

WILEY HANDBOOKS IN  
FINANCIAL ENGINEERING  
AND ECONOMETRICS



# HANDBOOK OF HIGH-FREQUENCY TRADING AND MODELING IN FINANCE

EDITED BY

Ionut Florescu  
Maria C. Mariani  
H. Eugene Stanley  
Frederi G. Viens

WILEY



# **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, Frederi 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*

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

## **2 Gaussian Inequalities and Tranche Sensitivities** *51*

*Claas Becker and Ambar N. Sengupta*

- 2.1** Introduction *51*
- 2.2** The tranche loss function *52*
- 2.3** A sensitivity identity *54*
- 2.4** 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



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

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

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

---

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

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

# 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 Invivo, 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*

## 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

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:

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

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,  $G_l$  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:

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

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:

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

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

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

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  $\{P_\nu\}$ ,  $\nu \in \mathbb{N}^*$ . In other words, the probability measure  $P_\nu$  for a given  $\nu > 0$ , 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  $\nu$ . 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  $\nu$ .

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

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

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

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

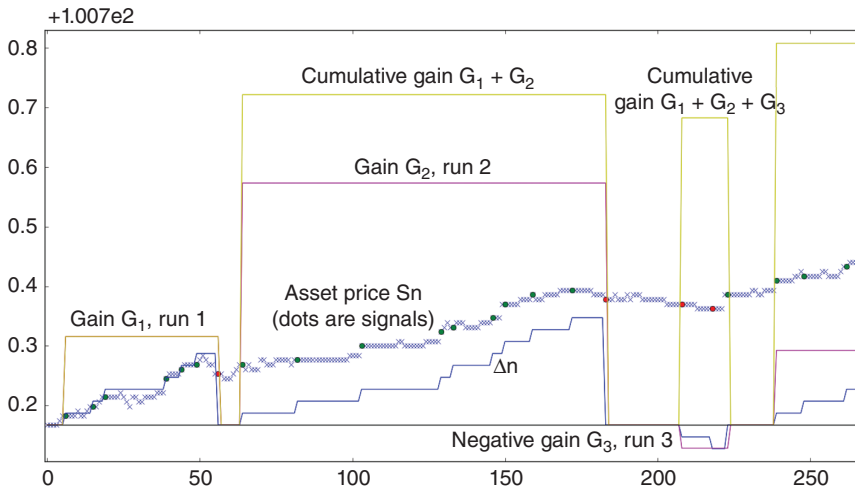
Thus, for  $k \geq 0$ , the CUSUM timing scheme for our trend-following trading strategy is defined by using (1.20) and (1.21) (and coming from (1.1) and (1.2)),

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

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.

