Preface

The main purpose of this volume is to investigate algorithmic methods based on machine learning in order to design sequential investment strategies for financial markets. Such sequential investment strategies use information collected from the past of the market and determine, at the beginning of a trading period, a portfolio, that is, a way to invest the currently available capital among the assets that are available for purchase or investment.

Our aim in writing this volume is to produce a self-contained text intended for a wide audience, including graduate students in finance, statistics, mathematics, computer science, and engineering, as well as researchers in these fields. Thus the material is presented in a manner that requires only a basic knowledge of probability.

In the approach that we adopt, the goal of the decider or investor is to maximize his wealth on the *long run*; however the investor does not have direct information about the underlying distributions which are generating the stock prices. In the area of mathematical finance most of the known theoretical results have been obtained for models which consider single assets in a single period, and they typically assume a parametric model of the underlying stochastic process of the prices.

In the last decade it has become clear that decision schemes that consider multiple assets simultaneously, and that try to consider decisions over multiple periods can increase the investor's wealth through judicious rebalancing of investments between the assets. Since accurate statistical modelling of stock market behavior is now known to be notoriously difficult, in our work we take an extreme point of view and work with minimal assumptions on the probabilistic distributions regarding the time series of interest. Our approach addresses the best rebalancing of portfolio assets in the sense that it maximizes the expected log-return. If the distributions of the underlying price processes are unknown then one has to "learn" the optimal portfolio from past data, and effective empirical strategies can then be derived using methods from nonparametric statistical smoothing and machine learning.

The growth optimal portfolio (GOP) is defined as having a maximal expected growth rate over any time horizon. As a consequence, this portfolio is sure to outperform any other significantly different strategy as the time horizon increases. This property in particular has fascinated many researchers in finance and mathematics, and created an exciting literature on growth optimal investment.

Thus Chapter 1 attempts to provide a comprehensive survey of the literature and applications of the GOP. In particular, the heated debate of whether the GOP has a special place among portfolios in the asset allocation decision is reviewed as this still seem to be an area where some misconceptions exist. The survey also provides a review of the recent use of the GOP as a pricing tool, for instance in the "benchmark approach".

Chapter 2 provides a survey of sequential investment strategies for financial markets. The GOP can be derived from the log-optimal criterion (called Kelly-criterion, too), which means that one chooses the portfolio maximizing the conditional expectation of the log-return given the past data. Under the memoryless assumption on the underlying market process of the assets' relative prices the best constantly rebalanced portfolio is studied, called log-optimal portfolio, which achieves the maximal asymptotic average growth rate. Semi-log optimal portfolio selection as a small computational complexity alternative of the log-optimal portfolio selection is studied both theoretically and empirically. For generalized dynamic portfolio selection, when the market process is stationary and ergodic, the challenging problem is whether or not it is possible to learn the conditional distributions from data, i.e., whether one can construct empirical (data driven) strategies achieving the optimal growth rate. It turns out that utilizing the current approaches of nonparametric estimates and machine learning algorithms such empirical GOPs exist. The empirical performance of the methods is illustrated for NYSE data.

The theoretical and empirical optimality of GOP will be based on some assumptions, the most important of which is that the transaction cost is ignored. In Chapter 3 the discrete time growth optimal investment with proportional transaction costs is considered. Here the market process is modelled by a first order Markov process. Assuming that the distribution of the market process is known, we show sequential investment strategies such that, in the long run, the growth rate on trajectories achieves the maximum with probability 1.

In the previous chapters the model does not include the possibility of short selling and leverage. Chapter 4, revisits the GOP on non-leveraged, long only markets for memoryless market process. We derive optimality conditions to frameworks on leverage and short selling, and establish no-ruin conditions. Moreover we investigate the strategy and its asymptotic growth rate from both theoretical and empirical points of view. The empirical performance of the methods is illustrated for NYSE data showing that with short selling and leverage the growth rate is drastically increasing.

For constructing empirical GOPs, the role of nonparametric estimates and machine learning algorithms is important. Chapter 5 is devoted to the application of these principles for the prediction of stationary time series. This chapter presents simple procedures for the prediction of a real valued time series with side information. For the regression problem, survey the basic principles of nonparametric estimates. Based on current machine learning algorithms, the predictions are the aggregations of several simple predictors. We show that if the sequence is a realization of a stationary random process then the average of squared errors converges, almost surely, to that of the optimum, given by the Bayes predictor. We offer an analog result for the prediction of gaussian processes. These prediction strategies have some consequences for the pattern recognition problem, too.

Chapter 6 is on empirical pricing American options, which can be viewed as an optimal stopping problem derived from a backward recursion such that in each step of the recursion one needs conditional expectations. For empirical pricing, Longstaff and Schwartz suggested to replace the conditional expectations by regression function estimates. We survey the current literature and based on nonparametric regression estimates, some new algorithms are introduced and investigated.

As we conclude this preface, we would like to acknowledge the contribution of many people who influenced this volume. A number of colleagues and friends have, often without realizing it, contributed to our understanding of rebalancing, nonparametrics and machine learning. In particular we would like to thank in this respect Paul Algoet, Andrew Barron, Tom Cover, Miguel Delgado, Luc Devroye, Jürgen Dippon, László Gerencsér, András György, Péter Kevei, Jussi Klemelä, Michael Kohler, Adam Krzyżak, Kasper Larsen, Tamás Linder, Gábor Lugosi, Gábor Molnár-Sáska, Eckhard Platen, Miklós Rásonyi, Christian Riis Flor, Walter Schachermayer, Dominik Schäfer, Wolfgang Stummer, Csaba Szepesvári, Frederic Udina, István Vajda and Sara van de Geer.

In addition, we gratefully acknowledge the research support of the Budapest University of Technology and Economics, the Hungarian Academy of Sciences (MTA SZTAKI), the Morgan Stanley, the MSCI Inc., the Universität Stuttgart and the University of Southern Denmark. This work is connected to the scientific program of the "Development of quality-oriented and cooperative R+D+I strategy and functional model at BME" project, which is supported by the New Hungary Development Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002).

László Györfi, Budapest, Hungary
György Ottucsák, Budapest, Hungary
Harro Walk, Stuttgart, Germany

January 26, 2011

Contents