Cairo University
Faculty of Economics and Political Science
Department of Statistics

ON FINANCIAL TIME SERIES DATA MINING

By
Mona Nazih Ali Abdel Bary

A Dissertation Submitted to the Faculty of Economics and Political Science,
Cairo University, in Partial Fulfillment of the Requirements
for the degree of Doctor of Philosophy in Statistics

Supervised by

Prof. Dr. Nadia M. Girgis

Professor of Statistics,
Faculty of Economics and Political Science
Cairo University

Prof. Dr. Ali S. Hadi

Distinguished University Professor,
The American University, Egypt, and
Professor Emeritus, Cornell University, USA.

Dr. Azza A. El Naggar

Assistant Professor of Statistics
Faculty of Economics and Political Science
Cairo University

Acknowledgements

Any attempt at any level cannot be satisfactorily completed without the support and guidance of learned people. I wish to express my sincere gratitude to Prof. Dr. Nadia M. Girgis, for her continuous advice, support and encouragement during this work.

I would like to express my deep gratitude to my teacher, Prof. Dr. Ali S. Hadi for his great patience, detailed instructions and insightful comments at every stage of this Dissertation What I have learned from him goes beyond the knowledge he taught me; his attitude toward research, carefulness and commitment has and will have a profound impact on my future academic activities.

I cannot thank enough my supervisor, Dr. Azza A. El Naggar for her continuous advice, support and encouragement during this work.

I am also grateful to external examiner; my teacher, Prof. Dr. Ibrahim H. Ibrahim and examination member Dr. Kamal S. Selem.

Furthermore I am thankful to Prof. Dr. Mohammed Ismail and Dr. Zenab H. Amin for their interest and support.

Finally I would like to thank my father for all what he has done for me. Words cannot express one millionth of my gratitude to him.

To my father, my sisters, and the memory of my mother.

On Financial Time Series Data Mining

Abstract

The portfolio selection problem has a venerable history. Markowitz (1952), one of the creators of the modern portfolio theory, formulates the problem as a trade-off between the expected return and the expected risk of a portfolio. For his path breaking work that has revolutionized investment practice, he won the Nobel Prize in 1990. In this Dissertation we propose two enhancements to the traditional portfolio selection problem. First we enhance the formulation of the problem by introducing four additional constraints that take into account the following: (a) the collinearity problem to decrease the portfolio risk, (b) the special preference to active stocks to increase the expected return and decrease the systematic risk, (c) the special preference to stocks with outstanding performance to increase the un-expected return, and (d) control the overall risk of the portfolio.

Second, one of the common algorithms for solving the portfolio selection optimization problem is the Genetic Algorithm (GA), which is a stochastic search that starts with an initial solution and then allocates increasing trials to regions of the search space found to improve the objective function. This algorithm can run into problems when the optimal solution is in a small region surrounded on all directions by regions of low value of the objective function. We propose an enhancement to the GA that avoids this problem.

Time series techniques such as ARIMA and GARCH models can be used to predict the return and risk for each stock. This predictions can be used as input to the model.

The forecast of financial time series is a fundamental problem due to its importance in risk management of the stocks. There are three major difficulties about accurate forecast of financial time series, (a) the patterns of financial time series are dynamic, i.e., there is no single model that works all the time, (b) an efficient model must be able to adjust its sensitivity as time goes by, (c) misleading information must be identified and eliminated. A Hidden Markov Model (HMM) aims to solve these problems. We propose trading rules using HMM to answer the question of how and when do investors trade the stock?

Keywords: Modern portfolio theory, Optimal Portfolio, Stock Selection, Genetic Algorithm, Quadratic Programming, Nonlinear Programming, Financial Time Series, Hidden Markov Model, Portfolio Rotating and Stock Exchanging .

A Dissertation Submitted to the Department of Statistics

2011

Supervised by

Prof. Dr. Nadia M. Girgis, Prof. Dr. Ali S. Hadi, and Dr. Azza A. El Naggar

Name:	Mona Nazih Ali Abdel Bary
Nationality:	Egyptian
Date and Place of Birth:	29-11-1975, Tanta, Egypt
Degree:	Ph.D.
Specialty:	Statistics
Supervisors:	Prof. Dr. Nadia M. Girgis
	Prof. Dr. Ali S. Hadi
	Dr. Azza A. El Naggar

On Financial Time Series Data Mining

Summary

In this Dissertation, we propose enhancements to the traditional portfolio selection problem and to the Genetic Algorithm for obtaining the optimal solution. We also compare various methods for obtaining the optimal portfolio. Additionally, we suggest trading rules for rotating the portfolio and exchange of stocks.

First, we enhance the formulation of the problem by introducing four additional constraints that take into account (a) the collinearity problem to decrease the portfolio risk, (b) the special preference to active stocks to increase the expected return, (c) the special preference to stocks with outstanding performance to increase the un-expected return, and (d) control the overall risk of the portfolio.

Second, one of the common method for solving the portfolio selection optimization problem is Quadratic Programming. But by adding the nonlinear constraint we cannot use it for optimizing the portfolio. But, we can use both Nonlinear Programming (NLP) and Genetic Algorithm (GA). GA is a stochastic search that starts with an initial solution and then allocates increasing trials to regions of the search space

found to improve the objective function. This algorithm can run into problems when the optimal solution is in a small region surrounded on all sides by regions of low value of the objective function. We propose an enhancement to the GA that avoids this problem. We use the simulation study to empirically assess the performance of the proposed optimal portfolio model according to a number of constraints, and to compare their performance to that of the classical optimal portfolio model. In addition, an application of the two enhancements to data from the Egyptian stock market, which is known to be inefficient, shows that the two enhancements together can lead to substantial improvement in the selection of optimal portfolio. In fact, it is possible to find a portfolio with a performance better than that of the stock market as a whole, an event which is impossible to occur in efficient markets.

Finally, the forecast of financial time series is a fundamental problem due to it's importance in risk management of the stocks and the portfolios. There are three major difficulties in accurate forecasting of financial time series: (a) the patterns of financial time series are dynamic, i.e., there is no single model that works all the time, (b) an efficient system must be able to adjust its sensitivity as time goes by, (c) misleading information must be identified and eliminated. A Hidden Markov Model (HMM) aims to solve these problems. We propose a trading rules using HMM to answer the question of how and when do investors trade the stock?

This dissertation is organized as follows: Chapter 1 provides a literature review of optimization models and methods, definition for the Egyptian Stock Market, data Mining and financial market, and the assumptions and objectives of the study.

Chapter 2 includes the main basics of Hidden Markov model, the difference between Markov model and Hidden Markov model, the main concepts in Hidden Markov model, types of Hidden Markov models, and steps involved in creating the structure a Hidden Markov model.

Chapter 3 introduces the main problem of the Modern Portfolio Theory and the dissertation's suggested solution. Additionally, it suggests method for enhancing the genetic algorithm. Two proposals for enhancing global Optimal portfolios are introduced. Finally, it refers to a frame for risk management of the stocks.

Chapter 4 includes the simulation to assess the performance of the proposed enhancement of the optimal portfolio frame. This includes the design of the simulation study, generation of the simulation data, and results of the simulation.

Chapter 5 introduces the application of the new portfolio models and methods to the Egyptian Stock Market.

Chapter 6 introduces the application of the risk management frame for determine the adequate time for buying and selling each stock.

Chapter 7 provides summary and further future work.

Contents

List of Tables