FIGURE 12.27** IBM conditional volatility Model diagnostics: Top are the ACF of the standardized and standardized squared residual. Bottom are the distribution and GED Q–Q plot for the fitted model.

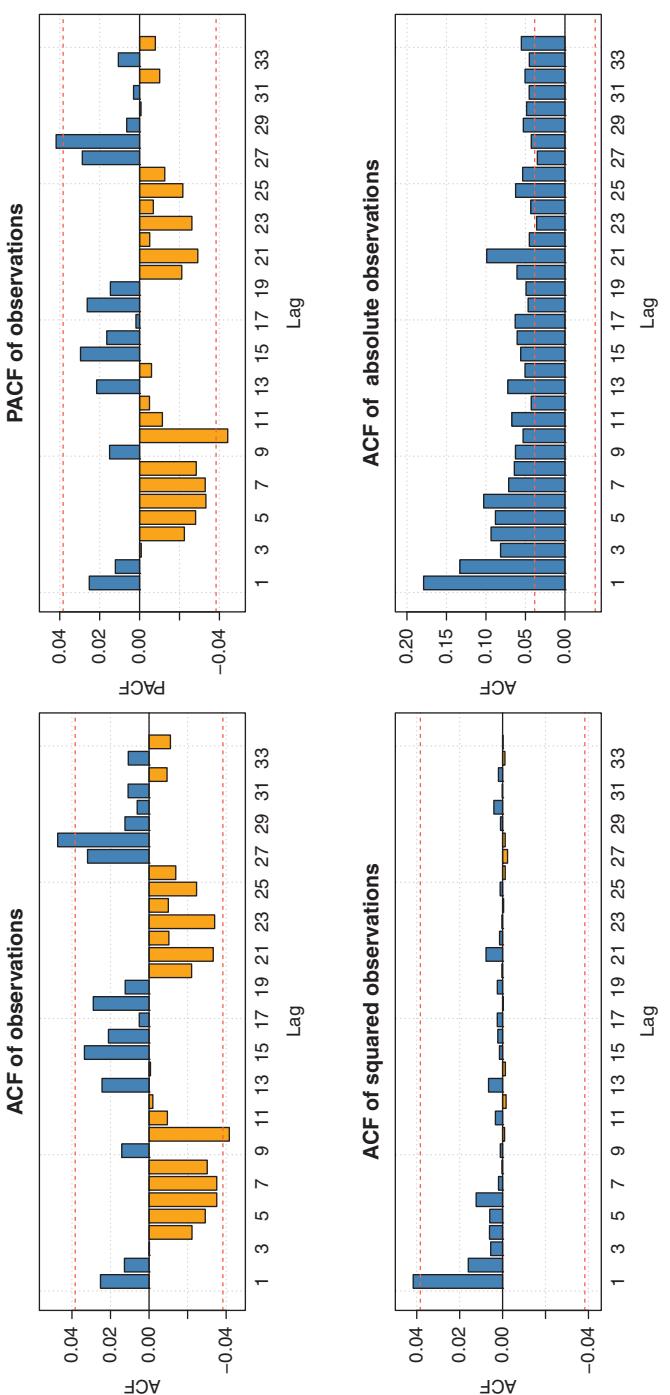


FIGURE 12.28 WMT high-frequency returns and squared and absolute returns ACF and PACF.

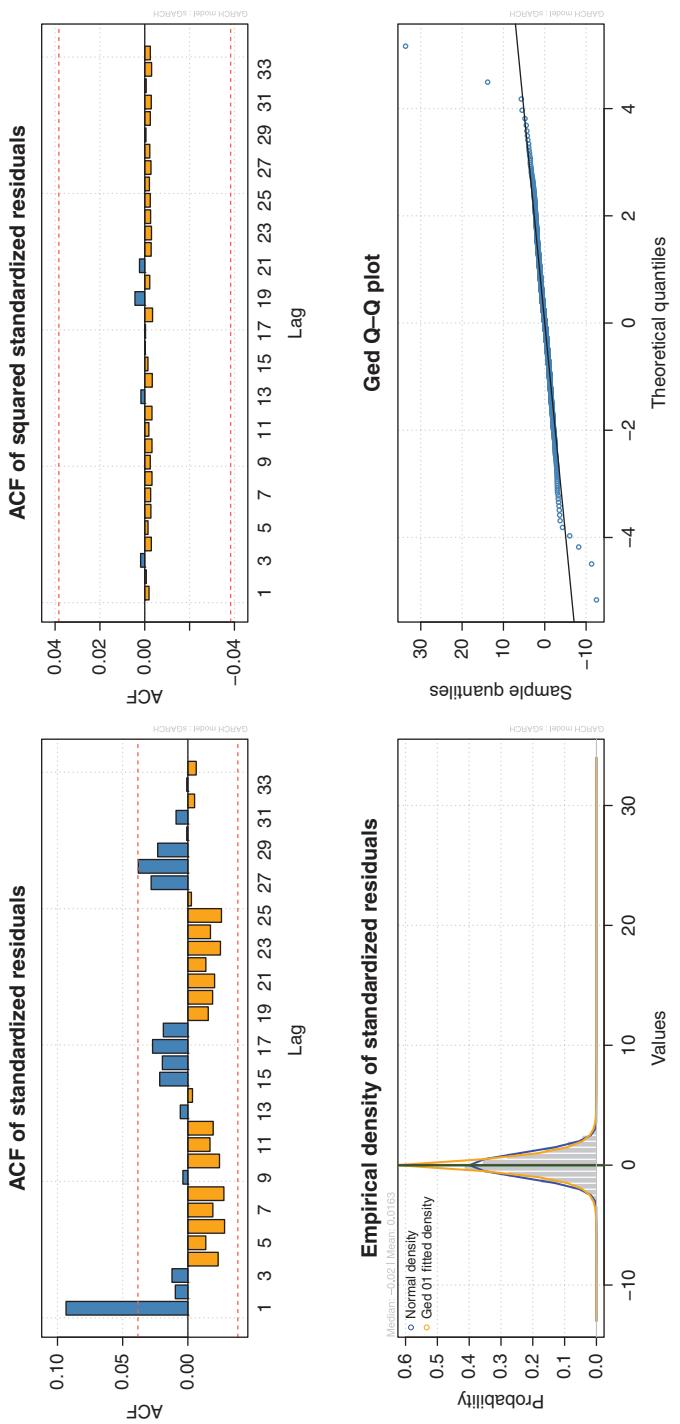


FIGURE 12.29 WMT high-frequency conditional volatility model diagnostics.

**TABLE 12.1** Results summary

Figure 12.9.	Time lag, $T$	$\alpha$	KS statistics	$p$ -value	TLF fit Kurt.	Sample Kurt.
a	1	1.646	0.0823	0.464	167	8.75
b	2	1.635	0.0698	0.68	176	13.51
c	4	1.745	0.0568	0.894	28.8	6.75
d	8	1.745	0.059	0.8767	14.8	8.59
e	16	2.0	0.1360	0.0664	0	4.229

cumulative distribution function corresponding to a standard normal random variable (zero mean, variance equal to 1). The red line represents the best fit for absolute value of random variable defined by the characteristic function (12.12). In other words, the best fit is provided by a standardized TLF.

Note that for small time lag  $T$  (consequent changes in diversity) the data are far from standard normal distribution and follow the truncated Lévy flights closely. For larger values of time lag  $T$ , the data are almost normally distributed, probably because interpolation over longer time periods “smooths out” shorter-term fluctuations in magnitude that lead to deviations from normality. Curiously, this is opposite to the behavior of stock market returns, where for large time lag we have Lévy stable distribution and with smaller time lag one obtains a close fit to a normal distribution [28, 29].

Table 12.1 shows the summary for Figures 12.9a, 12.9b, 12.9c, 12.9d, and 12.9e, together with the Kolmogorov–Smirnov (KS) test for hypothesis  $H_0 : GS_T$  has standardized TLF distribution. In the same table, we include the  $p$ -values for KS test and computed the kurtosis for TLF.

We can see from Table 12.1 that our sample ( $G_T$ ) has kurtosis above zero and modeling with normal distribution is not appropriate. The normal distribution can be easily rejected by the KS test. Note that the kurtosis for the TLF is also function of time lag  $T$ . For  $T \rightarrow \infty$ , we have that  $\gamma_2$  is of  $\mathcal{O}(T^{-2})$ . In Section 12.5, we presented a study of the statistical behavior of data arising in population models, of a financial index along with the rate of return of specific companies within the index, and of HFD by using a standardized truncated Lévy flight model. In all the cases, we obtained that the evolution of data can be described by the model. We can see that all the values obtained for the exponent  $\alpha$  are lower than 2.

In previous works (see, e.g., [30] and the references therein), it was found that the exponents calculated for market indices were strictly greater than 2. Weron [30] concluded that these values could be a consequence of working with finite samples. This behavior was compatible with a slow convergence to a Gaussian distribution but it was not possible to conclude that the Lévy distribution was the appropriate stochastic process for explaining the financial indices evolution. The authors believe that the standardized Lévy model that was used in this work, together with computation of the constants involved in the model, allowed them to more accurately complete a numerical analysis. This standardized Lévy model is suitable for better working with finite samples, and it offers a new way for analyzing financial indices, as well as other phenomena with similar behavior.

Figure 12.9 shows the log–log plot of the cumulative distribution of the normalized return for four different values of time scale. The red line is the best fit of the Lévy distribution. The green line indicates the Gaussian distribution. In the cumulative distribution curve of each fund, there are some outlying points. Those outlying points correspond to the significant drops in a very short period (1 or 2 days, or 1 or 2 min, depending on the data) that happened to the data. This is exactly the way in which a market crash is defined: A market crash is an outlying point in the cumulative probability distribution of the stochastic process described by the Lévy model. It should be noted that outlying points exactly reflect the crash of the corresponding financial indices/stocks.

Finally, the underlying volatility processes in earthquake series, high-frequency financial data, financial indices, and explosive data were explored using various GARCH models in this thesis. The GARCH models applied include basic GARCH, IGARCH, ARFIMA (0,d,0)-GARCH, and FIGARCH specifications. The methodology is not new; however, the major contribution of this work comes in the realm of applications. The methodology was applied to three domains: geophysics (earthquake data), finance (high-frequency financial data, and indices), and explosives data. In all the applications, the methodology provides insight into features of the series volatility.

The results show that the indices (DJIA and S&P 500) returns and the explosives (EXP) series volatility had the highest persistence that were best described by using FIGARCH, a long-memory model. This result is in line with the previous work of Breidt et al. [4], Mariani et al. [31],

and Mike So [32] for the indices volatility. In his work, Mike So [32] applied the modified rescaled range test (R/S) proposed by Lo (1991) and the semiparametric test (GPH) proposed by Geweke and Porker-Hudak [33] to detect the existence of long-term dependence in volatility in the S & P 500 index and DJIA index. He used three proxies of the variability of returns to achieve this result: the absolute mean deviation, the squared mean deviation, and the logarithm of the absolute mean deviation. Mariani et al. [31], in their paper, used the Hurst exponent and the Detrended fluctuation analysis methodology to show the existence of long-memory effects in the international market indices, that is, Morgan Stanley Capital International Europe, Australasia, and Far East index (MSCI EAFE), and the Emerging market index, and compare their result with S&P 500. They also found that immediately before the crisis the estimate of the long memory increases, while during the time of the crisis, the stocks behave randomly. Finally, Breidt et al. [4] also used their proposed long-memory stochastic volatility model to indicate the existence of long memory in financial indices. Hence, our results of long memory in the indices volatility reinforced previous results by different methodology.

The ARFIMA(0,d,0)-GARCH specification was preferred for BAC and JPM high-frequency series whose volatility was found to be intermediate. The intermediate memory found in the HFD is consistent with previous results of Barany et al. (2011) [34] and Mariani et al. (2009) [35] except with the WMT and IBM HFD volatility that was best described using the GARCH model, which has short memory. Mariani et al. (2009) [35] employed the relationship between the Hurst parameter ( $H$ ) and the Detrended fluctuation analysis parameter ( $\alpha$ ) and compared with the value 0.5 to investigate the memory behavior of BAC and JPM HFD from March 10 to March 18, 2008 (Bear Stearns financial crisis). Their results gave Hurst estimate of 0.63 and 0.62 for BAC and JPM, respectively, which implies long memory. Our values for the fractional difference parameter estimate 0.1 and 0.236 for the BAC and JPM, respectively, also showed long memory.

The earthquake series was divided into two regions: symmetrically (EQ1) and nonsymmetrically (EQ2) distributed. Both regions showed intermediate memory. Our analysis indicated that the earthquake series showed long memory while the explosive series showed short memory. On the other hand, both the explosive and the earthquake series volatility showed long memory, but the persistence in the explosive volatility was higher

than that of the earthquake. The order of persistence (memory) in series volatility from highest to lowest is as follows:

1. Indices (DJIA and S&P 500) and explosives data
2. Earthquake data
3. BAC and JPM high-frequency financial data
4. WMT and IBM high-frequency financial data

The predictions made from the  $MA(1) - FIGARCH(1, d, 1)$  model for the DJIA index offered good results since all the actual observations were within the prediction limits. The earthquake series predictions from our assumed model were fairly accurate since we had 8 out of 10 earthquake's directions correctly predicted.

The outliers observed from the generalized error Q–Q plot for the indices correspond to significant drops in prices in a short time period (1 or 2 days). This is what happened on October 15, 2008, for both DJIA and S&P 500 indices.

In Table 12.1, we have the following set of parameters:

1.  $\alpha$ —parameter  $\alpha$  as in (12.12).
2. KS statistics—value of the KS statistics.
3.  $P$ -value— $p$ -value for the KS test and hypothesis  $H_0$ : Data are following TLF distribution.
4. TFL fit kurtosis—kurtosis computed according to the formula (12.13).
5. Sample kurtosis—sample kurtosis computed in the following way:

$$\frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2)^2}$$

### 12.6.1.1 ADF test

**TABLE 12.2 Augmented Dickey–Fuller (ADF) unit root test values.**

$H_0: I(1)$ , critical values:  $-2.567$  (1%);  $-1.941$  (5%);  $-1.616$  (10%) for the indices, HFD, earthquake series, and the explosive series.

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
Return	-12.321	-12.493	-14.75	-14.328	-14.752	-14.325	-11.352	-18.232
Sqd. Ret	-4.7648	-4.833	-12.34	-13.001	-12.455	-12.456	-11.367	-5.628
Abs. Ret	-4.8911	-4.845	-9.193	-9.6733	-10.032	-10.032	-11.352	-4.361

### 12.6.1.2 KPSS test

**TABLE 12.3** Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) unit root test values.  $H_0: I(0)$ , critical values: 0.739 (1%); 0.463 (5%); 0.347 (10%) for the indices, HFD, earthquake series, and the explosive series.

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
Return	0.166	0.208	0.167	0.146	0.089	0.089	6.641	0.007
Sqd. Ret	1.549	1.724	0.108	0.099	0.075	0.075	6.850	0.867
Abs. Ret	2.544	2.968	0.979	1.271	0.598	0.598	6.641	1.233

### 12.6.1.3 ARCH test

**TABLE 12.4** AutoRegressive Conditional Heteroscedastic test for correlation in squared and absolute series. Critical values: 32.91 (0.1%); 26.22 (1%); 21.03 (5%) for the indices, HFD, earthquake series, and the explosive series.

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
ARCH LM	731.136	703.342	15.454	8.198	21.741	5.649	177.227	1299
Ljung. Box	2508.581	2539.780	18.621	8.838	23.145	5.938	364.196	4266
Pierce. Box	2498.418	2529.511	18.570	8.824	23.088	5.929	363.255	4251

### 12.6.1.4 Whittle estimate of d

**TABLE 12.5** Whittle estimate of the long-memory parameter ( $d$ ) for the indices, HFD, earthquake series, and the explosive series where  $d \in (-0.5, 0.5)$ .

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
Return	-0.083*	-0.085*	0.031	0.091**	-0.019	0.005	0.136**	0.499**
Sqd. Ret	0.199**	0.202**	0.042	0.042	0.038	0.035	0.140**	0.419**
Abs. Ret	0.218**	0.215**	0.183**	0.204**	0.145**	0.159**	0.136**	0.385**

\*significant at 5%; \*\*significant at 1%.

### 12.6.1.5 AIC

**TABLE 12.6** AIC of models fitted to the indices, HFD, earthquake series, and the explosive series. The model with minimum AIC is selected.

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
GARCH(1,1)	-6.544	-6.417	-11.1720	-11.042	-12.632	-12.713	2.508	-3.473
ARFIMA+GAR	-6.545	-6.420	-11.1783	-11.051	-12.623	-12.694	2.484	-3.484
FIGARCH(0,d,0)	-6.506	-6.375	-11.1750	-11.044	-12.624	-12.695	2.517	-3.472
FIGARCH(0,d,1)	-6.541	-6.417	-11.1780	-11.043	-12.623	-12.694	2.518	-3.471
FIGARCH(1,d,1)	-6.553	-6.429	-11.1781	-11.041	-12.622	-12.693	2.510	-3.657
IGARCH(1,1)	-6.544	-6.417	-11.1730	-11.043	-12.624	-12.695	2.507	-3.472

### 12.6.1.6 BIC

**TABLE 12.7 BIC of models fitted to the indices, HFD, earthquake series, and the explosive series. The model with minimum BIC is selected.**

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
GARCH(1,1)	-6.534	-6.407	-11.162	-11.031	-12.612	-12.682	2.517	-3.461
ARFIMA+GAR	-6.533	-6.407	-11.163	-11.038	-12.619	-12.701	2.495	-3.644
FIGARCH(0,d,0)	-6.499	-6.367	-11.169	-11.035	-12.615	-12.686	2.523	-3.463
FIGARCH(0,d,1)	-6.531	-6.406	-11.167	-11.032	-12.612	-12.682	2.526	-3.459
FIGARCH(1,d,1)	-6.541	-6.416	-11.165	-11.032	-12.609	-12.679	2.521	-3.469
IGARCH(1,1)	-6.537	-6.409	-11.164	-11.034	-12.614	-12.685	2.514	-3.464

### 12.6.1.7 *d* Estimate

**TABLE 12.8 Long-memory volatility models and their estimated long-memory parameter for the indices, HFD, earthquake series, and the explosive series.**

Symbol	DJIA	SP500	BAC	JPM	IBM	WMT	EQ2	EXP
ARFIMA+GAR	-0.032*	-0.045**	0.10**	.236**	.107**	0.152*	0.092**	0.861**
FIGARCH(0,d,0)	0.215**	0.214**	0.00	1.24	0.158	0.00	0.097	0.390**
FIGARCH(0,d,1)	0.307**	0.313**	0.00	-0.12	0.246	0.00	0.115	0.390**
FIGARCH(1,d,1)	0.602**	0.634**	0.00	0.00	0.443	0.00	0.331	0.782**

\*significant at 5%; \*\*significant at 1%.

## Acknowledgments

The authors thank Steve Wang for providing data sets on generic diversity in the fossil record. We are especially grateful to Dr. Ionut Florescu for having shared high-frequency data with us.

## 12.A Appendix A—Big ‘O’ notation

Big O notation describes limit behavior of a function. Either where we are taking limit toward infinity or some finite value. For example, if we say that

$$f(x) = \mathcal{O}(g(x)) \text{ as } x \text{ goes to infinity}$$

this precisely means that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = M, M < +\infty.$$

In other words,  $f(x)$  behaves like  $g(x)$  for large values of  $x$ .

Similarly we can think of an example in which  $\Delta x$  goes to 0:

$$f(\Delta x) = \mathcal{O}((\Delta x)^2) \text{ as } \Delta x \text{ goes to } 0$$

this precisely means that

$$\lim_{\Delta x \rightarrow 0} \frac{f(\Delta x)}{(\Delta x)^2} = M, M < +\infty.$$

In other words,  $f(\Delta x)$  behaves like  $(\Delta x)^2$  for  $x$  going to 0.

## References

1. J.J. Sepkoski (2001). *Mass Extinctions, Concept*. Encyclopedia of Biodiversity, Volume 4, Academic Press.
2. R.K. Bambach, A.H. Knoll, S. Wang (2004). *Origination, extinction, and mass depletions of marine diversity*. Paleobiology 30, 522–542.
3. S.M. Ross (2009). *Introduction to Probability Models*. Academic Press.
4. F.J. Breidt, N. Crato, P. De Lima (1998). *The detection and estimation of long memory in stochastic volatility*. Journal of Econometrics 83, 325–348.
5. P.M. Robinson. (1991) *Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression*. Journal of Econometrics 47, 67–84.
6. N. Shephard. (1996) *Statistical aspects of ARCH and stochastic volatility*. In D. R. Cox, D. B. Hinkley, and O. E. Barndorff-Nielsen, editors. *Time Series Models: In Econometrics, Finance and Other Fields*. Chapman Hall, London.
7. I.N. Lobato, N.E. Savin. (1998) *Real and spurious long-memory properties of stock-market data*. Journal of Business & Economic Statistics 16, 261–283.
8. R.T. Baillie. (1996) *Long memory processes and fractional integration in econometrics*. Journal of Econometrics 73, 5–59.
9. O.E. Barndorff-Nielsen, N. Shephard. (2001). *Modelling by Levy processes for financial econometrics*. In *Levy Processes*. Birkhauser, Boston, MA, pp. 283–318.

10. G.E.P. Box, G.M. Jenkins, G.C. Reinsel (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall.
11. R.F. Engle (1982). *Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation*. Econometrica, 50 (4), 987–1007.
12. T. Bollerslev (1986). *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics 31, 307–327.
13. J.M. Zakoian (1994). *Threshold heteroskedastic models*. Journal of Economic Dynamics and Control 18, 931–955.
14. R.F. Engle, V.K. Ng (1991). *Measuring and testing the impact of news on volatility*. Journal of Finance 48(5), 1749–1778.
15. D.B. Nelson. (1991). *Conditional heteroskedasticity in asset returns: A new approach*. Econometrica 59, 347–370.
16. A.C. Harvey, E. Ruiz, N. Shephard. (1994). *Multivariate stochastic variance models*. Review of Economic Studies 61, 247–265.
17. R.T. Baillie, T. Bollerslev, H. O. Mikkelsen. (1996). *Fractionally integrated generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics 74, 3–30.
18. P.M. Robinson, M. Henry (1999). *Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels*. Econometric Theory 15, 299–336.
19. R.F. Engle, T. Bollerslev (1986). *Modelling the persistence of conditional variances*. Econometric Reviews 5, 1–50, 81–87.
20. G. Casella, R.L. Berger (2002). *Statistical inference*, Duxbury advanced series in statistics and decision sciences. Thomson Learning.
21. W.J. Stewart (2009). *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*, Princeton University Press.
22. P. Lévy (1925). *Calcul des probabilités*, Gauthier-Villars, Paris.
23. A. Ya. Khintchine, P. Lévy (1936). Sur les lois stables, C. R. Acad. Sci. Paris 202, 374.
24. J. Voit (2005). *The Statistical Mechanics of Financial Markets*. Springer.
25. R.N. Mantegna, H.E. Stanley (1994). Physical Review Letters 73, 2946.

26. I. Koponen (1995). *Physical Review E* 52, 1197.
27. R.H. Shumway, D.S. Stoffer (1999). *Time Series Analysis and Its Application*. Springer Science, 2011.
28. M.C. Mariani, Y. Liu (2007, April 15). *Normalized truncated Levy walks applied to the study of financial indices*. *Physica A: Statistical Mechanics and Its Applications* 377(2), 590–598, ISSN 0378-4371, 10.1016/j.physa.2006.11.066.
29. R.N. Mantegna, H.E. Stanley (1999). *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge.
30. R. Weron (2001). *Levy-stable distributions revisited: Tail index 2 does not exclude the Levy-stable regime*. *International Journal of Modern Physics C* 12, 209–223.
31. M. Mariani, I. Florescu, M.P.B. Varela, and E. Ncheuguim (2010). *Study of memory effect in international market indices*. *Physica A* 2010, 389(8), 1653–1646.
32. M.P. So (2010). *Long-term memory in stock market volatility*. *Applied Financial Economics* 10(5), 519–524.
33. J. Geweke, S. Porter-Hudak (1983). *The estimation and application of long memory time series models*. *Journal of Time Series Analysis*, 4(4), 221–238.
34. E. Barany, M.P. Varela, I. Florescu, I. Sengupta (2011). *Detecting market crashes by analysing long-memory effect using high-frequency data*. *Quantitative Finance* 12(4), 515–518.
35. M. Mariani, I. Florescu, M.P.B. Varela, E. Ncheuguim (2009). *Long correlations and Levy models applied to the study of memory effects in high frequency (tick) data*. *Physica A* 388(8), 1659–1664.

# Index

*Note:* Page numbers followed by f and t refer to figures and tables, respectively.

- ABD test, 138–139, 142–146
  - list of studies on, 147t
- ACF, *see* autocorrelation function (ACF)
- additive regression model, 254
- ADF test, *see* augmented Dickey–Fuller test (ADF test)
- adiabatic quantum computing, 79–80
- age, generic diversity in fossil record, 372
- AIC of models, 335t, 420t
- algorithms, 77
  - classes, 78
  - complexity, 77–78
  - performance, 78–79
- ARCH effect, 311, 311t, 312f, 313, 334t, 420t
- ARFIMA(p,d,q)-GARCH(r,s) process, 299–300
- ARFIMA(p,d,q) model, 299
- ARMA(2,1) model, 154, 155t, 168t, 169
- ARMA(p,q) model, 297
- augmented Dickey–Fuller test (ADF test), 302–303, 334t, 419t
- autocorrelation function (ACF), 64, 65f
- backward-looking test, 139, 158–161, 160, 162–164
- Bai–Perron test, 64–65
  - of coal/WTI log prices ratio, 64f, 65
- Bayes methodology, 109
- BIC of models, 335t, 421t
- bicubic interpolation, 276–277
- biharmonic interpolation, 277, 278f–279f, 279, 280f–281f, 281
- bilinear interpolation, 272–276
  - algorithm, 272–274, 273f
  - nonlinear, 274, 275f–276f, 276
  - unit square, 274
- bipower variation (BP), 142
- BNS tests, 140–142
- BP, *see* bipower variation (BP)
- Brownian distance correlation, 62–63, 67–68
  - and Granger causality test, 68
  - significance level of, 67t
- Cauchy (Lorentz) distribution, 368
- CDO, *see* collateralized debt obligation (CDO)
- Chimera graph, 86
- classical optimization techniques, 74

---

*Handbook of High-Frequency Trading and Modeling in Finance*, First Edition.

Edited by Ionut Florescu, Maria C. Mariani, H. Eugene Stanley and Frederi G. Viens.

© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

- class NP algorithms, 78  
 class NP-complete algorithms, 78  
 class NP-hard algorithms, 78  
 class P algorithms, 78  
 collateralized debt obligation (CDO), 51–52  
     tranche loss function and, 52–53  
 combinatorial optimization, 74, 82–83  
 conditional mean model (returns), 309  
 correlation sensitivities, for tranche  
     loss, 55–58  
 critical region selection,  
     high-frequency prices jumps, 152–153  
 cumulative sum (CUSUM) timing  
     examples, 12–24, 12f–15f  
     lazy simple random walk, 21–24, 23f, 25–26  
     random walk expected gain over  
         subperiod, 15–18  
     simple random walk, 18–20, 19f  
 overview, 1–3  
 process, 7–10  
 scheme, 10–12, 11f  
 stopping time, 7–10  
 two-sided, 10  
 CUSUM statistic process, 7–10  
 CUSUM stopping rule, 8–9  
 CUSUM strategy Monte Carlo, 24–27
- data-driven testing procedure, for  
     high-frequency prices jumps  
     microstructure noise, 146, 148–149  
     SPY data, 146, 148–149, 148f
- d* Estimate, long-memory volatility  
     models, 335t, 421t
- deterministic model, governing  
     equations for, 342–345  
     application to geophysical  
         (earthquake data), 343–344, 346f–352f, 353t  
     results, 344–345, 345f
- distance correlation, 63–64  
 distance covariance, 63  
 diversified portfolio, 82  
 diversity, in fossil record, 372  
 DJIA, *see* Dow Jones Index (DJIA)  
 Dow Jones Index (DJIA) returns, 306, 307f–309f  
 D-wave simulator, 98  
 D-wave system, 78–79  
     mapping for, 86–87  
     results from, 98
- electricity generation from fossil  
     fuels, 68–69, 68t  
 embedding, 87, 87f  
 Expectation Maximization (EM)  
     algorithm, 112  
 explosive series, analysis of, 394, 395f, 396, 397f–408f, 409, 410f–415f  
*ext*, generic diversity in fossil record, 372
- filtered empirical distributions at  
      $t_1, \dots, t_T$ , 110–112  
     evolution step for, 111  
     selection step, 112
- financial model for optimization, 81  
 forward-looking tests, 139, 158  
 fossil fuels prices, 61–62  
     Bai–Perron test for, 64–66  
     data for, 64, 64f–65f  
     electricity generation and, 68–69, 68t  
     log prices, correlation matrix of, 66t  
 fractional Brownian motion, 224–226  
 fractional IGARCH (FIGARCH)  
     model, 296, 300–302, 373
- functional central limit theorem  
     for multivariate linear Hawkes  
         processes, 190  
     for univariate nonlinear Hawkes  
         processes, 190

- gain over subperiod, 5–7  
calculation for, 6–7  
random walk expected, 15–18
- GARCH(p,q) model, 296–298, 373  
geophysics and finance using, 295–337
- Gaussian copula model, 51–52  
tranche loss function and, 52–53
- Gaussian inequalities, 54–58
- generalized extreme value (GEV) distribution, 139–140, 152–153
- generalized method of moments, 196
- generic diversity in fossil record, analysis of introduction, 371–373
- Johnson transformation (JT) function for, 378, 379f–380f, 380, 381f–383f
- Lévy distribution for, 380  
data analysis with TLF distribution, 389–390, 390f
- stable distributions, characterization of, 383–384
- sum of random variables with different parameters, 390–394, 391f–393f
- truncated, 384–389
- statistical and numerical analysis, 377–380, 378f
- statistical preliminaries and results, 373–377  
general version for  $n$  random variables, 376–377  
sum of exponential random variables with different parameters, 374–377
- genetic algorithm formulation, 76
- geophysics and finance using GARCH models, volatility structures in, 295–337
- data collection, analysis, and result, 306–335
- ARCH effect, 311, 311t, 312f, 313
- ARMA(2,2), 309, 310f, 311
- conditional mean model (returns), 309
- on Dow Jones Index (DJIA) returns, 306, 307f–309f
- on high-frequency, earthquake, and explosives series, 320f–329f, 330, 331f–332f, 333
- model diagnostic of conditional returns with conditional variance, 318, 319f, 330t
- model diagnostics, 314–315, 315f
- model selection and specification, 306, 309t, 311t, 313, 313t
- one-step ahead prediction of last 10 observations, 330, 330t
- returns and variance equation, 315, 316f, 317, 317t
- standardized residuals test, 314, 314f, 317–318, 318f
- long memory models, 298–305
- ARFIMA(p,d,q)-GARCH(r,s), 299–300
- ARFIMA(p,d,q) model, 299
- detection and estimation of, 302–305
- short memory models, 297–298
- ARMA(p,q) model, 297
- GARCH(p,q) model, 297–298
- IGARCH(1,1) model, 298
- geophysics (earthquake data) deterministic model, application to, 343–344
- Lévy flights application to, 345, 353
- MATLAB program code description for, 362–364
- using GARCH models, 295–337

- Granger causality, 62–63, 65, 68  
and Brownian distance correlation, 68  
definition of, 63  
significance level of, 67  
testing of, 63  
vector autoregressive (VAR) model for, 67
- graph-theoretic combinatorial optimization models, 82–83
- Grey relational analysis approach, 76
- Gumbel distribution, 139, 144–145, 145t
- HAR-MA model, least-square estimation of, 167t, 168
- Hawkes processes, 186–191  
applications of, 198–206  
measuring endogeneity (reflexivity), 205–206  
modeling jump-diffusion, 205  
modeling order arrivals, 199–200  
modeling price jumps, 200–204, 202f
- branching structure representation, 188
- brief history of, 211–212
- convergence of, 189  
functional central limit theorem for multivariate linear Hawkes processes, 190
- functional central limit theorem for univariate nonlinear Hawkes processes, 190
- law of large numbers for multivariate linear Hawkes processes, 189–190
- nearly unstable univariate linear Hawkes processes, 191
- estimation, 194–197  
expectation maximization, 194–196
- generalized method of moments, 196
- maximum likelihood estimation, 194
- nonparametric, 196–197
- in high frequency financial data modeling, 211–212
- hypothesis testing of, 197–198  
approximate thinning, 198  
random time change, 197–198
- linear, 187–188
- multivariate marked, 186–187
- simulation, 191–193  
by branching structure, 193  
inverse CDF transform, 192  
Ogata's modified thinning, 192–193
- stationarity, 188–189
- statistical inference of, 191–198
- Hermite multifractal random walk (HMRW), 231–234  
simulation of, 235–241, 239f–240f
- Hermite processes, 223  
financial applications, 234–243  
simulation of HMRW, 235–241, 239f–240f  
statistical properties, 241–243, 243f–246f
- fractional Brownian motion and, 224–226
- infinitely divisible cascading noise, 229–231
- multiplicative random walk driven by, 231–234  
definition, 231–233  
existence, 231–233  
properties of, 233–234
- properties of, 225–226
- Wiener integrals with respect to, 226–228
- Hermite random variable, 225

- heuristic optimization techniques, 74, 76
- hidden Markov process, 108
- high-frequency market data, application to methodology, 360–361, 362f results, 361
- high-frequency prices jumps, stochastic volatility, 137–175
- data-driven testing procedure microstructure noise, 146, 148–149
- SPY data, 146, 148–149, 148f
- empirical results, 161–165 backward-looking test, 162–164, 162t, 163f, 164t
- interpolated test, 165
- generalized testing procedure critical region selection, 152–153
- spot volatility estimation, 149–152, 149f, 150t–151t
- intraday jump tests
- ABD test, 142–146
  - BNS test, 140–142
  - LM test, 142–146
  - realized volatility measure test, 140–142
- simulation study
- model specification, 153–158, 155t–156t, 157f
  - results, 158–161, 159t–160t
- high-frequency (tick) data, analysis of, 394, 395f, 396, 397f–408f, 409, 410f–415f
- HMRW, *see* Hermite multifractal random walk (HMRW)
- idle time, of trading strategy, 25
- index of stability, 383
- infinitely divisible cascading noise, 229–231
- integrated GARCH (IGARCH) model, 296, 298, 373
- integrated volatility (IV), 141
- intermediate memory process, 300
- interpolated test, 160–161, 165
- interpolation methods, 271–287
- bicubic interpolation, 276–277
  - biharmonic interpolation, 277, 278f–279f, 279, 280f–281f, 281
- bilinear interpolation, 272–276
- algorithm, 272–274, 273f
  - nonlinear, 274, 275f–276f, 276
  - unit square, 274
- nearest-neighbor interpolation, 271
- numerical applications of, 285–287
- thin plate splines, 282–285
- physical analogy, 282
  - radial basis function, 283, 284f–285f
  - smoothness measure, 283
- intraday price jump tests
- ABD test, 142–146
  - BNS test, 140–142
  - LM test, 142–146
  - realized volatility measure test, 140–142
- intraday volatility pattern (IVP), 140
- inverse CDF transform, 192
- Ising model, 79, 83
- IV, *see* integrated volatility (IV)
- IVP, *see* intraday volatility pattern (IVP)
- Johnson transformation (JT) for diversity data, 378, 379f–383f
- kernel function, 111
- Kolmogorov–Smirnov (KS) test, 164, 416
- KPSS test, 303–304, 304t, 334t, 420t

- Lagrangian relaxation, 76  
 law of large numbers for multivariate linear Hawkes processes, 189–190  
 lazy simple random walk, CUSUM timing, 21–24, 23f, 25–26, 41t, 42t  
 least-square estimation of HAR-MA model, 167t, 168  
 Lévy distribution, for analysis of generic diversity in fossil record, 380  
 data analysis with TLF distribution, 389–390, 390f  
 stable distributions, characterization of, 383–384  
 sum of random variables with different parameters, 390–394, 391f–393f  
 truncated, 384–389  
 Lévy flights model  
 application to geophysics, 345, 353  
 to estimate crash dates, 366, 366t  
 results, 356–360, 357f–359f, 360t  
 truncated distribution, 353–356  
 Lévy-Smirnov distribution, 368  
 Limit Order Books (LOB), 183  
 characteristics of, 183–184  
 resiliency of, 200  
 Linear Hawkes process, 187–188  
 convergence of nearly unstable univariate, 191  
 functional central limit theorem for multivariate, 190  
 law of large numbers for multivariate, 189–190  
 LM test, 138–139, 142–146  
 list of studies on, 147t  
 LOB, *see* Limit Order Books (LOB)
- Local polynomial regression models, 255  
 multiple regression, 256  
 simple regression, 255  
 Loess curve, 257  
 long-memory generalized autoregressive conditionally heteroskedastic (LMGARCH) models, 296  
 long memory models, 298–302  
 ARFIMA(p,d,q)-GARCH(r,s), 299–300  
 ARFIMA(p,d,q) model, 299  
 detection and estimation of, 302–305  
 augmented Dickey–Fuller test (ADF test), 302–303  
 KPSS test, 303–304, 304t  
 Whittle method, 304–305  
 Lowess curve, 257  
 Lowess/loess method, nonparametric regression models, 257–258  
 marked point process (MPP), 185  
 Markov chain, 107  
 continuous time, 108  
 obtaining parameters of, 112–113  
 parameter estimation, 131  
 Markov processes, 187  
 Markowitz mean-variance model, 73  
 Matlab code  
 in geophysics, 362–364, 363f–364f  
 WMIS, 100–103  
 maximum independent set (MIS), 75, 82–83  
 maximum likelihood estimation (MLE), 194  
 minimum error thresholding method, 112  
 minorize-maximization (MM) algorithm, 197

- MIS, *see* maximum independent set (MIS)
- mixed models, 84
- MLE, *see* maximum likelihood estimation (MLE)
- Monofractal process, 222
- Monte Carlo, CUSUM strategy, 24–27
- MPP, *see* marked point process (MPP)
- MRW, *see* multifractal random walk (MRW)
- multifractality, 222
- multifractal random walk (MRW)
- defined, 222
  - by Hermite processes, 231–234
  - definition, 231–233
  - existence, 231–233
  - properties of, 233–234
  - overview, 221–224
  - properties, 223
  - statistical properties of, 241–243
- multivariate marked Hawkes process, 186–187
- multivariate point process, 185, 199
- nearest-neighbor interpolation, 271
- negative signals, 4
- nonparametric estimation of Hawkes process, 196–197
- nonparametric regression models, 253–271
- application to
  - financial data sampled with high frequency, 260, 266–269
  - geophysics, 259–260, 261f–265f, 266t–268t, 269f–270f
  - highlights and discussions, 270–271, 270t
  - generalized, 254
  - local polynomial regression, 255
- multiple regression, 256
- simple regression, 255
- lowess/loess method, 257–258
- NP algorithms, 78
- NP-complete algorithms, 78
- NP-hard algorithms, 78
- Ogata’s modified thinning method, 192–193
- optimization, portfolio, 76–77
- background of, 75–80
  - binary optimization on, 77
  - discussion, 95–97
  - diversified, 82
  - financial model for, 81
  - future research on, 98–100
  - graph-theoretic combinatorial models for, 82–83
  - hardware limitations in, 97
  - implementation limitations in, 98
  - Ising model for, 83
  - methods for, 84–88
  - input data, 85
  - mapping implementation, 86–88, 87f
  - mean-variance calculations, 85–86
  - model implementation, 85
  - risk measures implementation, 86
- mixed models for, 84
- model limitations in, 97–98
- models, 80–84
- QUBO model for, 83
- results of, 88–95
- restricted minimum-risk model, 91–94, 92f–93f
  - simple correlation model, 88–91, 89f–91f
- WMIS minimum-risk, max return model, 94–95, 94f–95f

- Orig, generic diversity in fossil record, 372
- Ornstein–Uhlenbeck (OU) process, 153, 155t
- P algorithms, 78
- Phanerozoic, 372
- point processes, 184–186. *See also* Hawkes processes
- definition, 207
  - marked, 185, 209
  - moments, 208
  - multivariate, 185
- Poisson processes, 186
- random time change, 211
  - stochastic intensities and, 185–186, 209–210
- Poisson processes, 186, 198
- portfolio optimization, 76–77
- background of, 75–80
  - binary optimization on, 77
  - discussion, 95–97
  - diversified, 82
  - financial model for, 81
  - future research on, 98–100
  - graph-theoretic combinatorial models for, 82–83
  - hardware limitations in, 97
  - implementation limitations in, 98
  - Ising model for, 83
  - methods for, 84–88
    - input data, 85
    - mapping implementation, 86–88, 87f
    - mean-variance calculations, 85–86
    - model implementation, 85
    - risk measures implementation, 86
  - mixed models for, 84
  - model limitations in, 97–98
- models, 80–84
- QUBO model for, 83
- results of, 88–95
- restricted minimum-risk model, 91–94, 92f–93f
  - simple correlation model, 88–91, 89f–91f
  - WMIS minimum-risk, max return model, 94–95, 94f–95f
- positive signals, 3–4
- proximal interpolation, *see* nearest-neighbor interpolation
- quadratic unconstrained binary optimization (QUBO), 74, 83
- quadratic variation (QV), of price, 141
- QUBO, *see* quadratic unconstrained binary optimization (QUBO)
- random walk expected gain over subperiod, 15–18, 43t–46t
- realized volatility measure test, 140–142
- regime switching volatility model, 107–133
- restricted minimum-risk model, 91–94, 92f–93f
- Rosenblatt process, 223, 225
- scale invariance model, application to methodology, 360–361, 362f results, 361
- scale invariance model, application to high-frequency market data methodology, 360–361, 362f results, 361
- scatterplot smoothing, *see* nonparametric regression models
- short memory models, 297–298
- ARMA(p,q) model, 297

- GARCH(p,q) model, 297–298  
 IGARCH(1,1) model, 298  
 Sidàk correction, 144, 145t  
 signals, 3–5  
   negative, 4  
   positive, 3–4  
   properties, 4–5  
   sequence of, 4  
   and subperiod, 5–7  
 simple correlation model, 88–91,  
   89f–91f  
 simple random walk, CUSUM timing,  
   18–20, 19f  
 simulation  
   of Hawkes processes, 191–193  
   by branching structure, 193  
   inverse CDF transform, 192  
   Ogata's modified thinning,  
   192–193  
   of HMRW, 235–241, 239f–240f  
 speed of reaction, CUSUM, 39  
 spot volatility estimation,  
   high-frequency prices jumps,  
   149–152, 149f, 150t–151t  
 square-root effect, threshold  
   parameter, 38  
 stable distributions  
   characterization of, 367–368  
   introduction of, 366–367  
 standard maximum likelihood  
   method, 153  
 stochastic intensities, point processes  
   and, 185–186  
 stochastic volatility model, 107–133  
   in analysis of  
     during earthquake, 125–127,  
       125f–126f  
     earthquake signal, 123–124,  
       123f–124f, 125f  
     end of earthquake signal,  
       aftershocks, 127  
   seismometer readings during  
     earthquake, 121–123,  
       122f  
   empirical testing of, 132–133  
   in finance area, 109–110, 113–119,  
       115f–118f  
   with hidden Markov process,  
       108–133  
   historical note on, 109–110  
   jumps in high-frequency prices  
       under, 137–175  
   methodology, 110–113  
   obtaining filtered empirical  
     distributions at  $t_1, \dots, t_T$ ,  
       110–112  
   obtaining parameters of Markov  
     chain, 112–113  
   in physical data application,  
       119–121, 120f, 120t  
   and problem, 108–109  
   theoretical results of  
     about convergence, 129–131  
     about parameter estimation,  
       129–131  
   Markov chain parameter  
     estimation, 131  
   particle filter, working of,  
       128–129  
   stock indices, analysis of, 394, 395f,  
       396, 397f–408f, 409,  
       410f–415f  
   stopping time, CUSUM, 7–10  
   subperiod, 4  
     gain over, 5–7  
     signal and, 5–7  
 SV2FJ\_2 $\rho$  model, 154, 156, 157f  
   calibration of  $\xi$  under, 170–175,  
       170t–175t  
   minimized loss function loss for,  
       169, 169t  
   parameter values of, 156t

- thin plate splines (TPSs), 282–285  
physical analogy, 282  
radial basis function, 283,  
284f–285f  
smoothness measure, 283  
threshold GARCH (TGARCH)  
model, 296, 373  
TPSs, *see* thin plate splines (TPSs)  
tranche, 53  
tranche loss function, 52–53  
correlation sensitivities for,  
55–58  
trend-based trading strategy, 3–7  
gain over subperiod, 5–7  
signaling, 3–5  
threshold parameter, effect of,  
27–28, 28f, 29t–33t, 34, 34f,  
35t–37t, 38–39  
trends, 3–5  
for US treasury notes, 11–12,  
11f–12f  
trends, 3–5  
truncated Lévy flights (TLF)  
distribution, 353–356,  
384–389  
data analysis with, 389–390,  
390f  
infinite divisibility, 388–389  
kurtosis, 387–388  
two-sided CUSUM (2-CUSUM), 10  
US Clean Air Act, 69  
US treasury notes, trading strategy  
for, 11–24, 11f–12f  
variance–covariance matrix, 85–86  
vector autoregressive (VAR) model,  
67  
volatility, 113  
clustering, 222  
structures in geophysics and  
finance using GARCH  
models, 295–337  
data collection, analysis, and  
result, 306–335  
long memory models, 298–305  
short memory models, 297–298  
weighted MIS (WMIS), 82–83, 88  
Matlab Code, 100–103  
White/Terasvirta test, 67t  
Whittle estimate of long-memory  
parameter ( $d$ ), 420t  
Whittle method, 304–305, 334t  
Wiener–Hopf equation, 197  
Wiener integrals with respect to  
Hermite processes, 226–228  
Wiener process, 224–225  
WMIS, *see* weighted MIS (WMIS)  
WMIS minimum-risk, max return  
model, 94–95, 94f–95f

Wiley Handbooks in

## FINANCIAL ENGINEERING AND ECONOMETRICS

Advisory Editor

**Ruey S. Tsay**

*The University of Chicago Booth School of Business, USA*

---

The dynamic and interaction between financial markets around the world have changed dramatically under economic globalization. In addition, advances in communication and data collection have changed the way information is processed and used. In this new era, financial instruments have become increasingly sophisticated and their impacts are far-reaching. The recent financial (credit) crisis is a vivid example of the new challenges we face and continue to face in this information age. Analytical skills and ability to extract useful information from mass data, to comprehend the complexity of financial instruments, and to assess the financial risk involved become a necessity for economists, financial managers, and risk management professionals. To master such skills and ability, knowledge from computer science, economics, finance, mathematics and statistics is essential. As such, financial engineering is cross-disciplinary, and its theory and applications advance rapidly.

The goal of this Handbook Series is to provide a one-stop source for students, researchers, and practitioners to learn the knowledge and analytical skills they need to face today's challenges in financial markets. The Series intends to introduce systematically recent developments in different areas of financial engineering and econometrics. The coverage will be broad and thorough with balance in theory and applications. Each volume will be edited by leading researchers and practitioners in the area and covers state-of-the-art methods and theory of the selected topic.

### Published Wiley Handbooks in Financial Engineering and Econometrics

Bauwens, Hafner, and Laurent · *Handbook of Volatility Models and Their Applications*

Brandimarte · *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics*

Chan and Wong · *Handbook of Financial Risk Management: Simulations and Case Studies*

Cruz, Peters, and Shevchenko · *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*

Florescu, Mariani, Stanley, and Viens · *Handbook of High-Frequency Trading and Modeling in Finance*

James, Marsh, and Sarno · *Handbook of Exchange Rates*

Peters and Shevchenko · *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*

Viens, Mariani, and Florescu · *Handbook of Modeling High-Frequency Data in Finance*

Szylar · *Handbook of Market Risk*  
Veronesi · *Handbook of Fixed-Income Securities*

### Forthcoming Wiley Handbooks in Financial Engineering and Econometrics

- Bali and Engle · *Handbook of Asset Pricing*  
Chacko · *Handbook of Credit and Interest Rate Derivatives*  
Jacquier · *Handbook of Econometric Methods for Finance: Bayesian and Classical Perspectives*  
Longin · *Handbook of Extreme Value Theory and Its Applications to Finance and Insurance*  
Starer · *Handbook of Equity Portfolio Management: Theory and Practice*  
Szylar · *Handbook of Hedge Fund Risk Management and Performance: In a Challenging Regulatory Environment*  
Szylar · *Handbook of Macroeconomic Investing*