Preface

The idea of writing this book arose in 2000 when the first author was assigned to teach the required course STATS 240 (Statistical Methods in Finance) in the new M.S. program in financial mathematics at Stanford, which is an interdisciplinary program that aims to provide a master’s-level education in applied mathematics, statistics, computing, finance, and economics. Students in the program had different backgrounds in statistics. Some had only taken a basic course in statistical inference, while others had taken a broad spectrum of M.S.- and Ph.D.-level statistics courses. On the other hand, all of them had already taken required core courses in investment theory and derivative pricing, and STATS 240 was supposed to link the theory and pricing formulas to real-world data and pricing or investment strategies. Besides students in the program, the course also attracted many students from other departments in the university, further increasing the heterogeneity of students, as many of them had a strong background in mathematical and statistical modeling from the mathematical, physical, and engineering sciences but no previous experience in finance. To address the diversity in background but common strong interest in the subject and in a potential career as a “quant” in the financial industry, the course material was carefully chosen not only to present basic statistical methods of importance to quantitative finance but also to summarize domain knowledge in finance and show how it can be combined with statistical modeling in financial analysis and decision making.

The course material evolved over the years, especially after the second author helped as the head TA during the years 2004 and 2005. The course also expanded to include a section offered by the Stanford Center for Professional Development, with nondegree students in the financial industry taking it on-line (http://scpd.stanford.edu). The steady increase in both student interest and course material led to splitting the single course into two in 2006, with STATS 240 followed by STATS 241 (Statistical Modeling in Financial Markets). Part I of this book, Basic Statistical Methods and Financial Appli-
cations, is covered in STATS 240 and has six chapters. Chapters 1 and 2 cover linear regression, multivariate analysis, and maximum likelihood. These statistical methods are applied in Chapter 3 to a fundamental topic in quantitative finance, namely portfolio theory and investment models, for which Harry Markowitz and William Sharpe were awarded Nobel Prizes in Economics. Whereas the theory assumes the model parameters are known, in practice the parameters have to be estimated from historical data, and Chapter 3 addresses the statistical issues and describes various statistical approaches. One approach is deferred to Section 4.4 in Chapter 4, where we introduce Bayesian methods after further discussion of likelihood inference for parametric models and its applications to logistic regression and other generalized linear models that extend the linear regression models of Chapter 1 via certain “link functions.” Chapter 4 also extends the least squares method in Chapter 1 to nonlinear regression models. This provides the background for the nonlinear least squares approach, which is used in various places in Part II of the book. Another important topic in quantitative finance that has attracted considerable attention in recent years, especially after the 2003 Nobel Prizes to Robert Engle and Clive Granger, is financial time series. After introducing the basic ideas and models in time series analysis in Chapter 5, Chapter 6 extends them to develop dynamic models of asset returns and their volatilities. The six chapters that constitute Part I of the book provide students in financial mathematics (or engineering) and mathematical (or computational) finance programs with basic training in statistics in a single course that also covers two fundamental topics in quantitative finance to illustrate the relevance and applications of statistical methods.

Part II of the book, Advanced Topics in Quantitative Finance, is covered in STATS 241 at Stanford. It introduces nonparametric regression in Chapter 7, which also applies the methodology to develop a substantive-empirical approach that combines domain knowledge (economic theory and market practice) with statistical modeling (via nonparametric regression). This approach provides a systematic and versatile tool to link the theory and formulas students have learned in mathematical finance courses to market data. A case in point is option pricing theory, the importance of which in financial economics led to Nobel Prizes for Robert Merton and Myron Scholes in 1997, and which is a basic topic taught in introductory mathematical finance courses. Discrepancies between the theoretical and observed option prices are revealed in certain patterns of “implied volatilities,” and their statistical properties are studied in Chapter 8. Section 8.3 describes several approaches in the literature to address these discrepancies and considers in particular a substantive-empirical approach having a substantive component associated with the classical Black-Scholes formula and an empirical component that uses nonparametric regression to model market deviations from the Black-Scholes formula. Chapter 9 introduces advanced multivariate and time series methods in financial econo-
metrics. It provides several important tools for analyzing time series data on interest rates with different maturities in Chapter 10, which also relates the statistical (empirical) analysis of real-world interest rates to stochastic process models for the valuation of interest rate derivatives in mathematical finance. The finance theories in Chapters 8 and 10 and in mathematical finance courses assume the absence of arbitrage. “Statistical arbitrage,” which has become an important activity of many hedge fund managers, uses statistical learning from market prices and trading patterns to identify arbitrage opportunities in financial markets. Chapter 11 considers statistical trading strategies and related topics such as market microstructure, data-snooping checks, transaction costs, and dynamic trading. Chapter 12 applies the statistical methods in previous chapters and also describes new ones for risk management from the corporate/regulatory perspective, protecting the financial institution and its investors in case rare adverse events occur.

Since M.S. students in financial mathematics/engineering programs have strong mathematical backgrounds, the mathematical level of the book is targeted toward this audience. On the other hand, the book does not assume many prerequisites in statistics and finance. It attempts to develop the key methods and concepts, and their interrelationships, in a self-contained manner. It is also intended for analysts in the financial industry, as mentioned above in connection with the Stanford Center for Professional Development, and exposes them to modern advances in statistics by relating these advances directly to financial modeling.

Because Part I is intended for a one-semester (or one-quarter) course, it is presented in a focused and concise manner to help students study and review the material for exercises, projects, and examinations. The instructor can provide more detailed explanations of the major ideas and motivational examples in lectures, while teaching assistants can help those students who lack the background by giving tutorials and discussion sessions. Our experience has been that this system works well with the book. Another method that we have used to help students master this relatively large amount of material in a single course is to assign team projects so that they can learn together and from each other.

Besides students in the M.S. program in financial mathematics at Stanford, the course STATS 241, which uses Part II of the book, has also attracted Ph.D. students in economics, engineering, mathematics, statistics, and the Graduate School of Business. It attempts to prepare students for quantitative financial research in industry and academia. Because of the breadth and depth of the topics covered in Part II of the book, the course contents actually vary from year to year, focusing on certain topics in one year and other topics in the next. Again, team projects have proved useful for students to learn together and from each other.
After the second author joined the faculty at Columbia University in 2005, he developed a new course, W4290, in the M.S. program in statistics on the applications of statistical methods in quantitative finance. Since students in this course have already taken, or are concurrently taking, courses in statistical inference, regression, multivariate analysis, nonparametrics, and time series, the course focuses on finance applications and is based on Chapters 3, 6, and 8 and parts of Chapters 10, 11, and 12. For each finance topic covered, a review of the statistical methods to be used is first summarized, and the material from the relevant chapters of the book has been particularly useful for such a summary. Besides the core group of students in statistics, the course has also attracted many other students who are interested in finance not only to take it but also to take related statistics courses.

The Website for the book is:

The datasets for the exercises, and instructions and sample outputs of the statistical software in \texttt{R} and \texttt{MATLAB} that are mentioned in the text, can be downloaded from the Website. The Web page will be updated periodically, and corrections and supplements will be posted at the updates. We will very much appreciate feedback and suggestions from readers, who can email them to xing@stanford.edu.

\textbf{Acknowledgments}

First we want to thank all the students who took STATS 240 and STATS 241 at Stanford and W4290 at Columbia, for their interest in the subject and comments on the earlier draft of the book. We also want to express our gratitude to the teaching assistants of these courses for helping us improve the course material and develop exercises and projects. In particular, we want to thank Antje Berndt, Zehao Chen, Wei Jin, Yuxue Jin, Andreas Eckner, Noureddine El Karoui, Horel Guillaume, Jingyang Li, Wenzhi Li, Haiyan Liu, Ruixue Liu, Paul Pong, Bo Shen, Kevin Sun, Ning Sun, Zhen Wei, Chun-Yip Yau, and Qingfeng Zhang. The first author was invited to give a lecture series by the Institute of Statistical Science at Academia Sinica in December 2001. This was the first time he drafted a set of lecture notes on statistical methods in finance. He wants to thank the attendees of the lectures for their enthusiastic interest. He is particularly grateful to his coauthor on the substantive-empirical approach, Samuel Po-Shing Wong, who helped him prepare the lecture notes and delivered a lecture describing the approach. Two years later, he teamed up with Wong again to give another lecture series on statistical methods in finance at the Center of Mathematical Sciences of Zhejiang University, and he wants to thank the participants for their interest and feedback, which helped him refine the lecture notes. Comments and suggestions from friends and former students working in the financial industry have been especially helpful in
shaping the contents of Part II. We are particularly grateful to Ying Chen, David Li, Viktor Spivakovsky, and Zhifeng Zhang in this connection.

Department of Statistics, Stanford University
Tze Leung Lai
Department of Statistics, Columbia University
Haipeng Xing
December 2007