DYNAMIC ASSET ALLOCATION
BY STOCHASTIC PROGRAMMING METHODS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
MANAGEMENT SCIENCE AND ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Alexis Collomb
December 2004

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Walter Murray Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Gerd Infanger

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

James Primbs

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Hervé Kieffel

Approved for the University Committee on Graduate Studies.

Abstract

This thesis addresses two research questions: (i) the predictability of equity returns and (ii) the variations of optimal asset allocations, within a multi-stage stochastic programming framework, with respect to the statistical modeling of returns.

The predictability and serial correlation of equity returns has been for many years a controversial subject among academics and professional financial analysts. We propose here a novel methodology for detecting local bursts of serial correlation. Moreover, by analyzing historical data and conditional returns, we show evidence of market predictability for equity indices over short-term horizons of a few days, with both momentum and reversal effects.

Stochastic programming is widely used for many asset and liability management ("ALM") applications. Within ALM models, we look at the limited setting of a multi-stage asset allocation problem and restrict ourselves to a multi-stage stochastic programming framework. As for any solution technique used for solving a problem with uncertain parameters, the stochastic modeling of the uncertainty may drive the character of the solution. We compare the optimal asset allocations obtained from a geometric Brownian motion ("GBM") model and a vector autoregressive ("VAR") model of asset class returns. In the process, we show clear evidence of serial correlation for the returns on Treasury bonds and bills and compare the forecasting performances of the GBM and the VAR models. In the VAR case, we show that the allocation results vary significantly depending on the initial conditions. For both the GBM and VAR models, we also show that results may be very sensitive to the historical samples used for calibrating the models. To address this instability, we use a third statistical model of asset class returns, a Bayesian vector autoregressive ("BVAR") approach. We show it provides better out-of-sample forecasts of returns, especially when the different statistical models are calibrated using small samples. We conclude that our multi-stage stochastic programming model (for which it is trivial to include transaction costs) is particularly appropriate for assisting allocation decisions with limited information and significant transaction costs, as is the case for funds of funds.

Acknowledgements

First of all, I want to thank my Reading Committee. In particular, I am greatly indebted to Professor Walter Murray who made me feel welcome and advised me throughout the years. Day after day, Professor Murray provided me with very valuable research guidance and colorful conversational topics (from optimization to Monty Python, and all kinds of sports and politics in between). His support made my time here both intellectually productive and very pleasant... Just as inspirational and enthusiastic was Professor Gerd Infanger. I owe him a lot with respect to stochastic programming insights and I very much enjoyed his contagious optimism throughout the process. I also want to thank Professor James Primbs for his very useful and pointed comments and his great responsiveness. I am also greatly indebted to Professor Hervé Kieffel who provided me with profusely detailed and disciplined feedback.

For the great ride my years in the Systems Optimization Laboratory have been, I would like to thank (in chronological order): Byunggyoo Kim, Kien-Ming Ng, Michael Friedlander, Maureen Doyle, Che-Lin Su, Uday Shanbag, Samantha Infeld and Kaustuv. I am also very grateful to Michael Saunders for his incredibly pointed and efficient corrections, out of pure benevolence, and to Professor Richard Cottle for his dry humor, on the fly. In addition, I was honored to be hooded by Professor Peter Glynn (representing Professor Murray). I should also mention I was extremely inspired by Professor Darrell Duffie for his quasi-encyclopedic knowledge of finance (and other topics), his class on dynamic asset pricing theory, and his always sharp and warm responses to my emails. Also, I want to thank Professor Ken Singleton for allowing me to audit his equally excellent class on empirical asset pricing and sit in his research and reading group. I am not mentioning for the sake of conciseness many other people who all taught me a lesson or a proof, at some point in time, but I hope they will forgive me. I wish I had consulted with all above mentioned more often.

I would like to thank my family (direct and extended) and numerous friends (too many to name here) for their continued moral support. Three people I would have liked to thank

directly for the role models they were in getting me to this point (in very different, yet equally meaningful, ways) are my grandfathers, Pierre-Charles Wirth and Charles Collomb, and my godfather, Pierre Faurre. This dissertation is dedicated in part to their memories.

Last but not least, this thesis is dedicated to my parents. I am most thankful for their support and enduring love over the years. I am greatly indebted to my father, for being such a father, and to my mother, for being such a mother! I know they will read this thesis, will be very critical, and will encourage me to do much more!

I hope as I finish writing these lines that the fantastic time I had at Stanford will turn out to have been not an end in itself, but a beginning and a gate to the end of education which (I was told) is creation. I hope this work qualifies as the first building block, however humble, of a hard-working career of trying to make useful contributions.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, the quantitative management of financial cash flows for either individuals or organizations has become increasingly important. The last score of years have seen a significant development of equity markets around the globe (fuelled to a large extent by the US markets) and the size of the US Pension Asset Reserves has dramatically increased from about \$2 trillion in 1985 to more than \$9 trillion in 2000 as Figure 1.1 shows. This rise in the size of pension assets under management can partly be explained by the significant development of individual retirement accounts ("IRAs") and other types of retirement plans such as 401(k) (this is illustrated by Figure 1.2). Today more than ever, the quantitative management of assets and liabilities is of primary importance for either individuals or institutions.

For individuals, the increasing access in many countries (especially the US) to online brokerage, financial news channels and the like, has significantly contributed to developing a taste for financial planning. The many sophisticated tools available to financial planners have become increasingly available to the general public and this increased availability has contributed in turn to building the general public awareness of the subject. In the US alone, the proportion of wealth invested in equity products and other types of financial securities has greatly increased since 1985 (as shown by Figure 1.3). This trend has been supported in part by the general public perception that these products are the most appropriate means for funding (or supplementing) retirement.

For organizations, the recent development and use of sophisticated information technologies (e.g. corporate intranets) has usually translated into improved financial accounting and reporting. This greater and more accurate volume of financial information available has in turn emphasized to many corporate or institutional executives the need for better financial

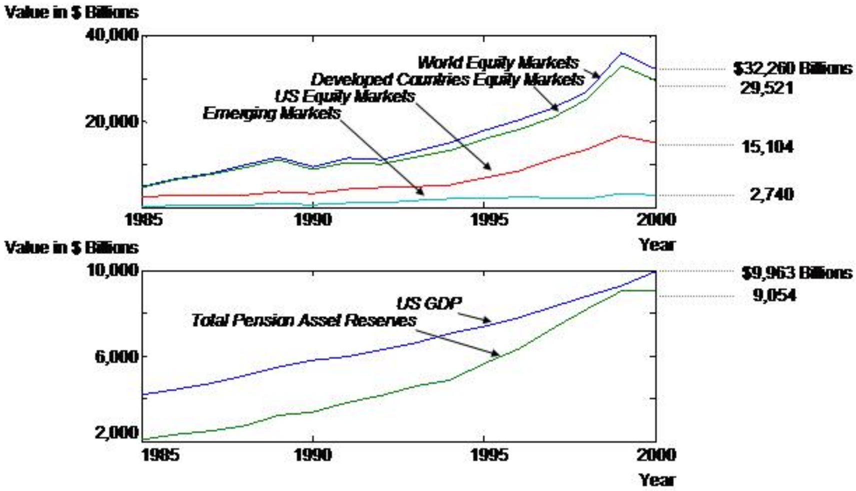


Figure 1.1: Growth of US Pension Asset Reserves. Source: 2001 Securities Industry Fact Book.

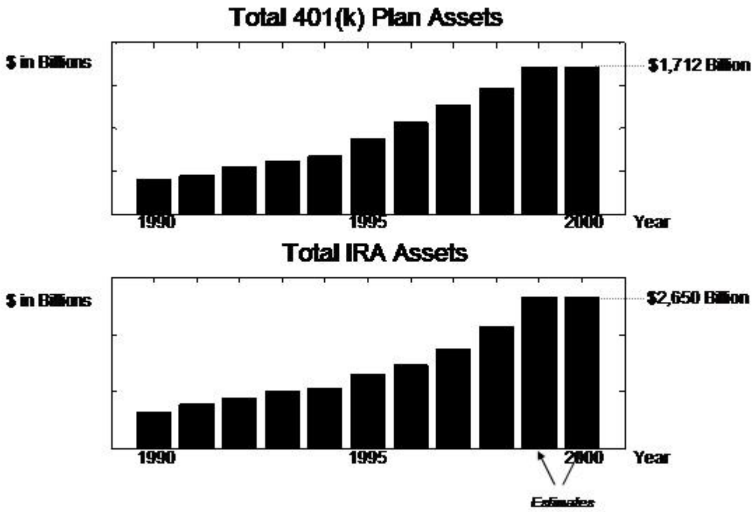


Figure 1.2: Growth of 401(k)'s and IRAs. Source: 2001 Securities Industry Fact Book.

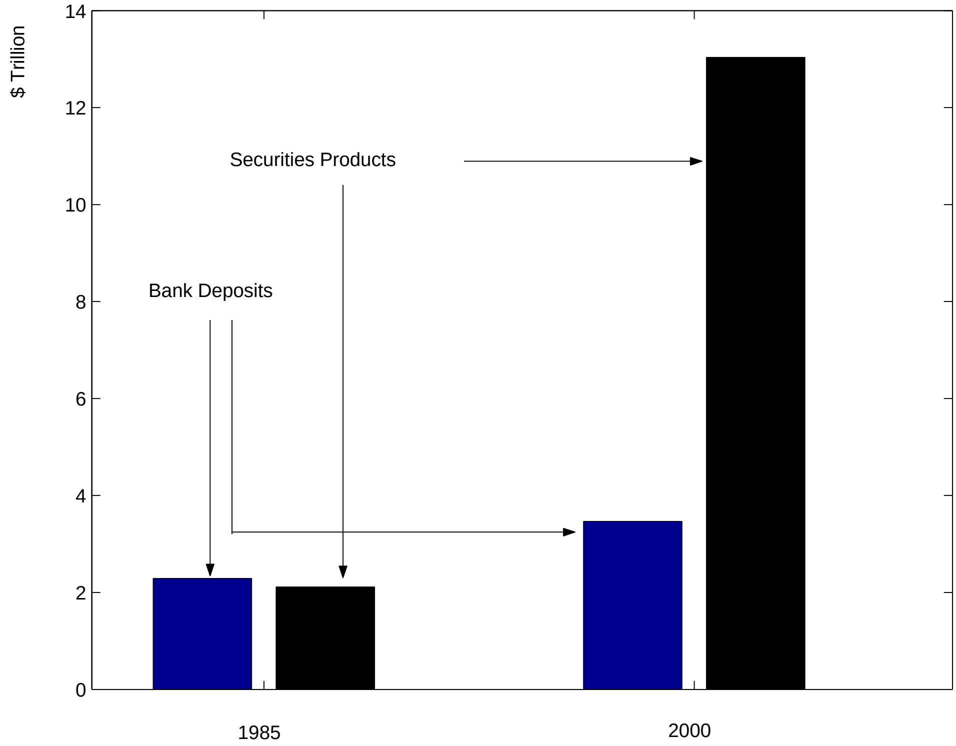


Figure 1.3: This figure shows the progression from 1985 to 2000 of the dollars invested by US households in liquid financial assets. Over that period, the bank deposits grew from about \$2.3 to \$3.5 trillion (a compound annual growth rate of about 2.85%) while the investments in securities products rose much more rapidly from \$2.1 to \$13.1 trillion (a compound annual growth rate of almost 13%). Source: 2001 Securities Industry Fact Book.

management. In this case, the paradigm seems to be that accounting and financial information flowing seamlessly to decision makers has helped them identify inefficiencies and improve the firm's overall financial management.

In general, the use of financial planning tools has been somehow demystified. Hence it is reasonable to assert that, today more than ever, financial planning and the management of cash flows is of concern to a large public.

In this chapter, we first provide a general review of Asset and Liability Management ("ALM") models and identify from first principles some of the general issues that they involve. Second, we introduce the particular framework of our research and map out our specific contributions.

1.1 Representation of an ALM model

ALM models usually require drawing from different fields such as statistics, economics, financial mathematics and optimization. In a schematic and generic way, we can organize the different elements of ALM models along three separate poles of interest and a common set of structural considerations.

Building an ALM model almost always requires performing three tasks before starting the optimization process. The first task consists in a prior elicitation of future asset and liability cash flows that are to be expected. The second task is to clearly define the optimization objectives and make sure they are appropriate for the investing entity (whether an individual or an organization). The third task is to choose the set of possible investment vehicles that we want to include in the optimization process. Common to these three tasks are structural considerations that need to be accounted for such as whether to have a continuous- or discrete-time model, how new information is revealed (i.e. the so-called information filtration) and other general assumptions. Figure 1.4 schematizes these points.

1.1.1 Asset and Liability Cash Flows

It is important to understand here the different uses that we can make of the word "asset". One usage of "asset", as in "asset cash flow", refers to the underlying asset that is generating revenues for the firm or the individual of concern. In other words, if we are dealing with an individual's financial planning model, the asset cash flow referred to above would be the income stream this individual expects to receive during his working career. Another usage of the word "asset", as in "asset class", refers to a specific class of investments. In developing an ALM model, the nature of the asset and liability streams needs to be carefully estimated.

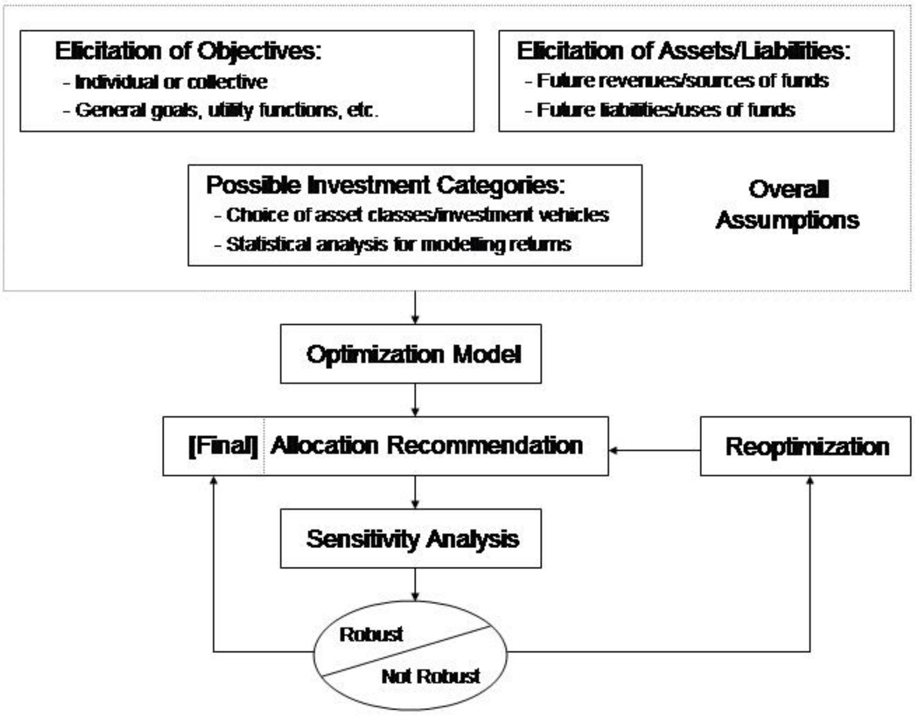


Figure 1.4: Optimization Framework for an ALM Model. This graph, which we suggest, provides an overview of the three important types of assumptions that optimizing an ALM model requires defining: (i) the future asset and liability cash flows; (ii) the objectives and (iii) the possible asset classes or investment vehicles.

The precise timing of cash flows may be known in advance or not. In many cases, we may only have an approximate idea of their temporal distribution. Similarly, the magnitude of these cash flows may be known in advance precisely or in a fuzzy way. In many (if not most) cases, we may only have a probabilistic assessment of what to expect. All these early assumptions need to be carefully assessed as they may be extremely influential on the optimal investment policy recommendation we are seeking to calculate.

1.1.2 Optimization Objectives

The second task at hand is to properly elicit the optimization objectives. The representation of these objectives will vary a lot depending on whether we are dealing with an individual investor or an organization.

Individual Decision Making

The modeling of optimization objectives for an individual decision maker is far more complex than it seems at first glance. As put by Bell, Raiffa, and Tversky (Bell, Raiffa, and Tversky 1988), we can distinguish between three kinds of theories of decision making under uncertainty. Normative theories state how decision makers or economic agents should behave (if they were fully "rational"). Descriptive theories tell us how agents do behave in fact (and usually underline "deviations" from rationality). Last but not least, prospective theories attempt to offer advice as to how decision makers should behave when faced with incomplete information and/or complex situations that cannot be fully grasped by our limited cognitive abilities.

Normative School The normative school of thought focuses on what rational decision making should be and assumes that an individual's preferences over a set of alternatives can pass logical tests of consistency derived from a few axioms. In other words the decision maker is said to be "rational". For instance, this approach assumes completeness over the different set of alternatives, meaning that it presupposes that the decision maker is always able to rank two alternatives (or state his indifference between the two). Assuming they are consistent, the decision maker's preferences are then translated into a utility function supposed to reflect the possible satisfaction the decision maker can derive from wealth (Friedman and Savage 1948, Friedman 1957, Markowitz 1952b). For an individual (who is usually risk averse), representing the investor's objectives would then translate into fitting an increasing and strictly concave utility function by asking the individual to choose between a series of monetary deals, and this until a satisfactory bounding of the risk aversion parameter(s) of interest has been achieved.

Descriptive and Prospective Schools The normative school has been seriously questioned by different investigations in behavioral economics, behavioral finance and psychology that demonstrated the inconsistency and irrationality of many economic agents. This is a vast topic and we limit ourselves here to providing some references in the different fields of behavioral finance (Shefrin and Statman 1985), economics (Arrow 1986, Kahneman and Tversky 1979, Thaler 1980, Thaler 1981, Tversky and Kahneman 1991), decision sciences (Bell, Raiffa, and Tversky 1988), marketing (Thaler 1985) and psychology (Kahneman and Tversky 1973, Kahneman and Tversky 1984).

Organizational Decision Making

By "organizational decision making", we refer to any decision making situation that involves a complex organization (e.g. a corporation, an academic institution, etc.) In such a case, eliciting and quantifying clear objectives can be a significant (and sometimes open-ended) task. For instance in the case of a large corporation, defining the ALM optimal policy may be a complex iterative process between different stakeholders such as the executive management, the auditors, some large shareholders, etc. Tradeoffs between short-term and long-term objectives, between conflicting interests and different opinions and assessments of financial and operational risks need to be agreed upon. Thus, our end purpose of clearly eliciting quantified objectives that can later be "plugged" into an ALM model may never be met. There has been a vast and still growing literature on designing organizational mechanisms to make sure that different groups in a complex organization have incentives to cooperate. Designing such incentive-compatible mechanisms requires a game theoretic mindset and Thaler suggests using "prescriptive" game theoretic approaches (Thaler 1992). We also refer the interested reader to the review by Kreps on "mechanism design" (Kreps 1990). Also, Collomb provides a thorough treatment on how to formulate "reasonable" objectives when facing the possibility of divergent criteria between various management groups (Collomb 1971).

1.1.3 Set of Possible Investments

There are many possible investment vehicles with different risk profiles that can be used for investing purposes. We distinguish here between "primary" investment vehicles such as individual stocks or bonds and "secondary" investment vehicles, such as mutual funds.

There are even higher-order investment vehicles, such as fund of funds[1] (that we could call "tertiary" investment vehicles according to our classification).

Among the traditional primary vehicles of investment are stocks, bonds (whether issued by a corporation or a government), currencies, commodities, etc. According to our classification, secondary investment vehicles comprise all funds that invest in the former, such as mutual funds, exchange-traded funds ("ETFs"), index funds, so-called "hedge" funds, etc. As an example of one of these funds, a mutual fund is simply a financial intermediary that allows a group of investors to pool their money together with a predetermined investment objective. The mutual fund will have a fund manager who is responsible for investing the pooled money into specific securities (usually stocks or bonds) and investing in a mutual fund will consist in buying shares (or portions) of the mutual fund and becoming a shareholder of the fund. What is important to emphasize here is the increasingly cascading nature of the possible investment vehicles that can be used for investing activities, from individual securities to high-level funds of funds.

From the investor's perspective, all these investment opportunities have very different risk characteristics, performance features, accessibility and fees, among others. Of fundamental importance is the representation that an investor has of the dynamics of these investment vehicles. This usually involves significant statistical work for analyzing their past performance in an effort to model their uncertain future performance (to the extent that past performance bears upon future performance).

In the bulk of our work we limit our investing universe to certain asset classes and their representative indices (considered "secondary" as they represent a pool of individual securities). In particular we use broad-based equity indices, government bonds and short-term treasury bills (used as a proxy for cash). In the last chapter, we address the issue of funds of funds in the context of a particular application.

1.1.4 Structural Assumptions

There are general assumptions that need to be made across most financial models in general, and ALM models in particular. First, an important issue is whether a continuous- or a discrete-time model will be used and, if discrete, how fine a time mesh should be chosen. Another important consideration is whether financial agents can be considered as "price takers" in a Walrasian context or not. Other general considerations of interest include: the information filtration (i.e. the way new information will be delivered), observability conditions of asset prices and/or of new data (e.g. signal/noise ratios), term structure

[1] A "fund of funds" is simply a fund that invests in funds instead of individual securities.

dynamics, inflationary expectations and other macroeconomic factors.

Continuous and Discrete Times

There is a rich financial literature of both continuous- and discrete-time models. For instance, Duffie provides a dense review of both types of models (Duffie 2001) and Merton provides an overview and synthesis of finance theory from the perspective of continuous-time analysis (Merton 1990). Usually it can be shown that for many purposes, such as contingent claim prices and optimal consumption-portfolio policies, the results derived from the discrete-time models converge to their corresponding continuous-time limits (He 1989). However, the fact that financial data used by practicioners is only available on a discrete basis raises serious practical implications for using continuous-time models. For instance Brigo and Mercurio have shown the clear limit of approximating discrete time with continuous time for pricing options (Brigo and Mercurio 2000). And choosing which continuous model is the most appropriate given discretely sampled data is far from being an obvious question as the work by Aït-Sahalia and Mykland shows (Aït-Sahalia and Mykland 2004). In addition, often solving continuous models requires strong assumptions that are unwarranted by empirical evidence if closed-form solutions are to be obtained. If these assumptions are not made and numerical solutions are required, discrete methods have to be used.

Walrasian Context

In most financial articles, the asset allocation optimization is performed under the assumption that the optimizing agent (whether an individual or an organization) is a "price-taker" (i.e. the agent's market orders and control actions are too small to influence the markets). Hence the investment decisions taken are separate from and have no bearing on the dynamics of the investment vehicles and the stochastic processes used to represent them. In the case of an individual investor, this assumption is (in most cases) largely justified. However, in the case of a large fund or of an institution executing sizable trades, this Walrasian assumption may be erroneous and inappropriate as feedback mechanisms may exist (it is difficult to execute large trades in a security without influencing the market in that security).

1.2 Our Research Framework and Contributions

1.2.1 Research Focus

Our original research intent was, in a broad sense, to investigate the predictability (if any) of financial markets and analyze the impact that any such predictability of financial returns would have on ALM models. We soon realized this original goal was far too ambitious for our purposes. First, the issue of whether, and to what extent, some asset class returns are predictable or not has been a vastly studied subject over the years. While it remained a controversial topic in the case of equity returns, finding anything significantly new there (assuming there was anything to be found) appeared challenging. Second, ALM models can vary a great deal depending upon the needs and applications they originate from and hence defining a generic model for our analysis seemed unlikely to be really fruitful.

Our research eventually settled on two different specific questions that stemmed from this original interest of studying market predictability and its impact on ALM models. First, we focused on analyzing short-term market momentum and reversals by introducing a new (and deceivingly simple) methodology which, to the best of our knowledge, had never been carried out before on equity indices. Second, to assess the predictability and serial correlation of different asset class returns, we estimated -against the same large sample of data- different statistical models of returns. In particular, and in contrast to the standard log-normal model of returns derived from geometric Brownian motion ("GBM"), we used a vector autoregressive ("VAR") and a Bayesian vector autoregressive ("BVAR") framework for analyzing asset class returns. We then assessed the forecasting performance of these different statistical models of returns and compared the allocation results these different statistical models of returns would yield.

1.2.2 Our Contributions

First, we revisited the issue of predictability of equity returns and have provided a new framework for detecting it. Our empirical work shows strong evidence of serial correlation in daily returns of equity indices. It underscores statistically significant momentum and reversal effects for equity indices, conditional upon their past cumulative number of consecutive upward or downward movements.

Second, we have analyzed the impact of different statistical models of asset class returns within a multi-stage stochastic programming framework. For a given set of data comprised of stock, bond and cash returns (for both yearly and monthly data), we have assessed the different forecasting performance of various stochastic models, including vector autoregressive ("VAR") processes and Bayesian vector autoregressive processes ("BVAR"). We have clearly shown that a VAR framework can be useful for predicting different bond returns (and the yield curve in general). We have then shown that for the same sample of data and the same programming horizon, the differences in the optimal asset allocations calculated can be significantly different depending upon which model of asset class returns is used.

Third, our BVAR multi-stage stochastic programming framework provides an appropriate model for funds of funds investing. In particular our model addresses two significant issues that funds of funds have to face: (i) significant transaction costs and (ii) scarce or limited historical performance data that makes it hard to quantify precisely the allocation process.

Chapter 2

Asset Allocation Methods and Strategies

In discrete time, there is a variety of techniques used by professionals and academics to solve dynamic asset allocation models but essentially we can categorize them as belonging either to a dynamic programming methodology or a stochastic programming methodology (or both).

Dynamic programming was first introduced by Bellman (Bellman 1957) and provides, among other things, a solution methodology for solving dynamic problems with a graph structure. Stochastic dynamic programming is essentially the same idea except that the dynamic structure of the problem may not be spelled out explicitly (as with an exhaustive graph structure describing the universe of possible scenarios) but rather be represented by continuous distributions.

Stochastic programming is defined (Birge and Louveaux 1997) as the branch of mathematical programming (whether linear or nonlinear) where some of the problem data might be stochastic (as opposed to fixed parameters).

There exists also other types of formulation of ALM models such as the network formulation for investment planning as developed by Mulvey and Vladimirou (Mulvey and Vladimirou 1989).

Early use of a dynamic programming formulation for solving investment problems can be found in Mossin (Mossin 1968) and Samuelson (Samuelson 1969).

Similarly, stochastic programming formulations of investment planning can be found in Dantzig and Infanger (Dantzig and Infanger 1993) and Infanger (Infanger 1993).

Our purpose in this chapter is to review the main existing asset allocation techniques and emphasize the pros and cons of the stochastic programming approach to investment planning.

In order to fulfill this goal, we first need to review the different utility functions traditionally used. They are crucial as they can sometimes drive the optimal investment policy.

2.1 Synopsis of Utility Functions

The concept of a utility function has axiomatic foundations that date back to von Neumann and Morgenstern (von Neumann and Morgenstern 1944) and Savage (Savage 1954). Utility functions were originally designed as a tool for choosing between alternatives that would produce different random wealth variables. The two main principles in their construction are non-satiation (i.e. more wealth should be preferred to less) and risk aversion (at least for most individuals) that respectively translate into having an increasing and concave mapping of wealth to the utility it yields.

2.1.1 Risk Aversion and Risk Tolerance

The degree of risk aversion exhibited by a utility function is related to the curvature of the function. Traditionally, it is formally defined by the Arrow-Pratt absolute risk aversion ("ARA") coefficient (cf (Luenberger 1998)) which is:

$$ARA(x) = -\frac{U''(x)}{U'(x)}. \tag{2.1}$$

The relative risk aversion ("RRA") coefficient is:

$$RRA(x) = -W\frac{U''(x)}{U'(x)}. \tag{2.2}$$

Risk tolerance is defined as the reciprocal of risk aversion. Thus, the absolute and relative risk tolerance coefficients are:

$$ART(x) = \frac{1}{ARA(x)}, \tag{2.3}$$

$$RRT(x) = \frac{1}{RRA(x)}. \tag{2.4}$$

2.1.2 Hyperbolic Absolute Risk Aversion Utility Functions

Usually absolute risk aversion decreases as wealth increases. A general class of utility functions that meets this criterium is the set of hyperbolic absolute risk aversion ("HARA") utility functions. A general formulation for such functions is

$$U(\mathrm{C}) = \frac{1-\gamma}{\gamma}\left(\frac{\beta \mathrm{C}}{1-\gamma} + \eta\right)^{\gamma}, \tag{2.5}$$

with appropriate values for the parameters and C. This family is rich as by varying the parameters, a utility function with absolute or relative risk aversion increasing, decreasing, or constant, can be obtained (Merton 1971). Particular cases of utility functions often used include:

- Linear or risk-neutral function: $U(x) = x$
- Quadratic: $U(x) = x - \frac{1}{2}cx^2$
- Exponential: $U(x) = -e^{-ax}$
- Power: $U(x) = cx^{\gamma}$
- Logarithmic: $U(x) = \ln x$

It is important to appreciate that in a one-period setting where the utility function is quadratic or the asset returns are all normal, maximizing the expected value of the utility function is equivalent to solving the Markowitz mean-variance problem (Luenberger 1998).

2.2 Synopsis of Asset Allocation Strategies

We review here the different standard allocation strategies by increasing degree of complexity.

Buy-and-Hold Strategy This is the simplest one-period strategy where an initial asset allocation is chosen and held through until the end of the planning horizon. A simple allocation choice can be derived by assuming normal returns and calculating the mean-variance portfolio.

Fixed-Mix Strategy This is used for a multi-period setting. The asset allocation weights are fixed and the assets rebalanced to those initial weights at each decision point. An attractive feature of this allocation policy is that it is somehow equivalent to a certain form of "volatility pumping" as at every period assets are sold "high" (i.e. when they have done better than most of their peers) and bought "low". The theoretical properties of fixed-mix strategies are discussed, among others, by Merton and Dempster, Evstigneev and Schenk-Hoppé (Merton 1990, Dempster, Evstigneev, and Schenk-Hoppé 2003). Infanger makes the point that fixed-mix strategies perform well even when some of the assumptions required for their theoretical justification are relaxed (Infanger 2002).

Fixed-mix strategies are important because they are often used to benchmark portfolio managers. Different studies have been made to assess the contribution of active asset selection and management and the value it added on top of the benchmarks. Such studies can be found in the work of Brinson, Hood and Beebower (Brinson, Hood, and Beebower 1986, Brinson, Hood, and Beebower 1991), Hensel, Ezra and Ilkiow (Hensel, Ezra, and Ilkiow 1991) and Blake, Lehmann and Timmermann (Blake, Lehmann, and Timmermann 1999).

Because fees are related to performance, usually relative to a benchmark or a peer group, Blake et al. in their study of the U.K. market showed that:

- U.K. pension fund managers have a weak incentive to add value and face many constraints if and when they try to do it.
- Fund managers know that relative, rather than absolute, performance determines their long-term survival in the industry.
- Fund managers earn fees related to the value of assets under management, not to their relative performance against a benchmark or their peers, with no specific penalty for underperforming or reward for outperforming.[1]

All these studies show the importance of fixed-mix strategies since they are often used as benchmarks in the fund management industry.

But as Samuelson (Samuelson 1969) and Merton (Merton 1990) show, there is also a theoretical justification for fixed-mix strategies if:

- Asset return distributions are independent and identically distributed.
- The utility function exhibits constant relative risk aversion.

[1] This is significantly different from the case of hedge fund managers whose compensation usually includes a performance fee of about 20% of the upside.

- Only investment income is considered.
- No transaction costs are allowed.

If the utility function is logarithmic, it is worth noticing that non-iid asset returns also result in a constant allocation strategy.

Stochastic Programming The two previous strategies have the drawback that they do not use new information from return realizations in their determination. The study by Cariño and Turner (Cariño and Turner 1998) shows that the recourse flexibility provided by a multiperiod stochastic programming model improves performance results. Because a multi-stage stochastic program allows the asset mix at any future stage to adjust to the current wealth, the objective is improved. Ziemba (Ziemba 2003) shows results to that effect.

However, Fleten, Høyland and Wallace (Fleten, Høyland, and Wallace 2002) show more mitigated results due to the differences between an in-sample and out-of-sample analysis for the portfolio model of a Norwegian life insurance company. In their analysis, the authors show that the stochastic programming approach works significantly better than a fixed-mix strategy in an in-sample analysis but only slightly better in an out-of-sample analysis. If, as the analyst proceeds forward through the time series of asset returns, there is a strong disconnect between the realized returns and the forecasted ones, the stochastic programming approach may loose its precision and its advantage over a fixed-mix approach and the gap between the two strategies closes. As Ziemba puts it, *"the stochastic programming model loses its advantage in optimally adapting to the information available in the scenario tree"*. However, we would still expect the stochastic programming approach to yield some improvement over a pure fixed-mix strategy.

Stochastic Dynamic Programming Stochastic dynamic programming is another solution technique for solving a multi-stage scenario tree. Whereas, multi-stage stochastic programming requires solving directly (or indirectly through sampling and decomposition techniques) a large-scale linear (or nonlinear) program, stochastic dynamic programming makes it possible to solve a multi-stage scenario tree by backward induction (provided the state space is of manageable size). Early use of dynamic programming for solving investment problems go back at least to Mossin (Mossin 1968) and Samuelson (Samuelson 1969).

Whether to use dynamic programming rather than stochastic programming is not of vital interest to us at this stage. For small problems, it is not particularly relevant, as both methods should give similar results. The more interesting question is how we forecast future asset returns in both cases. If the forecasts are out of line with the actual returns, an out-of-sample analysis of the results is likely to show that a multi-stage model (whether solved by dynamic or stochastic programming) is of little use versus a fixed-mix or even a buy-and-hold strategy.

However, especially in the presence of transaction costs, a multi-stage stochastic framework offers a natural and flexible formulation, especially in the case where the investment planner has strong views about particular dynamic scenarios.[2]

2.3 Idiosyncrasies of Stochastic Programming

The stochastic programming approach provides, in our view, a flexible and more realistic approach to dynamic asset allocation models. Among the important reasons for this flexibility is the fact that the modeling of either the utility function or the asset class returns do not require the use of smooth functions or highly regular distributions. In other words, a numerical approach to the multi-stage stochastic program, usually done by pre-sampling from the distributions, yields satisfactory results. Some of the idiosyncracies of the stochastic programming approach, as pointed out by Ziemba (Ziemba 2003) are:

- The often artificial elicitation of a smooth utility function is not needed. For individual planning, all that is needed is to understand future liabilities and how important the decision maker considers them. For doing so, the traditional mean-variance objective can be abandoned for a more tailored objective that the decision maker can easily understand. In our numerical results, we use a utility function that is piecewise linear and strictly concave. Each linear segment of the utility function has its own slope, supposed to reflect the relative penalty cost the decision maker associates with ending the planning horizon within a certain wealth range, long or short of a desired objective.

- Scenarios supposed to reflect particular beliefs about various elements of the models (e.g. a particular shock to returns in the event of a major geopolitical setback, etc.) can be easily included or adjusted. This form of scenario-dependent knowledge is also usually easy to extract from the decision maker. Adjusting certain scenarios may be vital in some cases. For instance, in standard asset models, the asset returns correlation matrix is often assumed to be constant throughout the model. This may be a

[2]It is worth noting that, in the presence of transaction costs, the problem as formulated becomes path-dependent and the current wealth cannot be used as the only state variable in the dynamic programming formulation. The state space needs to be enlarged in a significant way to reflect transaction costs, and the dynamic programming approach looses some of its attractiveness.

dangerous simplification. If most of the time, asset returns can be rightly assumed to follow a certain correlation structure, in times of particular stress on the financial markets (such as what happened in August 1998 with the Russian government defaulting on its debt), asset returns tend to be much more correlated and for the worse. Representing this alteration of the correlation structure is relatively easy to implement in the case of a scenario-based multi-stage stochastic program.

- It is easy to include transaction costs and different regularity requirements such as cash reserve constraints and turnover limits to a stochastic program.

The first two points listed above are also valid for a dynamic programming methodology. The first point pertains to the formulation of the utility function, which is independent of the solution technique. The second point relates to the formulation of the multi-stage scenario-based tree, which would be the same for dynamic or stochastic programming. However the third point is idiosyncratic to a stochastic programming formulation.

However, the stochastic programming approach has its limitations. In particular, the investment problem represented as a multi-stage program usually chooses the timing of these stages arbitrarily. A stochastic control approach would perform better. The stochastic programming approach also places two kinds of constraints on the optimal allocation policy. First it places an upper bound on the number of reallocation points. If there are significant transaction costs, it is very well possible that the solution of the "free problem" would only require a few trades and hence this constraint might not be too drastic. However, the second constraint is the time distribution of these stages. Assuming there are (n-1) reallocation stages in addition to the original allocation, we have to be careful in choosing their temporal distribution. If this distribution reflects "natural" points of reallocation (e.g. the decision maker only looks at his/her portfolio every year at the beginning of the year), then presetting these reallocation points seems appropriate. However, if there is no a priori behavioral constraints relevant to the decision maker, we need to check that the distribution of reallocation times is not too arbitrary.

Chapter 3

Comparison of Asset Allocation Methods

In this chapter, we perform both a single-period and a multi-period analysis of different asset allocation strategies. For one-period models, we focus on a mean-variance efficient type of approach, as developed by Markowitz (Markowitz 1952a). For a multi-period model, we compare the different results obtained with buy-and-hold, fixed-mix and stochastic programming strategies.

3.1 Single-Period Model

A single-period model, despite its simplicity, is enough to capture a few important empirical points in asset allocation strategies. In general, the three following assumptions strongly influence the dynamic asset allocation results:

- The size of the historical window used for estimating the expected returns and covariance structure of the asset classes considered if indeed historical estimates are used.[1]

- The risk-free and borrowing rates.

- The investor's utility function.

[1]Though we do not address it here, it is important to bear in mind that other methodologies are more appropriate for estimating volatility such as using implied volatilities from equity option prices.

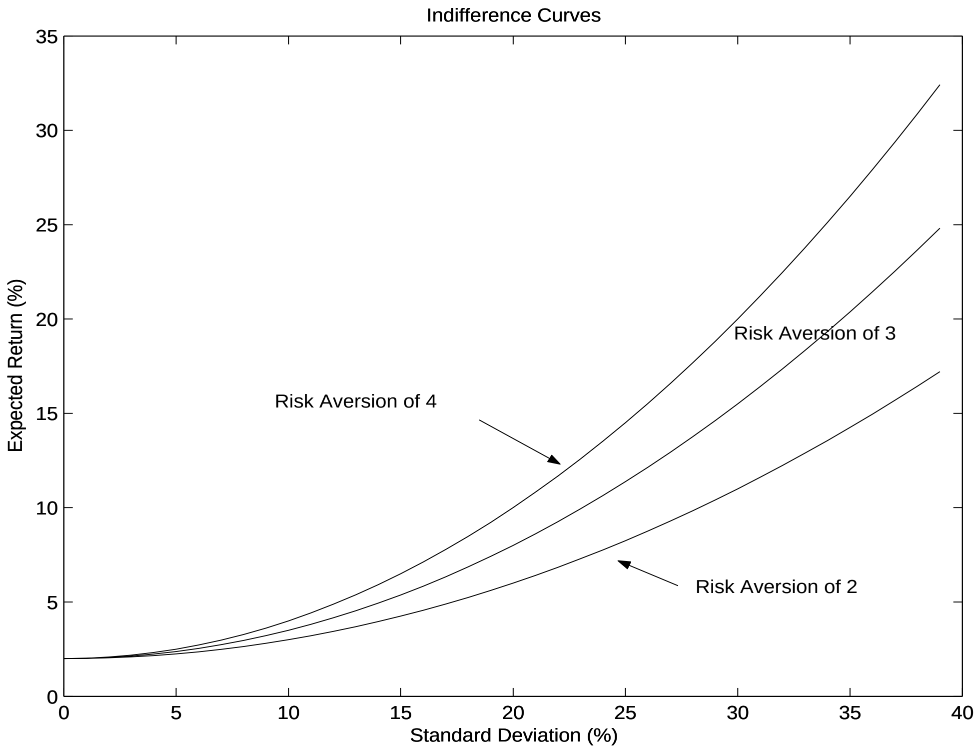


Figure 3.1: Utility Function Indifference Curves.

3.1.1 Sensitivity Results on Historical Estimates

The first analysis we do is on yearly data from 12/31/1946 to 12/31/2002 using as asset classes stocks and bonds. The equity class is represented in a first analysis by the NYSE index and in a second analysis by the NASDAQ index. For the government bond asset class, we use the 30-yr bond yearly returns.

Utility Function and Risk Aversion

For convenience, we use Matlab Financial Toolbox's integrated utility function:

$$U = E(r) - 0.5\, A\, \sigma^2 \tag{3.1}$$

where U is the utility function, $E(r)$ is the expected return, σ the standard deviation and A the risk aversion which will be the varying parameter in our analysis.

Figure 3.1 shows the indifference curves for such a utility function.

Results with NYSE Index

The risk aversion in the example is set at 7 for the utility function previously defined. The size of the historical window is 10 to 50 years. The risk free rate is at 2% and the borrowing rate is at 3%. These risk free and borrowing rate values are reasonable given that the historical window used ends in 2002. We proceed as if we were calculating at the beginning of 2003 a one-year forward allocation until the end of 2003 based upon historical data. At this point in time (i.e. January 2003), the interest rates were very low and our assumed values are reasonable.

Figure 3.2 shows the variations of the overall risk (defined as standard deviation) and the overall return to be expected with respect to the size of the historical window used for estimating both the stock and 30-yr bond expected returns, as well as their covariance matrix. The window size varies from 10 to 50 years. This means that we have computed the estimates using the last 10 to 50 years of historical returns, from the ending date of our sample.

We can observe that it takes about thirty years of historical data for both the expected return and risk to stabilize. However there is a inherent tension between the number of years needed to get seemingly stable results and the validity of using returns belonging to a distant past.

The next graphs represented in Figure 3.3 show the variations of the fraction of wealth to be allocated to the risky portfolio, along with the recommended fractions of total wealth allocated respectively to stocks and 30-yr bonds.

Based on NASDAQ Historical Performance

We perform the same analysis but this time with NASDAQ value-weighted index returns (including dividends) obtained from the CRSP. These NASDAQ returns are calculated from 1973 to 2003.

We can see from Figure 3.4 and Figure 3.5 that the variations are still quite significant. It is worth noticing that the inclusion of the first two years in our sample (1973 and 1974) reduces significantly the expected returns and the allocation to the risky portfolio.

3.2 Multi-Period Model

For multi-period models, we show the advantages of using stochastic programming over more traditional strategies, such as "buy-and-hold" and "fixed-mix" strategies.

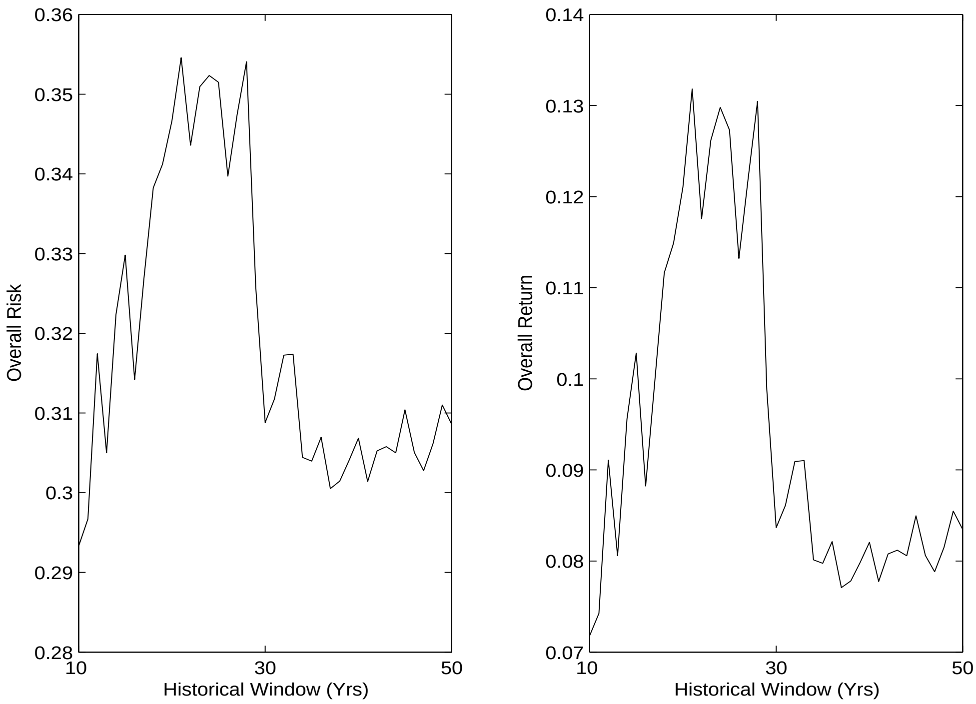


Figure 3.2: Variations of Overall Risk and Return (NYSE case). The NYSE yearly returns used are value-weighted and include dividends.

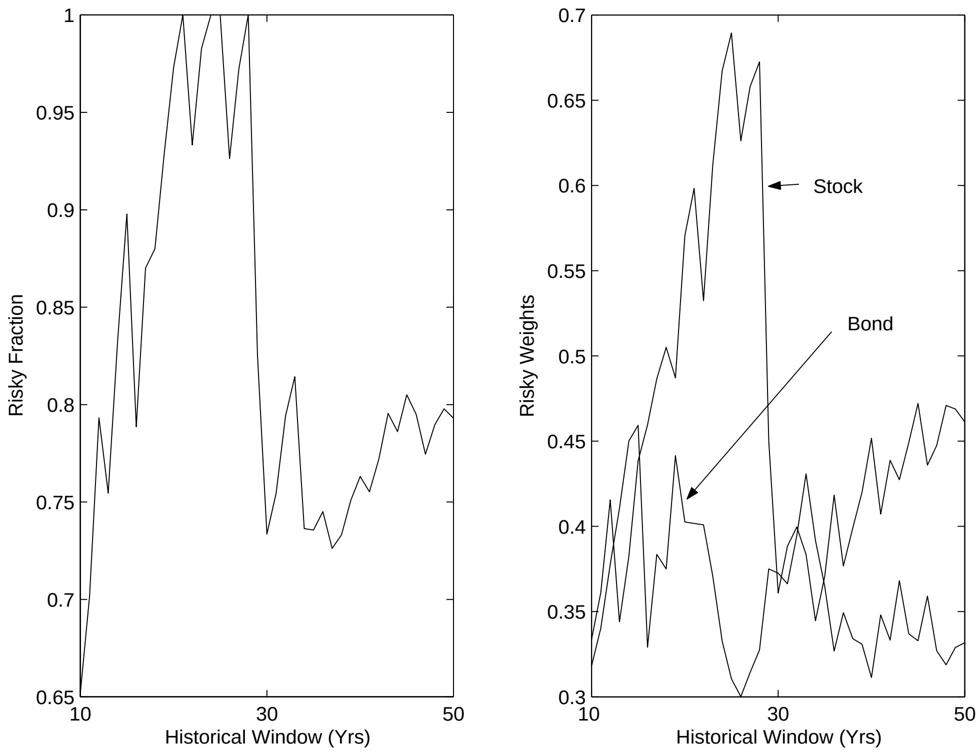


Figure 3.3: Allocation Weights (NYSE case).

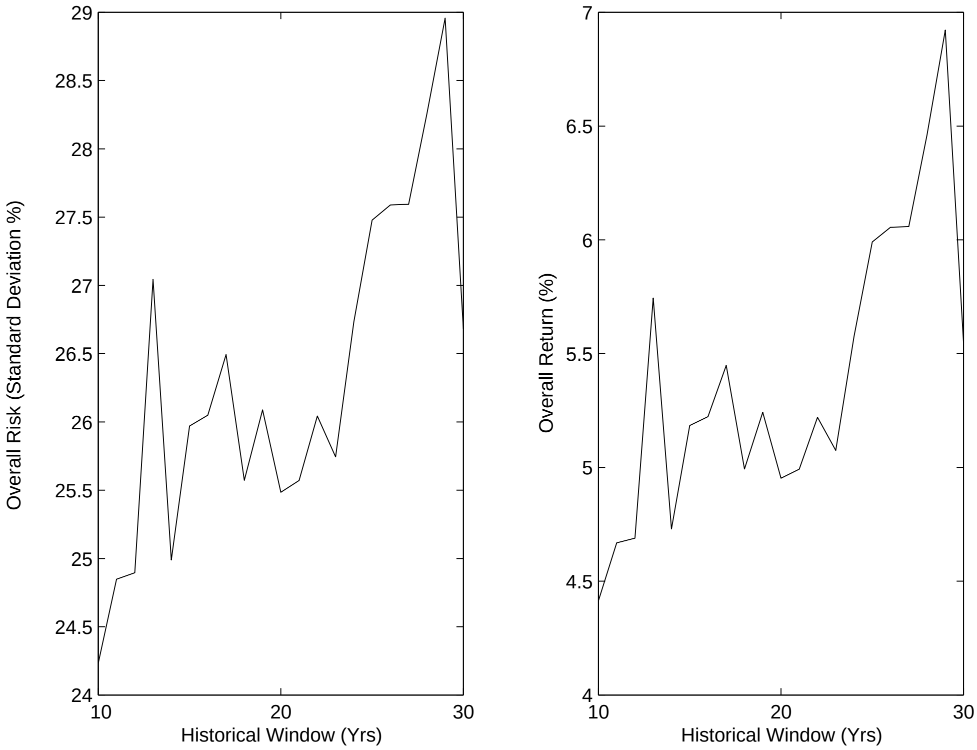


Figure 3.4: Variations of Overall Risk and Return (NASDAQ case). The NASDAQ yearly returns used are value-weighted and include dividends.

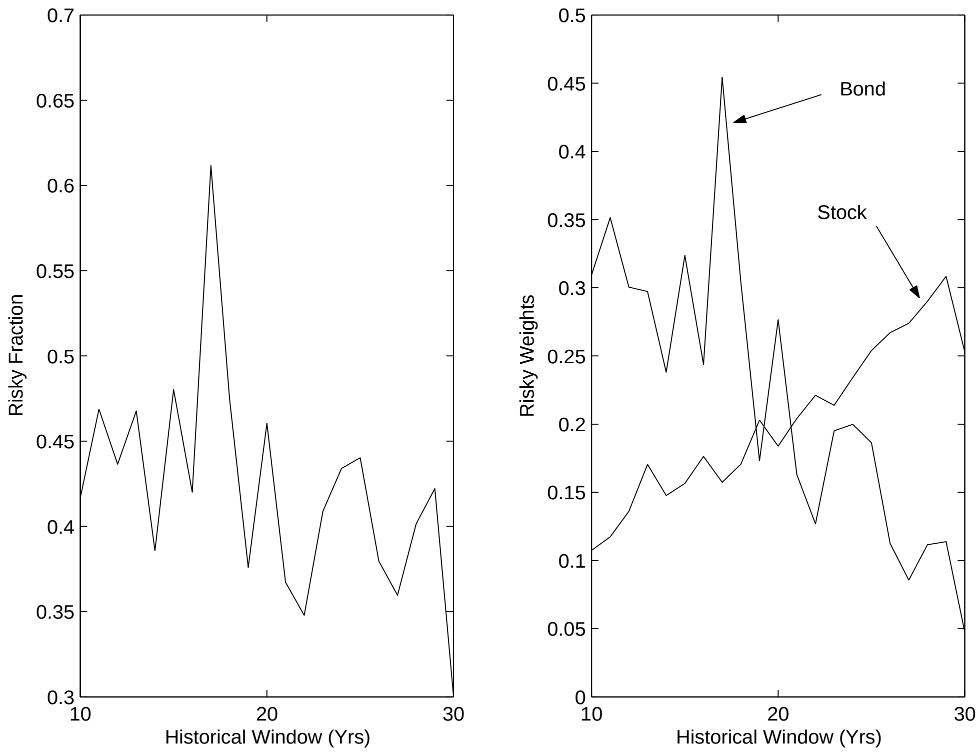


Figure 3.5: Allocation Weights (NASDAQ case).

3.2.1 Buy-and-Hold Strategy

This strategy is the simplest. As its name indicates, there is no rebalancing of the portfolio weights once the initial allocation decision has been made.

The initial equation is:

$$\sum_{i=1}^{n-1} x_0^i\,(1+btc^i) + x_0^{cash} = W_0, \tag{3.2}$$

where W_0 is the initial wealth available[2], $n-1$ is the number of asset classes other than money markets[3] and btc^i are the transaction costs for investing in asset class i.[4] Between $t-1$ and t the following equation applies:

$$x_{t-1}^{i,\omega_1,\ldots\omega_{t-1}}\,R_t^{i,\omega_t} = x_t^{i,\omega_1,\ldots\omega_t}, \quad i=1,...,n-1 \quad and \quad t=1,...,T-1. \tag{3.3}$$

The random variable R_t^{i,ω_t} is the random return of asset class i between times $t-1$ and t drawn in scenario ω_t. For the last period, we get:

$$\sum_{i=1}^{n-1} x_{T-1}^{i,\omega_1,\ldots,\omega_{T-1}}\,R_T^{i,\omega_T}\,(1-stc^i) + x_{T-1}^{cash,\omega_1,\ldots,w_{T-1}}\,R_T^{cash,\omega_T} = W_T^{\omega_1,\ldots,\omega_T}, \tag{3.4}$$

where stc^i are the transaction costs for selling out of the asset class i and $W_T^{\omega_1,\ldots\omega_T}$ is the terminal wealth thereby obtained in the scenario T-tuple $(\omega_1,...,\omega_T)$. The utility function is represented by slack variables $u_T^{\omega_1,\ldots,\omega_T}$ for being above the final goal G and $v_T^{w_1,\ldots,\omega_T}$ for being short of the goal, so that:

$$u_T^{\omega_1,\ldots,\omega_T} - v_T^{\omega_1,\ldots,\omega_T} = W_T^{\omega_1,\ldots,\omega_T} - G. \tag{3.5}$$

G is the desired goal. The final objective function, Z, is:

$$\mathrm{Z} = \frac{\sum_{(\omega_1,\ldots,\omega_T)\,\in\,\Omega_1\times\ldots\times\Omega_1}[u_T^{\omega_1,\ldots,\omega_T} - \psi\, v_T^{\omega_1,\ldots,\omega_T}]}{[|\Omega_1|\times\ldots\times|\Omega_T|]}. \tag{3.6}$$

[2]For the sake of clarity, "W", in W_0, is not to be confused -and has nothing to do- with "ω", in ω_t, which is a different letter and designates a time t scenario. We keep these possibly confusing notations however to follow existing literature.

[3]We assume there are no transaction costs for transferring money in and out of money markets and treat money markets as equivalent to "cash", which is our n^{th} asset class.

[4]In our model, the transaction costs are assumed to be proportional to the amount invested in the asset class.

In equation 3.6, ψ is the slope of the utility function below the goal G. It is a penalty factor for being short of this desired final goal.

3.2.2 Fixed-Mix Strategy

The fixed-mix strategy forces the additional constraints that at any reallocation time t, we rebalance the portfolio to fall back (post transactions costs) upon the initial allocation weights we had chosen at time $t = 0$. To deal with transaction costs, we have to introduce additional variables that represent how much of each asset class we have bought, identified by y_i, or sold, identified by z_i. We detail the full model, starting with the initial allocation equation of the initial wealth W_0 into the different asset classes:

$$\sum_{i=1}^{n-1} x_0^i \,(1 + btc^i) + x_0^{cash} = W_0,$$

and we calculate from it the fixed-mix proportions x_{FM}^i:

$$x_{FM}^i = \frac{x_0^i}{\sum_{i=1}^n x_0^i}, \quad i = 1...n. \tag{3.7}$$

Subsequently, at intermediate times, we have the following set of equations:

$$x_{t-1}^{i,\omega_1,\ldots,\omega_{t-1}}\, R_t^{i,\omega_t} + y_t^{i,\omega_1,\ldots,\omega_t} - z_t^{i,\omega_1,\ldots,\omega_t} = x_t^{i,\omega_1,\ldots,\omega_t}, \quad i = 1...n-1,$$

$$x_{cash}^{i,\omega_1,\ldots,\omega_{t-1}}\, R_t^{i,\omega_t} - \sum_{i=1}^{n-1} y_t^{i,\omega_1,\ldots,\omega_t}(1 + btc^i) + \sum_{i=1}^{n-1} z_t^{i,\omega_1,\ldots,\omega_t}(1 - stc^i) = x_t^{cash,,\omega_1,\ldots,\omega_t},$$

where $y_t^{i,\omega_1,\ldots,\omega_t}$ is the amount of asset class i bought in scenario $(\omega_1, ..., \omega_t)$ and z_t^i is the amount sold in scenario t-tuple $(\omega_1, ..., \omega_t)$. In addition, we have the fixed-mix constraints:

$$\frac{x_t^{i,\omega_1,\ldots,\omega_t}}{\sum_{i=1}^n x_t^{i,\omega_1,\ldots,\omega_t}} = x_{FM}^i, \quad i = 1...n. \tag{3.8}$$

The terminal equations are the same as previously:

$$\sum_{i=1}^{n-1} x_{T-1}^{i,\omega_1,...\omega_{T-1}} R_T^{i,\omega_T} (1 - stc^i) + x_{T-1}^{cash,\omega_1,...w_{T-1}} R_T^{cash,\omega_T} = W_T^{\omega_1,...\omega_T},$$

$$u_T^{\omega_1,...,\omega_T} - v_T^{\omega_1,...,\omega_T} = W_T^{\omega_1,...,\omega_T} - G,$$

$$\mathrm{Z} = \frac{\sum_{(\omega_1,...,\omega_T)\,\in\,\Omega_1\times...\times\Omega_T}[u_T^{\omega_1,...,\omega_T} - \psi\, v_T^{\omega_1,...,\omega_T}]}{[|\Omega_1| \times ... \times |\Omega_T|]}.$$

It is worth noticing that the introduction of the fixed-mix constraint imposes nonlinear constraints. Hence this nonlinear program ("NLP") is solved using MINOS (Murtagh and Saunders 1983). For further references on optimization, we refer the reader to the work by Gill, Murray and Wright (Gill, Murray, and Wright 1986).

3.2.3 Stochastic Programming Strategy

The stochastic program strategy is formulated in the same way as the fixed-mix strategy but without the fixed-mix constraints, i.e. equations (3.7) and (3.8). Because the feasible domain of the stochastic program encompasses the feasible domain of the fixed-mixed strategy, we know immediately that the stochastic programming strategy should yield a higher (or equal) objective value.

3.3 Computational Results

Figure 3.6 shows our multi-period framework for comparing different asset allocation strategies. We use a terminal utility function at T=3 and two reallocation times (when permitted by the strategy) at t=1 and t=2. Units of time are in years. The quantity of interest is the optimal asset allocation at t=0. The penalty factor, ψ, for being short of the goal, G, is set at 10. The number of scenarios is $|\Omega_1| = 50$, $|\Omega_2| = 30$, and $|\Omega_3| = 20$, for a total number of scenarios of 30,000. The initial wealth is set at a standard of $W_0 = 100$ and the final goal at $G = 115$.

Table 3.1 shows results for an in-sample analysis. The stochastic programming strategy performs better than the two others, as expected. What is more surprising is the very close results between the buy-and-hold strategy and the fixed-mix strategy. The solution time of the latter (about 8,000 seconds) is much larger than the solution time of the former (i.e. 176 seconds). It is not so surprising however given the fact that the fixed-mix strategy requires solving an NLP. We suspect that these relatively comparable results are due to a

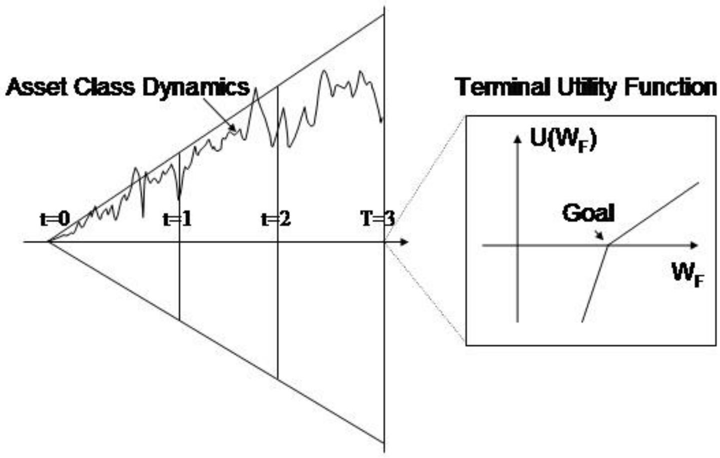


Figure 3.6: Framework for Computational Results.

Allocation Strategy	Stock (%)	Bond (%)	Cash (%)	Obj.	Sol. Time (sec.)
Buy-and-Hold	55.9%	19.9%	24.2%	-3.79	176.1
Fixed-Mix	55.0%	19.3%	25.7%	-3.79	7983.3
Stochastic Programming	67.4%	27.7%	4.9%	4.62	243.9

Table 3.1: Comparison of asset allocation strategies. Computations were carried out on a Intel Pentium 4 running at 1.6 GHz with 512 Mb of RAM.

large part to the small sample size used. Also, in this set of results, all transaction costs are set at zero. However, adding transaction costs will make the buy-and-hold strategy more attractive.

Chapter 4

Momentum of Equity Returns

In this chapter, we present a new methodology for analyzing the predictability of equity returns. Instead of trying to determine the best statistical fit for daily equity returns, we study directly the serial properties of daily returns in a restricted framework. Specifically we analyze how series of consecutive rising (or, conversely, diminishing) returns unfold. Our original work shows evidence that even in recent years, equity markets still harbored both momentum and mean-reversion effects.

This chapter is organized as follows. We first introduce the methodology used for our analysis. Second, we present our results for broad-based equity indices and underline the evidence supporting market momentum and market reversion.

Also, for further references on the subjects of market predictability, market efficiency, and the historical developments of models of equity returns, we refer the reader to the review of mathematical finance in the appendices.

4.1 Our Methodology

4.1.1 Issue of Market Predictability

Our interest in analyzing the serial properties of equity returns stemmed from the controversial issue of market predictability and the historical development of various models to account for the statistical properties of equity returns.

The issue of market predictability has been studied for a long time and by various communities, from speculators to academic members of the mathematical finance community. It is not hard to understand that if there was any pattern of predictability that could be found in equity markets, this pattern would be quickly exploited and we would expect arbitrageurs to wash it away. Historically however, the issue of predictability of equity returns has been a consistent topic of research. One of the reasons for this continued interest is the fact that the topic of market predictability (or lack thereof) has been closely tied to the issue of market efficiency (we refer the reader to our appendix on some historical elements of mathematical finance for further references on the subject). Many academics have contended that markets are unpredictable and many books and articles have been written on this topic. Starting with Bachelier in 1900 (Bachelier 1900), who introduced Brownian motion ("BM") to finance, models of equity returns have been laid out that describe the price of security as being essentially unpredictable in direction (at least once certain adjustments to returns have been made). We mean by direction that it would be impossible for a forecaster to guess whether the price is going to move up or down in the next time period. Later on, for various reasons, academics realized geometric Brownian motion ("GBM") was a more accurate model and Osborne provided detailed evidence to support this assertion (Osborne 1959). References on GBM and its usefulness for describing the stock market are given in Osborne's paper. In a discrete setting (which is our focus here as we will look at daily returns derived from daily closing prices), this implies that daily returns should follow a random walk. This expression of "random walk" was popularized to a large extent by Malkiel in his book "A Random Walk Down Wall Street" (Malkiel 2003). Malkiel defines it as:

> A random walk is one in which future steps or directions cannot be predicted on the basis of past actions. When the term is applied to the stock market, it means that short-run changes in stock prices cannot be predicted... On Wall Street, the term "random walk" is an obscenity. It is an epithet coined by the academic world and hurled insultingly at the professional soothsayers. Taken to its logical extreme, it means that a blindfolded monkey throwing darts at a newspaper's financial pages could select a portfolio that would do just as well as one carefully selected by the experts.

It is this issue of market predictability that we intend to analyze here, with a new methodology. It is a very important and controversial issue, and as Malkiel puts it:

> By the early 2000s, even some academics [have joined] the professionals in arguing that the stock market was at least somewhat predictable after all. Still, as [one] can see, there's tremendous battle going on, and it's fought with deadly intent because the stakes are tenure for the academics and bonuses for the professionals.

4.1.2 Our Restricted Formulation

We analyze market predictability in a restricted sense, that is by only looking at the "directional" moves of the market as a whole, or an individual security. What we mean by "directional move" is whether the market index or security price of concern is going to move up or down over the next time period.

We focus here on daily returns derived from closing prices. Formally, the closing price (or level) of the underlying (whether the underlying is taken to be a market index, a sectorial index or an individual security price) will be defined as P_t and the simple return r_t is defined as:

$$r_t = \frac{P_{t+1}}{P_t} - 1, \quad 0 \leq t \leq T-1, \tag{4.1}$$

where T is the time horizon of our analysis. From this time series r_t, we define:

$$b_t = \text{sign}(r_t), \tag{4.2}$$

where

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{otherwise.} \end{cases}$$

From the series b_t, we derive the series s_t as $s_0 = b_0$ and:

$$s_{t+1} = \begin{cases} s_t + \text{sign}(b_{t+1}) & \text{if } s_t b_{t+1} > 0 \\ \text{sign}(b_{t+1}) & \text{otherwise} \end{cases} \quad for \quad 0 \leq t \leq T-1.$$

The construction of the series s_t will allow us to track the number of consecutive days the price process P_t has been going up or down. Figure 4.1 shows how the series s_t is constructed.

By only looking at the directional moves of the price process P_t and focusing on the signs of the daily returns r_t, we lose the finer details of the distribution of r_t. To understand this, assume r_t to be drawn from the probability density function ("PDF") of a stationary distribution $f_{\mathbf{r}}(r)$ and to be independent of each other. Also assume that our goal was to predict over T observations the series $b_t = \text{sign}(r_t)$ with the greatest precision. Mathematically, if we call f_t[1] our forecast of the underlying move over the next period, we can define

[1]The notation used for the series of forecasts, f_t, is obviously not to be confused with the PDF $f_{\mathbf{r}}(r)$.

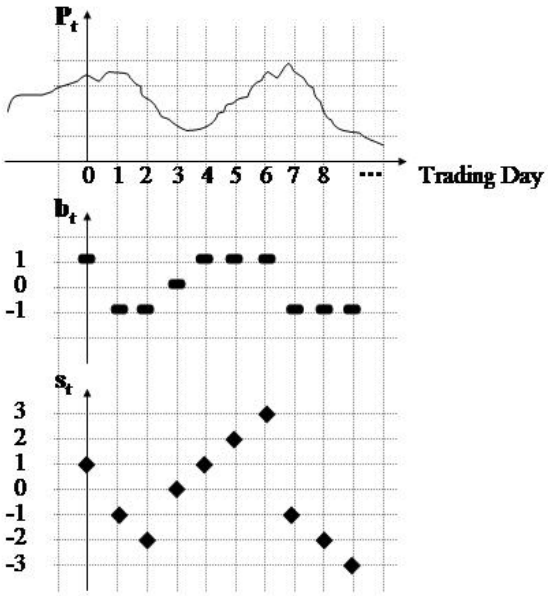


Figure 4.1: Mapping of the “underlying” price level into the series s_t that represents the number of cumulative and consecutive upward or downward movements of the original series P_t sampled discretely (at the end of each trading day). We have omitted on this graph the series r_t that is easily derived from the series P_t.

the forecaster's performance metric, Z_T, as:

$$Z_T = \frac{\sum_{t=0}^{T-1} 1_{\{f_t=b_t\}}}{T},$$

where $1_{\{f_t=b_t\}} = 1$, if $f_t = b_t$ and 0 otherwise. It is easy to show that the best forecasting strategy at each t, where $0 \leq t \leq T-1$, would be:

$$f_t = \begin{cases} -1 & \text{w.p. } p_{-1} \\ 0 & \text{w.p. } p_0 \\ 1 & \text{w.p. } p_1, \end{cases}$$

where:

$$p_{-1} = \int_{-1}^{0^-} f_{\mathbf{r}}(r)dr \tag{4.3}$$

$$p_0 = \quad Pr\{\mathbf{r} = 0\} \tag{4.4}$$

$$p_1 = \int_{0^+}^{+\infty} f_{\mathbf{r}}(r)dr. \tag{4.5}$$

So all we would need to know for our optimal forecast would be the cumulative distribution function ("CDF") value $F_{\mathbf{r}}(0^-) = \int_{-1}^{0^-} f_{\mathbf{u}}(u)du$ (which is p_{-1}), as well as the probability mass p_0 at 0 if there is any (and consequently $p_1 = 1 - p_{-1} - p_0$). So two CDFs $F_{\mathbf{r}}^1$ and $F_{\mathbf{r}}^2$ with the same "cutoff" value at 0 (and, if there exists one, the same punctual probability mass at 0) would give us the same forecasting policy.

Clearly, for many real applications (such as portfolio management), we would want to estimate in finer details the PDF $f_{\mathbf{r}}(r)$. For instance, if we assume over the time horizon T that the daily returns r_t will be independent and identically distributed as $r_t \sim N_{-\mathbf{1}}(0, \sigma^2)$, where $N_{-\mathbf{1}}$ is just a normal distribution that has been truncated below at -1 and renormalized, we would want to estimate the volatility σ. Hence, our analysis is restricted in the sense that we only focus on the sign of the returns r_t in our mapping of the series r_t into b_t. By doing so we disregard the information on the returns distribution embedded in the distribution moments (second-, third- and higher-order).[2] However, as our analysis will later show, the CDF "cutoff" value is all we need to display strong evidence of serial correlation in the returns of equity indices.

[2]We wish in no way to suggest to the reader that two probability distributions with the same moments *ad infinitum* have to be equivalent, as this is not the case.

4.1.3 Intuition for Methodology

There were different intuitive categories for undertaking this particular investigation. First, we believed that by and large markets did not always reflect all available information and that when markets close, we should expect to have a gradual diffusion of information among traders and the formation of consensual views from one day to the next. This, we expected, should be reflected by some market momentum over a couple of days. Second, we thought that market movements should be conditional to some extent upon their recent history so that if a market had been going up for a consecutive number of days x, the probability that it would go up for another day would decrease with x. This so-called mean reversion should be observable by looking at conditional market transitions. Last, from a statistical standpoint, focusing on the distribution of returns alone ignores the crucial information that is embedded in their development as a time series. This can be addressed by preserving the information contained in the autocorrelation functions of different orders. But even that is not enough as these summary statistics focus on the sample as a whole and provide average results for the sample, hence possibly missing "local patterns of serial correlation" in the sample. In other words, any sudden burst of correlation in the studied sample would go undetected if through the rest of the sample there negative correlation to counterbalance the local burst. This point is illustrated in Figure 4.2 and Figure 4.3.

Figure 4.3 shows two series with approximately the same first-order autocorrelation but strikingly different paths. Figure 4.2 shows the cumulative number of days with positive returns, and an artificially created spike in the first-order autocorrelation of the graph's time series A.

4.1.4 Previous Evidence of Momentum

We provide here an analysis on momentum in market indices. This will show that successive price variations are not necessarily independent. As Sornette shows (Sornette 2003), there is strong evidence of local burst of serial correlation by looking at drawdowns (defined as a "persistent decrease in the price over consecutive days"). Sornette writes that:

> Drawdowns are indicators that we care about: they measure directly the cumulative loss that an investment may suffer. They also quantify the worst-case scenario of an investor buying at the local high and selling at the next minimum. It is thus worthwhile to ask if there is any structure in the distribution of drawdowns absent in that of price variations... Their distribution [captures] the way successive drops can influence each other and construct in this way a

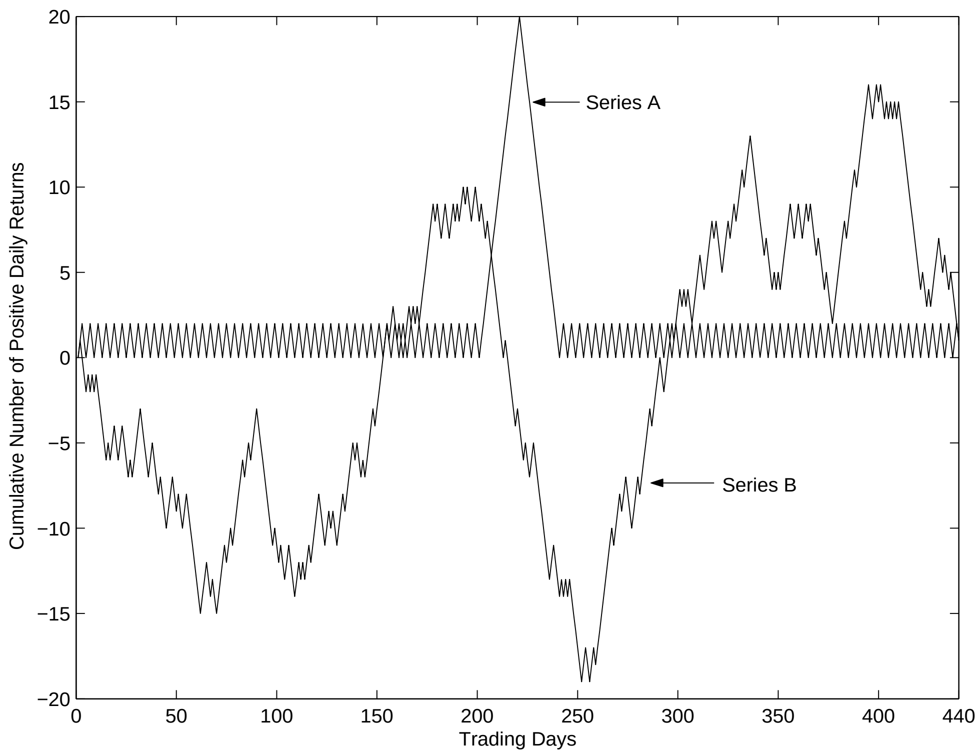


Figure 4.2: The two series were generated as follows. Series b_t^A (series A on the graph) was generated first with an artificially constructed local "burst" in serial correlation. The rest of the series was designed with zero first-order autocorrelation so that, all in all, the sample's first-order autocorrelation estimate is slightly positive. Series b_t^B (Series B on the graph) was generated by a constrained permutation so that the new sample's first-order autocorrelation, $\rho^1_{b_t^B}$, is similar to Series A's, i.e. $\rho^1_{b_t^A} \simeq \rho^1_{b_t^B}$.

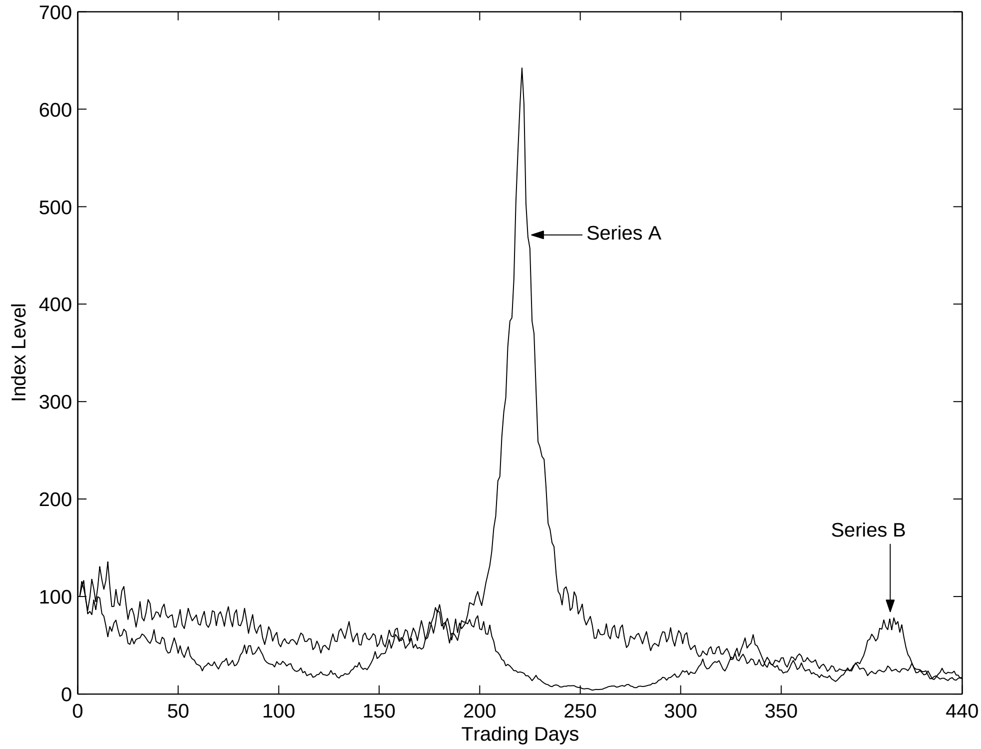


Figure 4.3: The index levels represented here correspond to the two series shown in Figure 4.2. The returns for series A were generated from r_t as $r_t \sim uniform(0; 0.2)$ if we wanted the corresponding $b_t > 0$ or conversely $r_t \sim uniform(-0.2; 0)$ to have $b_t < 0$. Series B was generated by a constrained permutation so that $\rho^1_{b^A_t} \simeq \rho^1_{b^B_t}$ and $\rho^1_{r^A_t} \simeq \rho^1_{r^B_t}$ and in both cases close to 0.

Rank	Starting Time	Index Value	Duration (days)	Loss
1	87.8	2508.2	4	-30.7%
2	14.6	76.7	2	-28.8%
3	29.8	301.2	3	-23.6%
4	33.5	108.7	4	-18.6%
5	32.2	77.1	8	-18.5%
6	29.9	238.2	4	-16.6%
7	29.8	273.5	2	-16.6%
8	32.6	67.5	1	-14.8%
9	31.9	90.1	7	-14.3%
10	32.7	76.5	3	-13.9%

Table 4.1: Characteristics of DJIA's 10 Largest Drawdowns in 20th Century. The "starting time", instead of being in standard date format, is expressed in "decimal years".

persistent process.

As Sornette points out, this persistence cannot be captured by the distribution of returns alone (i.e. by only counting the frequency of returns) as by unravelling the returns series in this way, we forget everything about the relative positions of returns as a function of time. Similarly, the local persistence of returns cannot be captured by a global analysis of the two-point correlation functions in the data sample studied as it measures an average linear dependence over the whole time series, while the dependence may only appear at special times, for instance for very large runs. Sornette provides in a study with Johansen significant evidence that "large stock market price drawdowns are outliers" (Johansen and Sornette 2001).

From the DJIA historical values, we can easily extract the characteristics of the 10 largest drawdowns of the DJIA in the twentieth century. Table 4.1 summarizes these values.

4.2 Our Results

Sornette shows in his study, by means of probabilistic arguments, that financial crashes are "outliers". However, Sornette focuses on the special cases of major drawdowns and crashes. We extend this analysis of local burst of persistence in the returns by searching for any kind of momentum, not just in the case of major market downturns.

The intuition and methodology behind our investigation are different from those of any study previously mentioned. Typically a financial crash will involve a few consecutive days of highly negative returns. Thus the persistence that a crash will exhibit will be very localized (over a few days) and rare (large negative returns). As we mentioned before, the basis for our idea was to investigate for momentum (and reversion) in both good and bad times, independently of the amplitudes of returns. Though in this work we will not dwell on identifying the causes of directional persistence in market movements, it could reflect a combination of factors such as the gradual formation and diffusion of new beliefs or other "herding" effects.

The data we have analyzed covers general stock indices, sectorial indices and individual stocks, though we will only present here results found for equity indices.

4.2.1 Conditional Probabilities of Transition

We look at daily data for three major indices, namely the Dow Jones Industrial Average, the S&P 500 and the Nasdaq.

Dow Jones Analysis

The data analyzed comprises the daily Dow Jones adjusted closing levels from 01/02/1970 to 12/31/2003. Figure 4.4 shows the daily momentum analysis for the Dow Jones.

S&P 500 Analysis

The data analyzed comprises the daily closing levels for the S&P 500 from 07/02/1962 to 12/31/2003. Figure 4.5 shows the daily momentum analysis for the S&P 500.

Nasdaq Analysis

The data analyzed comprises the daily levels of the Nasdaq from 12/14/1972 to 12/31/2003. Figure 4.6 shows the daily momentum analysis for the Nasdaq.

4.2.2 Conditional Returns

Table 4.2 shows the returns corresponding to particular transitions for the Dow Jones. Table 4.3 shows the returns corresponding to particular transitions for the S&P 500. Table 4.4 shows the returns corresponding to particular transitions for the Nasdaq.

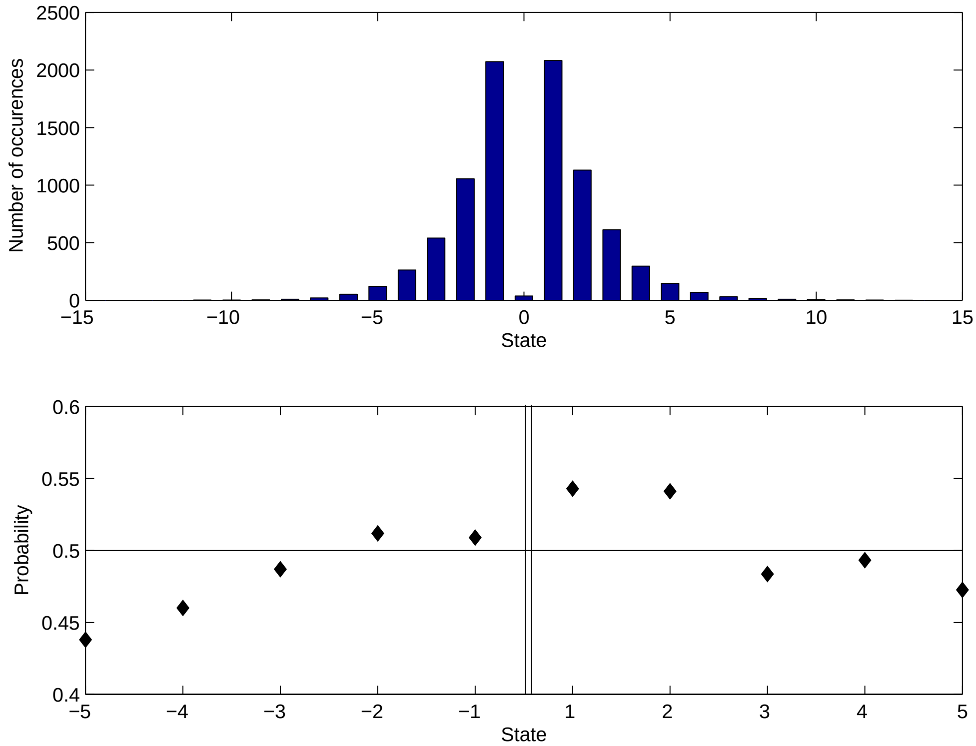


Figure 4.4: Daily Momentum Analysis of DJIA. The top graph shows the frequency of the visits of the series s_t to each state. The bottom graph zooms in on the estimated transitional probabilities from states -5 to state 5 excluding state 0. This graph shows for instance that, conditional on s_t being in state 1 (resp. 2), the probability that the Dow Jones will close up another day and s_{t+1} end up in state 2 (resp. 3) is almost 55% chance. Similarly we see that, conditional on s_t being in state -1 or -2, the probability that the DJIA will close down another day is higher than 50% as well. These transitional probabilities for states -2, -1, 1 and 2 suggest a momentum effect over three days for the DJIA index variations. Conversely, we can see that for states -5 to -3 and states 3 to 5, we have the opposite effect as the transitional probabilities for continuing the DJIA consecutive moves in one direction are less than 50%. This underlines market reversion beyond three days of consecutive moves for the DJIA equity index.

State s	$\mathrm{E}[r \mid s \to -1]$	$\mathrm{E}[r \mid s \to +1]$	$\mathrm{Pr}[r \mid s \to -1]$	$\mathrm{E}[r \mid s]$
1	-0.0062	0.0076	0.4642	0.0012
2	-0.0061	0.0072	0.4617	0.0011
3	-0.0062	0.0069	0.5542	-0.0002
4	-0.0068	0.0060	0.5274	-0.0007
5	-0.0066	0.0054	0.5268	-0.0009

Table 4.2: Dow Jones Expected Returns by Transition

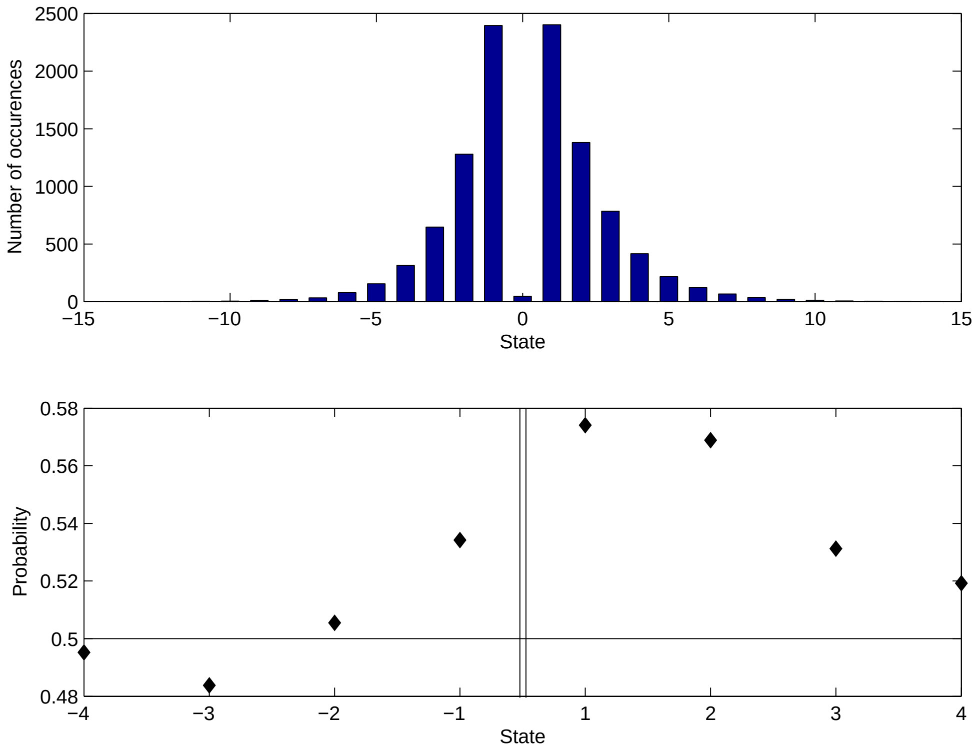


Figure 4.5: Daily Momentum Analysis of S&P 500. The top graph shows the frequency of the visits of the series s_t to each state. The bottom graph zooms in on the estimated transitional probabilities from states -4 to state 4 excluding state 0. This graph shows for instance that, conditional on s_t being in state 1, the probability that the Dow Jones will close up another day and s_{t+1} end up in state 2 is almost 58% chance. Similarly we see that, conditional on s_t being in state -1 or -2, the probability that the S&P 500 will close down another day is higher than 50% as well. These transitional probabilities for states -2, -1, 1, 2 (we're ignoring 3 and 4 here) confirm, as for the DJIA, a momentum effect of at least three days for the S&P 500 index variations. Conversely, we can see that for states -4 and -3, we have the opposite effect as the transitional probabilities for continuing the S&P 500 consecutive moves downwards are less than 50%. As for the DJIA, we have the bell-shaped curve centered around 0 that indicates momentum over the first few days followed by reversion.

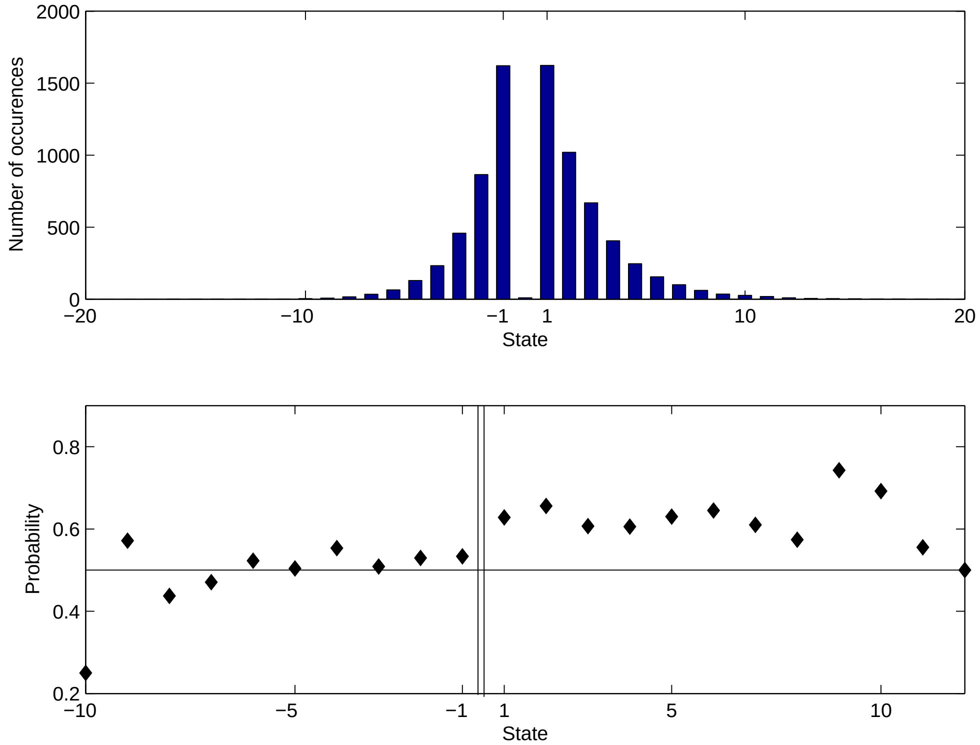


Figure 4.6: Daily Momentum Analysis of Nasdaq. The top graph shows the frequency of the visits of the series s_t to each state. The bottom graph zooms in on the estimated transitional probabilities from states -10 to state 10 excluding state 0. This graph shows that most transitional probabilities are above 50%. However the most significant states are states -3 to 3 (excluding 0). We see that conditional upon s_t being in state 1, 2 or 3, the probabilities for the NASDAQ to close up one additional day is about 60%.

State s	$\mathrm{E}[r \mid s \to -1]$	$\mathrm{E}[r \mid s \to +1]$	$\Pr[r \mid s \to -1]$	$\mathrm{E}[r \mid s]$
1	-0.0064	0.0068	0.4343	0.0010
2	-0.0055	0.0065	0.4448	0.0012
3	-0.0054	0.0063	0.5020	0.0004
4	-0.0058	0.0053	0.5054	-0.0003
5	-0.0053	0.0049	0.4670	0.0001

Table 4.3: S&P 500 Expected Returns by Transition

State s	$\mathrm{E}[r \mid s \to -1]$	$\mathrm{E}[r \mid s \to +1]$	$\Pr[r \mid s \to -1]$	$\mathrm{E}[r \mid s]$
1	-0.0085	0.0080	0.3767	0.0018
2	-0.0073	0.0065	0.3476	0.0017
3	-0.0064	0.0068	0.4122	0.0013
4	-0.0067	0.0063	0.4000	0.0011
5	-0.0050	0.0058	0.4044	0.0014

Table 4.4: Nasdaq Expected Returns by Transition

4.3 Analysis of Statistical Significance

4.3.1 Sensitivity Analysis

Figure 4.7 shows how sensitive transition probabilities are for higher states that are not often visited. In this graph we take an artificially constructed sample of market movements (Series A) that is such that the series s_t^A derived from it has a perfectly split distribution at every state. This is constructed as follows: suppose we have a process that is such that we have 1024 visits to state 1 (using our previous terminology). Half of these visits to state 1 are continued on to state 2, yielding 512 visits to state 2 and so forth until we reach 2 visits to state 10 and 1 visit at state 11. For this last visit to state 11, we assume this cumulative run up is then reverted (in other words, there would be a t in our sample such that $s_t = 11$ and $s_{t+1} = -1$). Hence, for this perfectly split construction, we would estimate conditional transition probabilities of 50% at every state except for the last state, state 11, where we would estimate $P(s_{t+1} = 12 \mid s_t = 11) = 0$. Now, suppose we perturb this process in two different ways to perform a sensitivity analysis of the conditional transition probabilities. First, we add to our original artificial series a run up to state 12 (this is our Series B). So Series B would have exactly the same number of visits than Series A at every state, increased by 1 (1025, 513, etc.). Similarly, we construct another series (Series C) by subtracting from the original series (Series A) the last visit to state 11. In other words, at every state, the number of visits of Series C would be the same as Series A, minus 1 (1023, 511, etc.). We compute the conditional transition probabilities for Series B and Series C and, as expected, can see that they are similar to Series A for the "low" states but become quite erratic for the "high" states.

4.3.2 Bootstrap Analysis

To measure the statistical significance of the runs or drawdowns observed, we perform a bootstrap analysis, which assumes independence of returns. There is a close connection

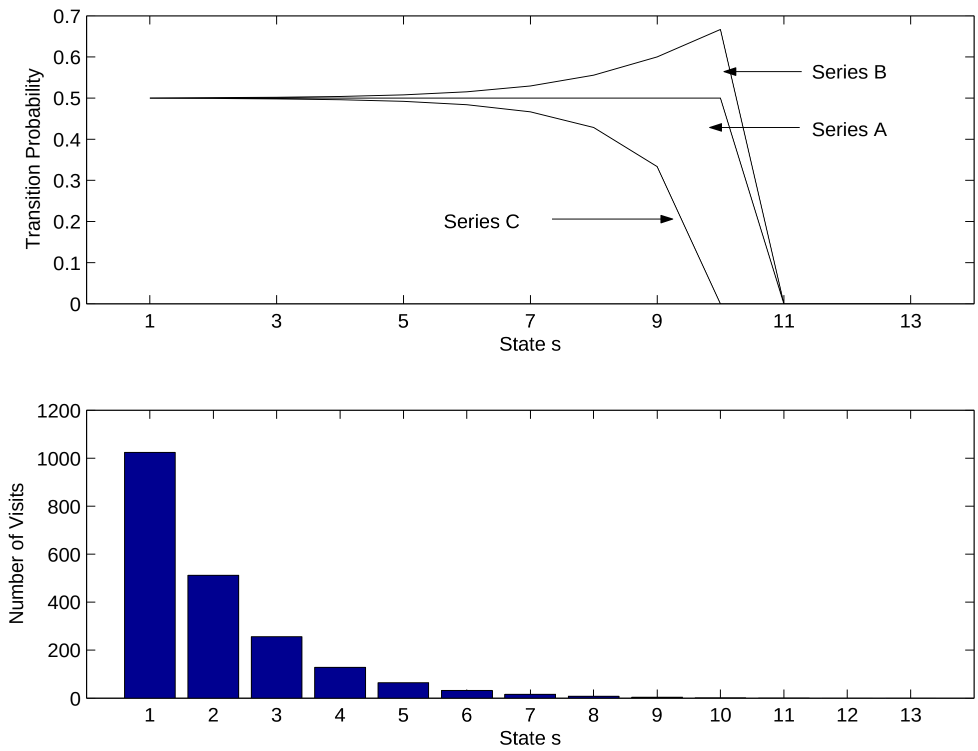


Figure 4.7: Sensitivity Analysis on Conditional Transition Probabilities. Series A is the artificially constructed series s_t^A with conditional transition probabilities of 50% (except for the last state 11. The bottom graph shows the number of visits in each state for Series A (1024 visits for state 1, 512 for 2, etc.) The top graph shows the conditional transition probabilities obtained for Series A and its derivatives Series B and Series C. Series B and Series C are obtained by adding or subtracting a visit to all states, respectively. So Series B conditional probabilities will be calculated using (1025, 513, 257, etc.) as the series representing the number of visits to each state.

between the bootstrap method and permutation tests of statistical significance and for further references, we refer the reader to the chapter contained in the work by Efron and Tibshirani (Efron and Tibshirani 1993). The principle involved is to reshuffle the returns randomly from the original sample to see if we observe such runs or drawdowns in the reshuffled process. If we do not, it implies that the assumption of independence cannot hold for the original series of historical returns, which is precisely what we want to prove. Figure 4.8 displays the results for 100 replications with the upper and lower limits observed for the transition probabilities. The results are conclusive except for the DJIA drawdowns. The S&P 500 clearly shows statistical significance for states 1 and 2 of runs and state 1 of drawdowns. However, the most significant of all is for the Nasdaq where states 1 and 2 for both runs and drawdowns are clearly outside of the min-max band.

4.3.3 Verification of the Markov Property

Though we do not need it to contend that equity indices display evidence of serial correlation, it is interesting to study if the process s_t could be considered a Markov chain. Formally what we would need to verify to make this claim is that

$$P(s_{t+1} = s_1 \mid s_t = s; s_{t'} = s_2) = P(s_{t+1} = s_1 \mid s_t = s; s_{t'} = s_3) \tag{4.6}$$

for all $t' < t$ and for all (s_1, s_2, s_3) in Z^3.

Given the finite size of our sample, it is hard to verify this formally for all possible histories of the process. What this property is simply saying is that the transition probabilities from s_t to s_{t+1} only depend on s_t and not on the past history of the process $(s_{t-1}, s_{t-2}, s_{t-3}, \dots)$. We verify this property for a memory of order 1. That is we verify that $P(s_{t+1} = s_1 \mid s_t = s; s_{t-1} = s_2) = P(s_{t+1} = s_1 \mid s_t = s; s_{t-1} = s_3)$ for all t and (s_1, s_2, s_3) in Z^3. Since we disregard transitions from state 0 (due to their low frequencies) and all other states only have one possible predecessor, we only need to verify this property for states -1 and 1. Table 4.5 and Table 4.6 show the empirical results for the DJIA index and we obtain comparable results for the S&P 500 and the Nasdaq indices.

4.4 Further Research

Further research could be done by performing the same analysis but with a different and translated sampling scheme. For instance, we could look at the daily opening prices (instead

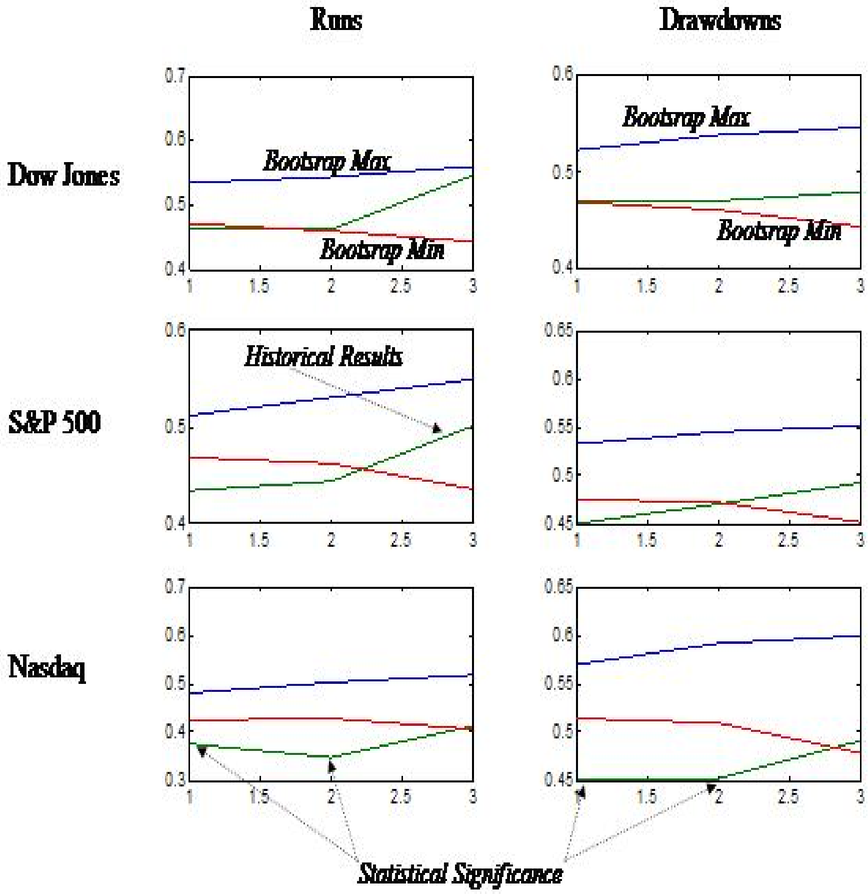


Figure 4.8: Bootstrap analysis of runs and drawdowns. On each graph, we represent the maximum and minimum transition probabilities obtained by our bootstrap procedure. That is, for each time series obtained by the bootstrap, we compute the transition probabilities as we have for the historical series (for each index). So we also represent on each graph the transition probabilities estimated from the historical realized returns. The historical transition probability is considered statistically significant if it lies outside of the "cone" defined by the bootstrap min-max. The transition probabilities represented here are the probabilities of "reversals". For instance for the "Dow Jones/Runs" graph, given that there has just been an upward market close (i.e. we are in state 1), the probability of having a downward market close the next day is about 45%. This is outside of the min-max cone defined by the bootstrap simulation and is considered statistically significant. The probabilities of "reversals" from states 2 and 3 are not considered statistically significant as they lie within the bootstrap min-max cone. Note that for the graphs of drawdowns, the states are represented in absolute values. In other words, state 1 should be read as state -1, etc.

s_{t-1}	$s_{t+1} = -2$	$s_{t+1} = 1$	$s_{t+1} = 0$	Continuation Ratio
0	8	4	0	0.6667
1	483	450	4	0.5155
2	263	251	3	0.5087
3	158	154	2	0.5032
4	72	76	0	0.4865
5	40	37	0	0.5195
6	21	18	0	0.5385
7	3	10	0	0.2308
8	4	3	0	0.5714
9	1	2	0	0.3333
10	2	1	0	0.6667
11	0	1	0	0
12	0	1	0	0
13	0	1	0	0
14	0	0	0	0

Table 4.5: Verification of the Markov Property for the DJIA. Conditional transitions from state $s_t = -1$. The left column shows the predecessor state to state -1 (possible predecessor states are $0, 1, 2, 3, \ldots$) The second column shows the number of transitions from state -1 on to state -2. The third column shows reversal from state -1 to state 1. The fourth column shows the number of transitions from state -1 to state 0. The last column computes the ratios of transitions to state -2 over the total number of transitions from state -1 with the particular predecessor state listed in the corresponding row.

s_{t-1}	$s_{t+1} = 2$	$s_{t+1} = -1$	$s_{t+1} = 0$	Continuation Ratio
-14	0	0	0	0
-13	0	0	0	0
-12	0	0	0	0
-11	1	1	0	0.5
-10	0	0	0	0
-9	1	0	0	1
-8	3	3	0	0.5
-7	5	7	0	0.4167
-6	17	15	0	0.5313
-5	41	27	0	0.6029
-4	79	61	0	0.5643
-3	140	132	4	0.5072
-2	279	227	7	0.5439
-1	549	454	5	0.5446
0	16	9	0	0.64

Table 4.6: Verification of the Markov Property for the DJIA. Conditional transitions from state $s_t = 1$. The left column shows the predecessor state to state 1 (possible predecessor states are $\ldots, -3, -2, -1, 0$) The second column shows the number of transitions from state 1 on to state 2. The third column shows reversal from state 1 to state -1. The fourth column shows the number of transitions from state 1 to state 0. The last column computes the ratios of transitions to state 2 over the total number of transitions from state 1 with the particular predecessor state listed in the corresponding row.

of closing prices) to see if the same results hold.

Chapter 5

Impact of Serial Correlation

The subject of optimal portfolio management is obviously very much intertwined with the more statistical subject of how to best model returns of financial assets. In this chapter, we want to study (for the same sample of data) how a VAR modeling of returns affects the initial allocation, as opposed to other statistical models of returns, such as GBM. As we have mentioned before, since Markowitz' ground breaking work on optimal portfolio allocation and its mean-variance efficient framework (Markowitz 1952a, Markowitz 1956, Markowitz 1987b, Markowitz 1987a), a vast body of literature has developed that uses multi-stage models in discrete time or that continuously calculates the optimal portfolio in continuous time (Campbell and Viceira 2002). Similarly, bearing a key interaction to the topic of strategic asset allocation, a vast body of literature was written throughout the last century on the topic of how to best model financial returns. We discussed this issue in the previous chapter and pointed out that this subject is closely connected to the traditionally controversial issue of whether financial returns can be deemed predictable or not. Even though this latter topic of the predictability of returns has been extensively studied in the financial community, the question of how any form of returns predictability affects portfolio optimization has not been so extensively studied. The aim of this chapter is to address this question in a restricted setting. Our restrictions are twofold: first, we assume that the problem of portfolio allocation can be properly treated within a multistage stochastic programming framework; second, we assume serial correlation of financial returns can be properly captured by means of a vector autoregressive process.

5.1 Review of Traditional Assumptions

Traditionally, the most significant assumptions that need to be made by researchers for computing dynamic portfolio allocations are: (i) how to model the investor's goals and the type of utility function that should be used; (ii) whether to include transaction costs or not, and (iii) how to best model the asset class returns involved in the problem.

5.1.1 Utility Function

As is assumed in Merton's early work (Merton 1969, Merton 1990), the utility function used is often assumed to exhibit either constant relative risk aversion or constant absolute risk aversion. However, this assumption while preserving the tractability of the allocation problem and enabling closed-form solutions to be determined is not always justified from a practical standpoint. In our approach, this assumption is relaxed. We consider a strictly concave piecewise linear utility function whose kinks correspond to explicit goals formulated by the decision-maker, at a prefixed horizon.

5.1.2 Transaction Costs

Taking transaction costs into account can be deemed a key reason for choosing a stochastic programming approach over a dynamic programming methodology. Within a dynamic programming framework, the introduction of transaction costs significantly increases the dimension of the state space. It is worth noticing that transaction costs have been historically decreasing. However, in the case of certain applications (e.g. a fund of funds investing in different hedge funds), they remain quite significant (e.g. some investment funds have significant fees when funds are withdrawn by investors and may also have upfront fixed fees when funds are invested). Thus, it is hard to a priori dismiss transaction costs and our approach seems all the more valuable in this case.

5.1.3 Stochastic Modeling of Underlying Assets

The third type of assumption needed is on choosing the best stochastic representation of asset class returns. We refer the reader to the appendix on historical developments of mathematical finance where the topic is reviewed.

Summary Statistic	Stock	Govt Bonds	Cash
Yearly Arithmetic Average	12.34%	6.09%	4.79%
Yearly Standard Deviation	16.47%	9.24%	3.05%
Monthly Arithmetic Average	0.97%	0.48%	0.39%
Monthly Standard Deviation	4.12%	2.05%	0.25%

Table 5.1: Data Summary Statistics.

5.1.4 Scope and Organization of the Chapter

Our goal in this chapter is not to take a position on this controversial discussion of if and to what extent markets can be deemed efficient and predictable, vast subjects that would go significantly beyond the scope of this chapter. Rather, we focus on two fundamental issues: first, can we detect any kind of serial correlation in financial returns by using a vector autoregressive framework; second, we want to measure the various impacts that different statistical fittings of the same data can have on optimal asset allocation decisions.

To perform our analysis, we focus mainly on two particular cases: (a) treating financial returns as if they were independent and identically distributed ("i.i.d.") and best described by geometric Brownian motion ("GBM") and (b) treating returns using a VAR model. Hence in the first case, we model asset returns as i.i.d. random variables and fit the latest historical data in a multivariate GBM framework. In the second case, we fit the same data in a VAR framework. We can them compare the two sets of results and analyze the impact of serial correlation on the optimal initial asset allocation.

The chapter is organized as follows. First, we describe the data and the estimation process used for both our GBM and VAR models. Second, we present our stochastic programming framework and its implementation details. Third, we report and discuss our results. Last, we emphasize the conclusions of this investigation.

5.2 Asset Returns Data and Model Estimates

5.2.1 Data

Table 5.2.1 provides a summary of the yearly and monthly data used, spanning from 12/31/1946 to 12/31/2003. This includes value-weighted returns including dividends for the NYSE stock index, yearly returns for the 10-year government bonds and yearly returns for the 30-day treasury bills (assumed to be a good proxy for cash). This data was obtained from the Center for Research in Security Prices database.

5.2.2 VAR Model

The vector autoregressive ("VAR") model posits a set of relationships between past lagged values of all variables in the model and the current value of each variable in the model. An introduction to this type of model can be found in the work by Chatfield (Chatfield 1996) and a more complete description in the works by Hamilton or Judge et al. (Hamilton 1994, Judge, Hill, Griffiths, Lutkepohl, and Lee 1988).

A compact way of writing a VAR model is:

$$\mathbf{y_t} = \mathbf{c} + \mathbf{B_1 y_{t-1}} + \cdots + \mathbf{B_p y_{t-p}} + \epsilon_\mathbf{t}, \tag{5.1}$$

where $\mathbf{y_t}, \mathbf{y_{t-1}}, \cdots, \mathbf{y_{t-p}}$ are n-dimensional vectors of the time series $\mathbf{y_t}$ and all its lagged values, up to order p. $\mathbf{c_t}$ represents a vector of constants and $\epsilon_\mathbf{t}$ an n-dimensional vector of independent disturbances.

So for a VAR model of lag 1, we get in expanded form:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \dots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \dots & \dots & \dots \\ b_{n1} & \dots & b_{nn} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \dots \\ y_{n,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \dots \\ \epsilon_{nt} \end{bmatrix}. \tag{5.2}$$

Using lag operators, another completely equivalent way of writing a VAR model is:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \dots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{bmatrix} + \begin{bmatrix} A_{11}(l) & \dots & A_{1n}(l) \\ \dots & \dots & \dots \\ A_{n1}(l) & \dots & A_{nn}(l) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ \dots \\ y_{nt} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \dots \\ \epsilon_{nt} \end{bmatrix}, \tag{5.3}$$

with the same notations as before and introducing $A_{ij}(l)$ such that $A_{ij}(l) = \sum_{k=1}^{p} a_{ij} l^k$, where l is the lag operator defined by $l^k(y_t) = y_{t-k}$ and p is the lag length specified by the modeler.

5.2.3 VAR Model Estimation and Choice of Lag

It can be showed that fitting a VAR model by a maximum likelihood procedure amounts to performing a set of OLS[1] regressions. For calculation details, we refer the reader to the work by Hamilton (Hamilton 1994). We conducted all subsequent estimations in MATLAB,

[1] We use here the common acronym for Ordinary Least Squares ("OLS").

using both the standard statistics package and an econometrics package developed by LeSage (LeSage 1999). For further references on MATLAB and its use for numerical applications in finance, the reader can also consult the work by Brandimarte (Brandimarte 2002).

An important issue for estimating a VAR model is choosing the time lag. We summarize below the procedure for doing so and detail some potential issues with this approach. Subsequently, we provide our VAR coefficient estimates as well as their related regression statistics.

Choosing the Time Lag

For determining the lag length, we follow the commonly used approach outlined by LeSage (LeSage 1999) that performs statistical tests of models with various lag lengths. As LeSage puts it:

> The longer lag models are viewed as unrestricted models in contrast to the shorter lag models, and a likelihood ratio statistic is constructed to test for the significance of imposing the restrictions. If the restrictions are associated with a statistically significant degradation in model fit, we conclude that the longer lag length model is more appropriate, rejecting the shorter lag model.

The test statistic used

$$LR = (T - c)(\log|\Sigma_r| - \log|\Sigma_u|) \tag{5.4}$$

is chi-squared distributed with degrees of freedom equal to the number of restrictions imposed. T is the number of observations and c is a correction factor for the degrees of freedom proposed by Sims, which is the number of variables in each unrestricted equation of the VAR model (Sims 1980). $|\Sigma_r|$ and $|\Sigma_u|$ respectively denote the determinants of the error covariance matrices from the restricted and unrestricted models.

Table 5.2 shows the likelihood ratio statistics for some sample data. This example shows the likelihood ratios as well as the marginal probability levels. Depending on the confidence level chosen (whether 95% or 99%), we would respectively choose a model of order 7 (that is with all lags up to, and including, the lag of order 7) or a model of order 4.

One of the problems with this procedure is that we may choose significantly different lags depending on the historical sample used. We show our results carrying out this procedure for both monthly and yearly data using a "rolling historical sample".

In Figure 5.1, each historical sample is made up of 120 monthly returns. The full sample of 696 monthly returns is covered as we roll the historical sample.[2] The graph shows the

[2]We use interchangeably the term "historical window" or "historical sample".

Lags Compared	LR Stat.	Probability
10-9	15.4443	0.0794
9-8	8.1336	0.5207
8-7	13.0347	0.1610
7 – 6	**20.4897**	**0.0151**
6-5	15.7754	0.0717
5-4	6.7818	0.6598
4 – 3	**30.0381**	**0.0004**
...	...	...

Table 5.2: Lag Analysis. This table shows the sequential procedure of determining the "statistically optimal" lag order by starting from a maximum lag order (specified by the modeler, which is 10 in this example) and implementing successive likelihood ratio tests down to a minimum order (also specified by the modeler, which would be 1 in this case). By comparing successively two consecutive lag orders, the modeler can stop the procedure depending on the confidence level chosen. In this case, if the desired confidence level is 95%, we would decide in favor of a model of order 7 (as the first time the marginal probability level is less than 1-95%=5% is for 1.51%, when comparing orders 7 vs. 6). If the desired confidence level is 99%, we would continue the sequential tests until we find a marginal probability level less than 1% (hence, in this case, we would decide in favor of a model of order 4).

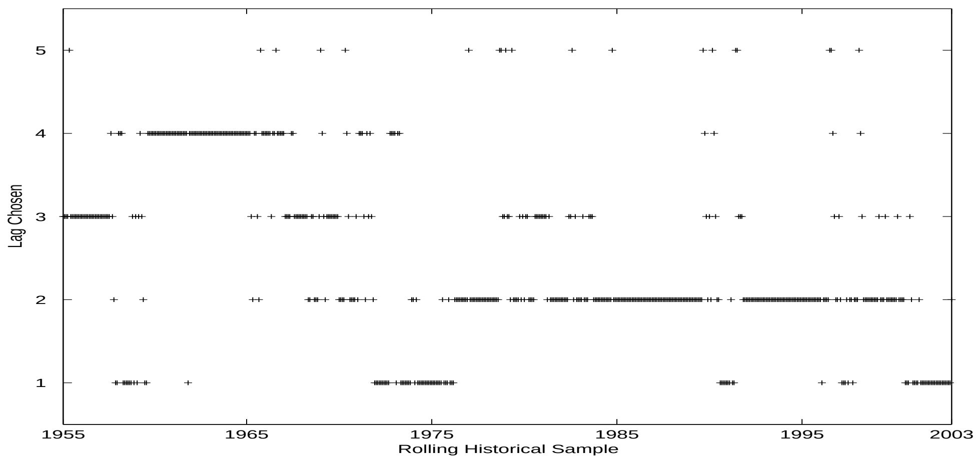


Figure 5.1: Variations of lag choices with respect to a "rolling historical sample". Each sample is composed of 120 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003. Hence the full sample is comprised of 696 monthly observations.

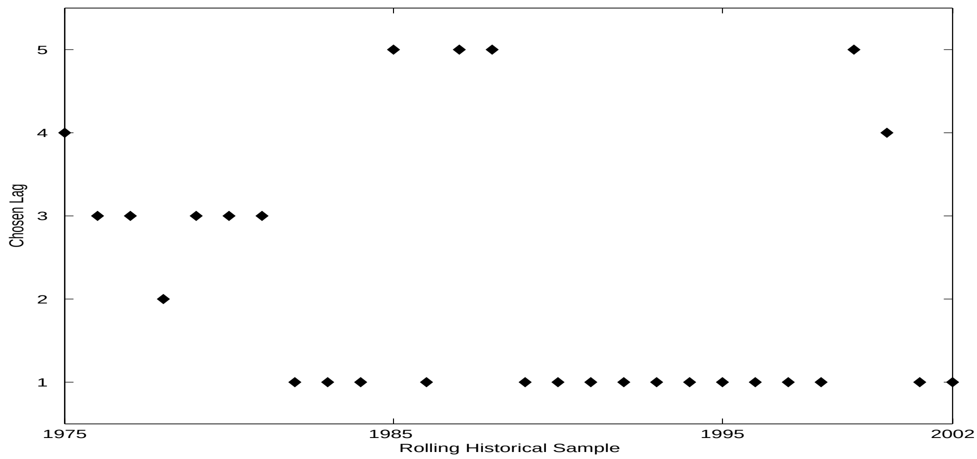


Figure 5.2: Variations of lag choices with respect to a “rolling” historical sample”. Each historical sample is composed of 30 consecutive yearly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from 1946 to 2002. Hence the full sample is comprised of 57 yearly observations.

variations of the lag orders that would be chosen by our procedure, starting with a maximum permissible lag of 5 down to a minimum lag of 1, and a confidence level of 99%. The graph shows that the lag order chosen varies a lot from one sample to another. Hence we settle for a lag order of 1. This allows us to fit a much smaller number of coefficients and diminishes the risk of overfitting.

Figure 5.2 provides the same analysis for yearly returns using a historical window size of 30 years, a maximum permissible lag of 5 (minimum lag being 1) and a confidence level of 99%. Similarly, the full sample of 57 yearly returns is covered by rolling the historical window. The variations are as significant as in the monthly case and justify (as previously with monthly data) our choice of a lag 1 model.

5.2.4 Coefficient Estimations

We provide here the coefficient estimates for both yearly and monthly data, using rolling historical windows. What our results underline are the strong variations in the estimated statistical significance of the coefficients depending on the historical sample used.

Stock Coefficients

Figure 5.3 shows the t-statistics for the stock coefficients. The graph shows that the coefficients used by the latest historical windows do not seem statistically significant (using 2.0 as the cutoff value for the absolute value of the t-statistic). Furthermore this graph shows a progressive degradation of the statistical significance (if any) of these coefficients. A possible interpretation for this trend is that throughout the sampled years (i.e. from 1946 to 2002), there has been an increase in market efficiency so that any form of stock returns predictability observed in the earlier part of our sample has slowly disappeared. Figure 5.3 graphs the lagged coefficients obtained for the stock returns.

We also include coefficient t-statistics for monthly returns for historical sample sizes of 30 months[3] (Figure 5.4) and 60 months (Figure 5.7).

Bond Coefficients

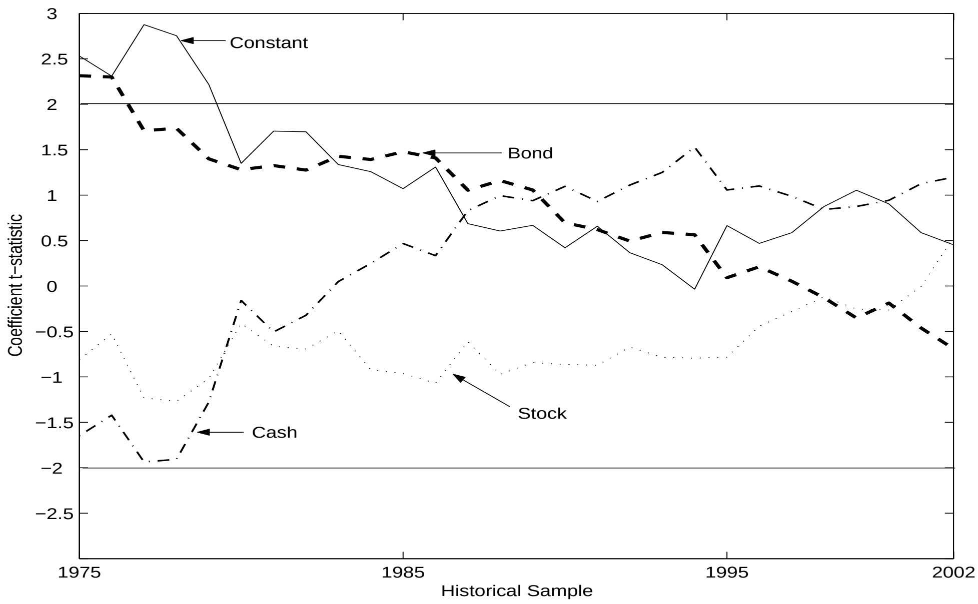
Figure 5.8 underlines the strong statistical significance of the cash returns lagged values for predicting next year's bond returns, as well as the autocorrelation of bond returns from one year to the next (though this effect is less clear). This is consistent with the traditional models of the yield curve where inter-temporal variations of the short-term bond returns induces serial correlation for longer-term bond returns. In other words, our estimates below seem to be consistent with the traditionally assumed persistence of the yield curve. Figure 5.9 graphs the lagged coefficients obtained for the bond yearly returns. Figure 5.10 and Figure 5.13 provide the t-statistics for monthly returns.

Cash Coefficients

Figure 5.14 shows we find the same statistical significance for the lagged-1 10-yr bond and cash returns on the current cash returns, as expected for the reason previously cited for bond returns. However, another surprising finding is the significance of the stock market lagged value for predicting the current cash returns. Intuitively one can understand that there would be a certain historical relationship in how the stock market does one year and the cash returns (assumed here to be equivalent to the returns on the 30-day treasury bills). Figure 5.15 underlines the fact that the autocorrelation of cash returns is significantly more important than the other effects.

Figure 5.16 and Figure 5.19 provide the t-statistics for monthly returns.

[3] We wanted to have the same number of data points as in the yearly case.



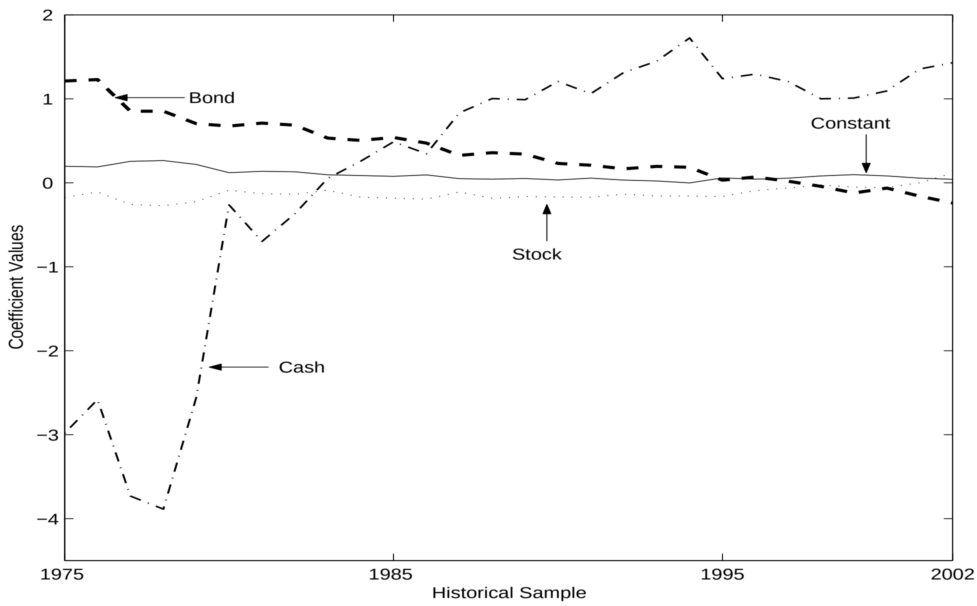


Figure 5.3: t-statistics and coefficient values for stock yearly returns. Each historical sample is composed of 30 consecutive yearly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from 1946 to 2002. Hence the full sample is comprised of 57 yearly observations.

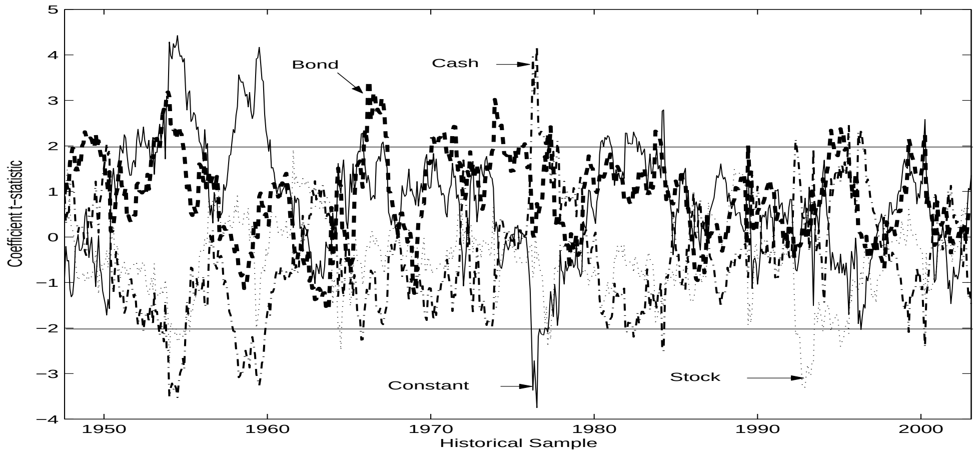


Figure 5.4: t-statistics of lagged coefficients on stock monthly returns. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003. Hence the full sample is comprised of 696 monthly observations.

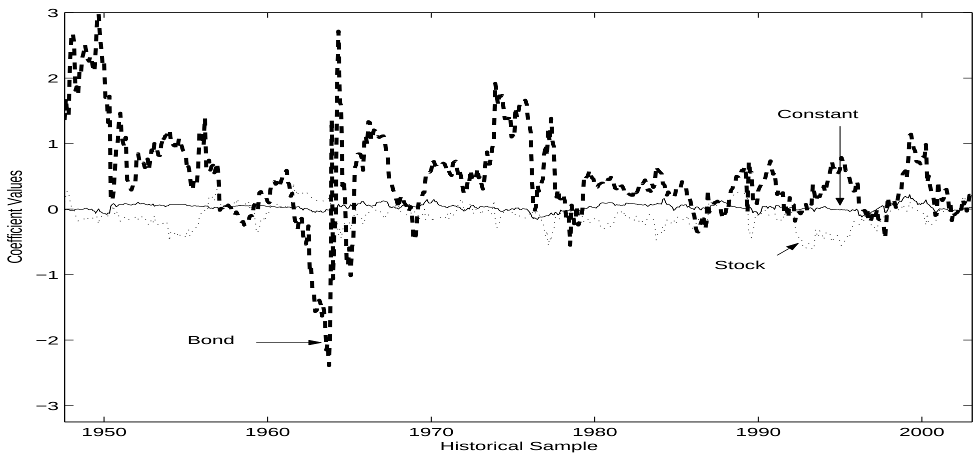


Figure 5.5: Values of lagged coefficients on stock monthly returns. This shows the values of the lags on stock and bond returns, as well as the constant. The cash coefficient value is shown separately. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003.

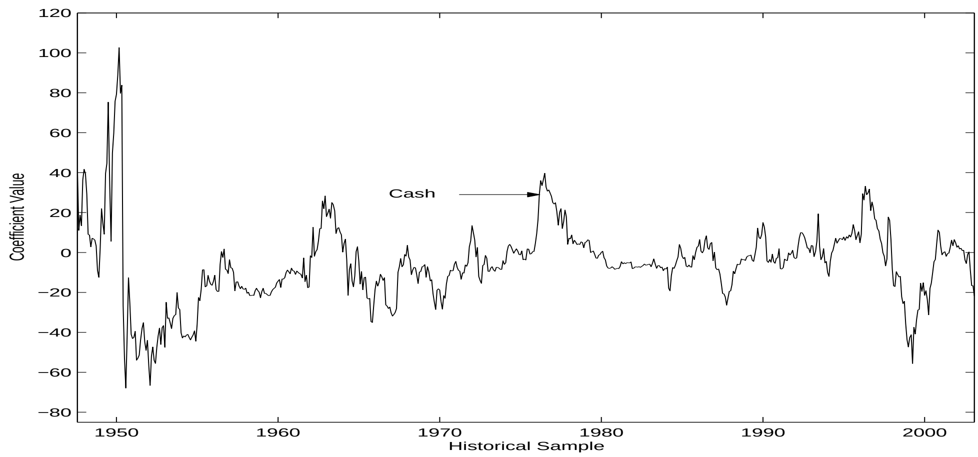


Figure 5.6: Value of lagged coefficient for cash on stock monthly returns. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003.

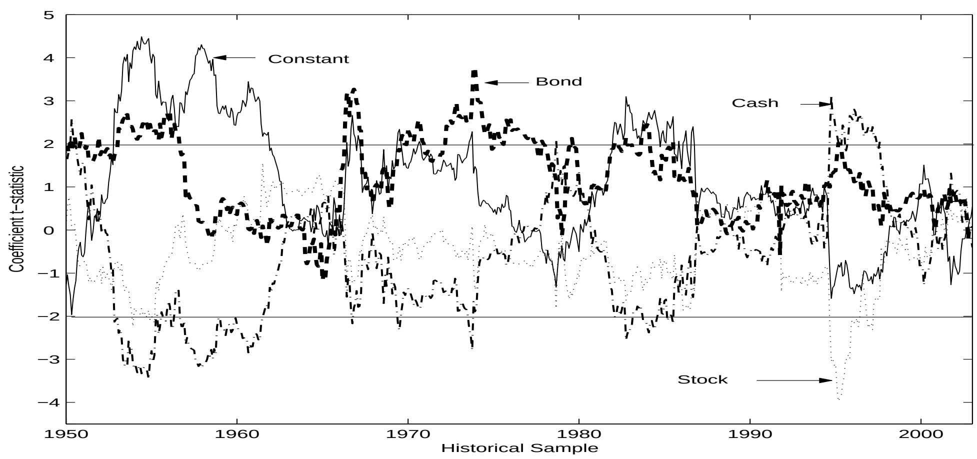


Figure 5.7: t-statistics of lagged coefficients on stock monthly returns. Each sample is composed of 60 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003. Hence the full sample is comprised of 696 monthly observations.

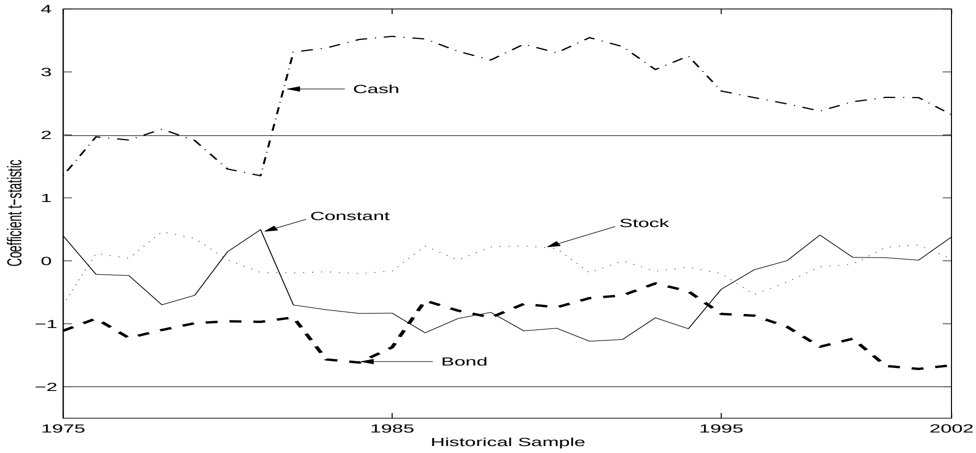


Figure 5.8: t-statistics for lagged coefficients on bond yearly returns. Each historical sample is composed of 30 consecutive yearly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from 1946 to 2002. Hence the full sample is comprised of 57 yearly observations.

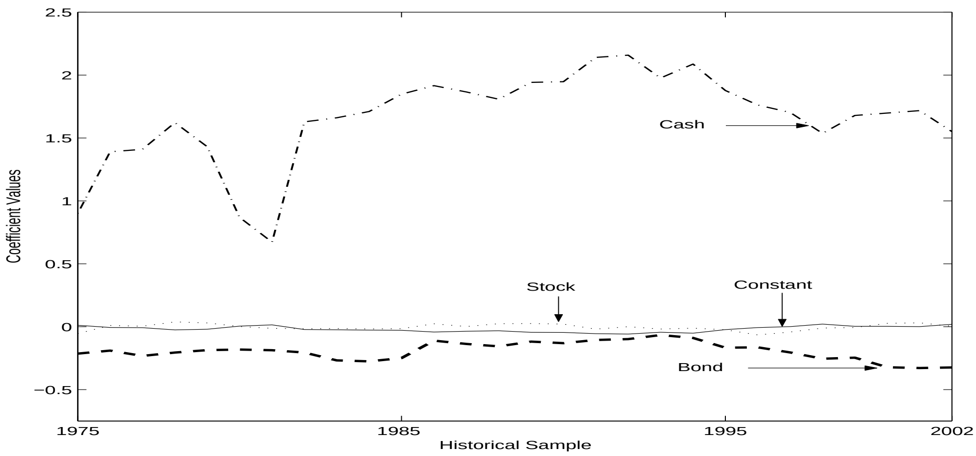


Figure 5.9: Values of lagged coefficients for bond yearly returns. Each historical sample is composed of 30 consecutive yearly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from 1946 to 2002. Hence the full sample is comprised of 57 yearly observations.

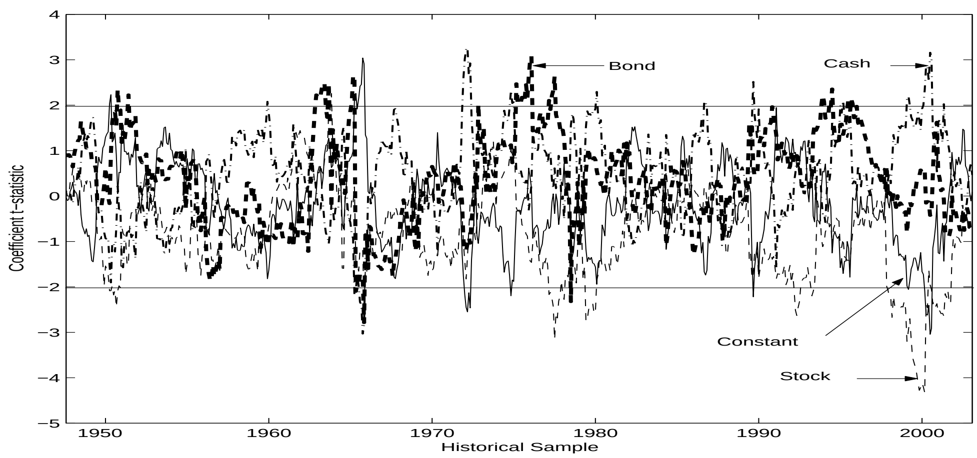


Figure 5.10: t-statistics for lagged coefficients on bond monthly returns. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003. Hence the full sample is comprised of 696 monthly observations.

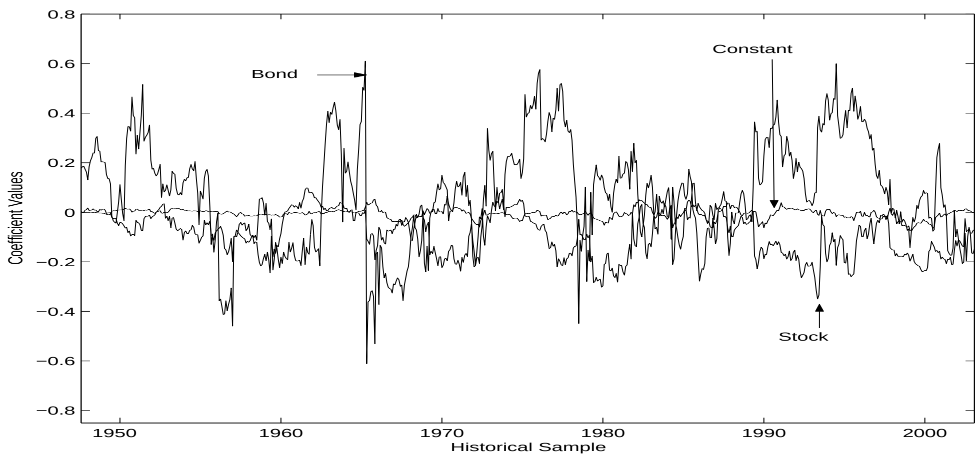


Figure 5.11: Values of lagged coefficients on bond monthly returns. This shows the values of the lags on stock and bond returns, as well as the constant. The cash coefficient value is shown separately. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003.

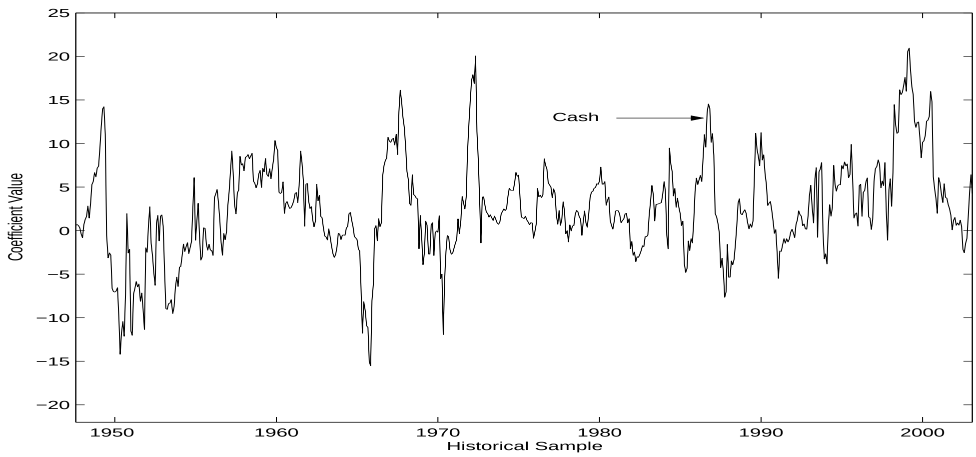


Figure 5.12: Value of lagged coefficient for cash on bond monthly returns. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003.

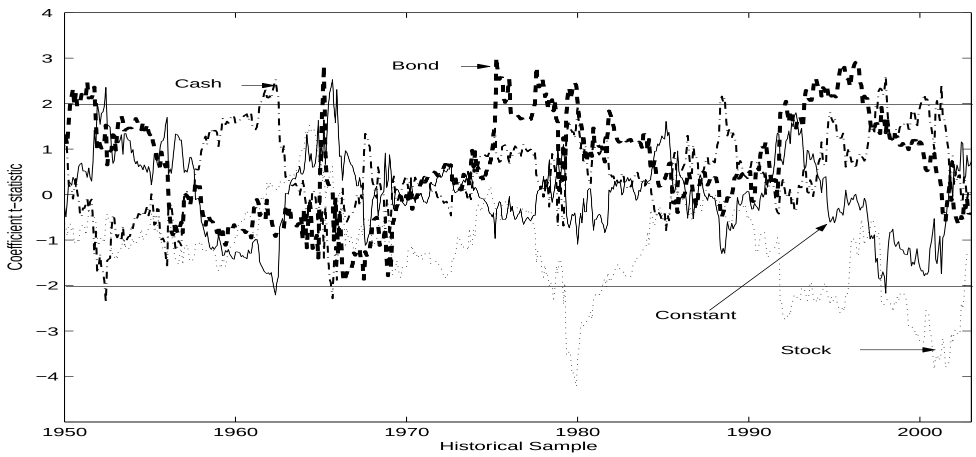


Figure 5.13: t-statistics for lagged coefficients on bond monthly returns. Each sample is composed of 60 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003. Hence the full sample is comprised of 696 monthly observations.

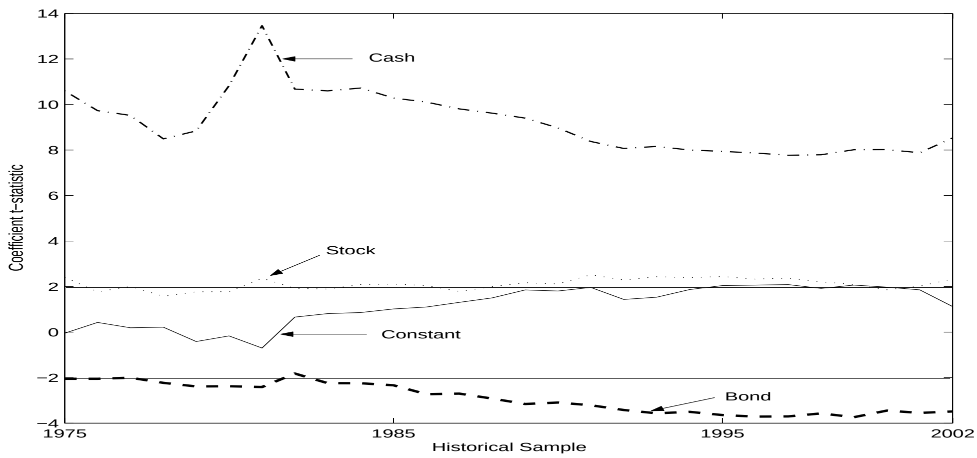


Figure 5.14: t-statistics for lagged coefficients on cash yearly returns. Each historical sample is composed of 30 consecutive yearly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from 1946 to 2002. Hence the full sample is comprised of 57 yearly observations.

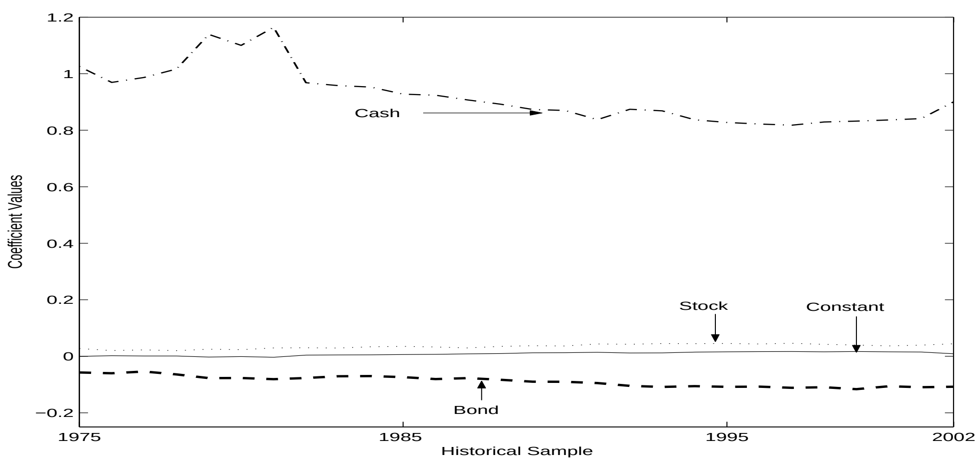


Figure 5.15: Values of lagged coefficients for cash yearly returns. Each historical sample is composed of 30 consecutive yearly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from 1946 to 2002. Hence the full sample is comprised of 57 yearly observations.

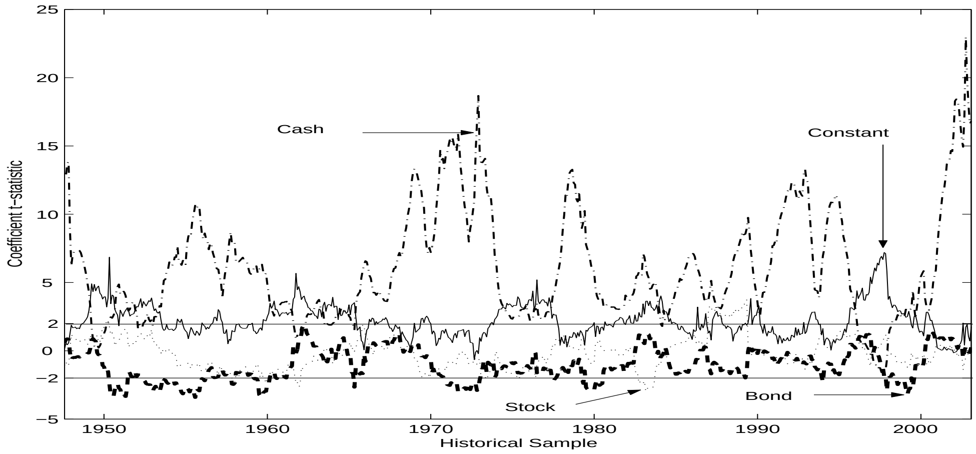


Figure 5.16: t-statistics for lagged coefficients on cash monthly returns. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003. Hence the full sample is comprised of 696 monthly observations.

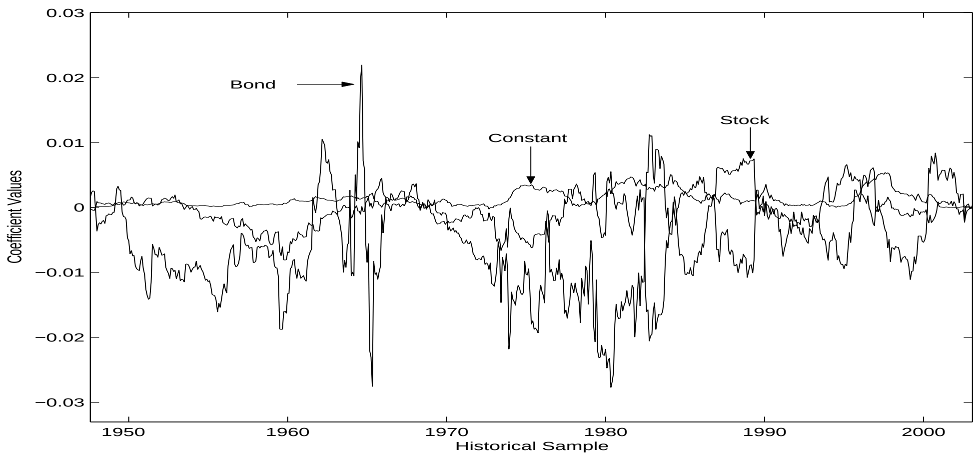


Figure 5.17: Values of lagged coefficients on cash monthly returns. This shows the values of the lags on stock and bond returns, as well as the constant. The cash coefficient value is shown separately. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003.

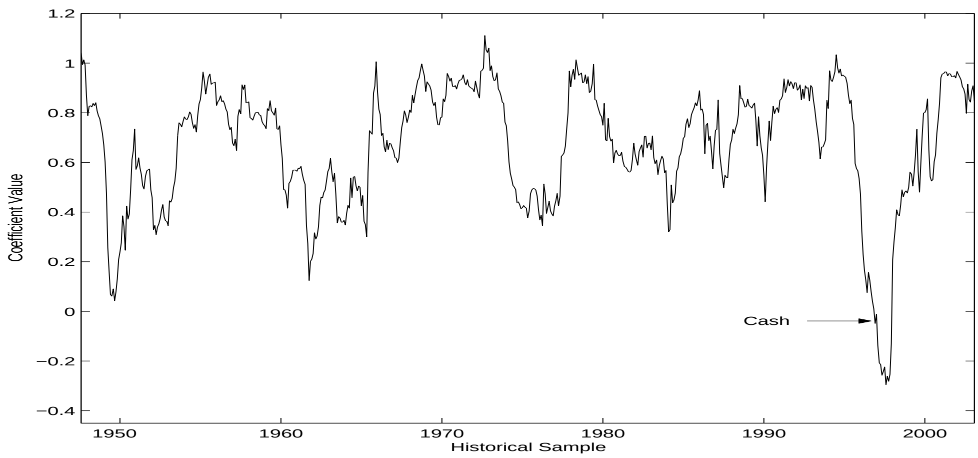


Figure 5.18: Value of lagged coefficient for cash on cash monthly returns. This shows the value of the lag on cash. Each sample is composed of 30 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003.

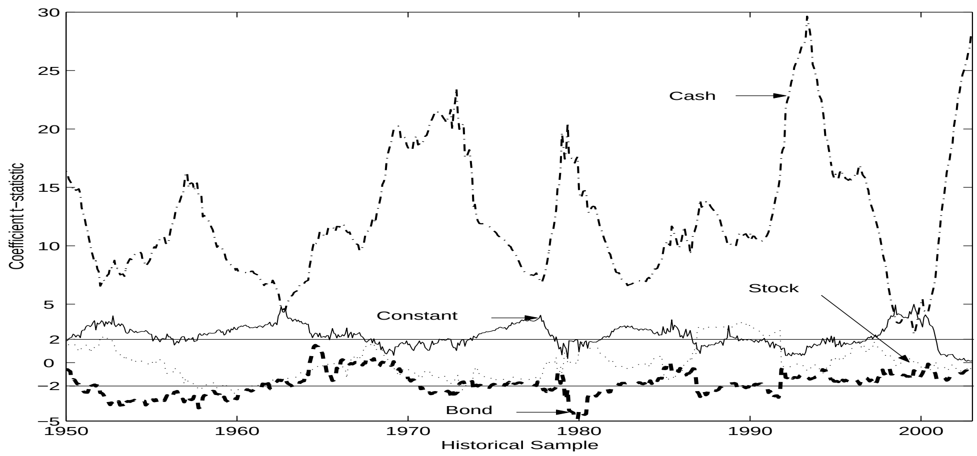


Figure 5.19: t-statistics for lagged coefficients on cash monthly returns. Each sample is composed of 60 consecutive monthly returns. The date indicated corresponds to the ending date of the sample. The full sample of returns ranges from January 1946 to December 2003. Hence the full sample is comprised of 696 monthly observations.

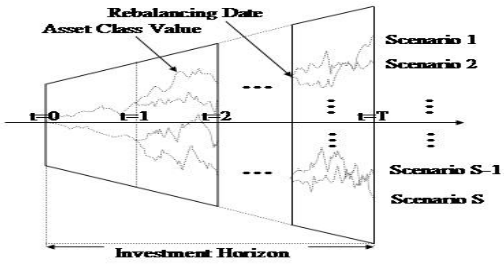


Figure 5.20: Scenario Tree

5.2.5 Bayesian Variants

A VAR model of higher order has empirical limitations that come from the large number of coefficients that usually need to be estimated. Large samples of observations involving time series that cover many years are usually needed to estimate the VAR model, and these are not always available. In addition there's a inherent tension between the large sample size needed to obtain more accurate estimates and the decreasing relevance of historical data as it is further distant in the past. Moreover, the independent variables represent lagged values, which tend to produce high correlations that lead to degraded precision in the parameter estimates. It is to overcome these problems that variants of VAR models using Bayesian prior information were introduced (Doan, Litterman, and Sims 1984). We address these latter models in the next chapter.

5.3 Model and Implementation

5.3.1 Our Stochastic Programming Framework

This formulation involves a sequence of decisions that react to outcomes (i.e. the asset class returns) that evolve over time.

We use a multi-stage stochastic linear programming framework as defined in (Infanger 1993). In general, multi-stage stochastic linear programs can be extremely large as the number of scenarios grows exponentially with the number of stages. Thus, we are limited in our application to a small number of stages. To deal with transaction costs, we have to introduce additional variables that represent how much of each asset class we have bought or sold.

The initial equation is:

$$\sum_{i=1}^{n-1} x_0^i \,(1 + btc^i) + x_0^{cash} = W_0, \tag{5.5}$$

where W_0 is the initial wealth available, $n-1$ is the number of asset classes other than money markets[4] and btc^i are the transaction costs for investing in asset class i.[5]

Between time $t-1$ and t, we have the following equations:

$$\begin{aligned} x_t^{i,\omega_1,\ldots,\omega_t} &= x_{t-1}^{i,\omega_1,\ldots,\omega_{t-1}} \, R_t^{i,\omega_t} + y_t^{i,\omega_1,\ldots,\omega_t} - z_t^{i,\omega_1,\ldots,\omega_t}, \quad i = 1...n-1, \\ x_t^{cash,,\omega_1,\ldots,\omega_t} &= x_{cash}^{i,\omega_1,\ldots,\omega_{t-1}} \, R_t^{i,\omega_t} - \sum_{i=1}^{n-1} y_t^{i,\omega_1,\ldots,\omega_t}(1 + btc^i) + \sum_{i=1}^{n-1} z_t^{i,\omega_1,\ldots,\omega_t}(1 - stc^i), \end{aligned}$$

where $y_t^{i,\omega_1,\ldots,\omega_t}$ is the amount of asset class i bought in scenario $(\omega_1, ..., \omega_t)$ and z_t^i is the amount sold in scenario t-tuple $(\omega_1, ..., \omega_t)$. The random variable R_t^{i,ω_t} is the random return of asset class i between times $t-1$ and t drawn in scenario ω_t.

For the last period, we get:

$$\sum_{i=1}^{n-1} x_{T-1}^{i,\omega_1,\ldots,\omega_{T-1}} \, R_T^{i,\omega_T} \,(1 - stc^i) + x_{T-1}^{cash,\omega_1,\ldots,w_{T-1}} \, R_T^{cash,\omega_T} = W_T^{\omega_1,\ldots,\omega_T}, \tag{5.6}$$

where stc^i are the transaction costs for selling out of the asset class i and $W_T^{\omega_1,..\omega_T}$ is the terminal wealth thereby obtained in the scenario T-tuple $(\omega_1, ..., \omega_T)$. The utility function is represented by slack variables $u_T^{\omega_1,\ldots,\omega_T}$ for being above the final goal G and $v_T^{w_1,\ldots,\omega_T}$ for being short of the goal, so that:

$$u_T^{\omega_1,\ldots,\omega_T} - v_T^{\omega_1,\ldots,\omega_T} = W_T^{\omega_1,\ldots,\omega_T} - G. \tag{5.7}$$

G is the desired goal.

[4] We assume there are no transaction costs for transferring money in and out of money markets and treat money markets as equivalent to "cash", which is our n^{th} asset class.

[5] In our model, the transaction costs are assumed to be proportional to the amount invested in the asset class.

The final objective, Z is:

$$Z = \frac{\sum_{(\omega_1,\ldots,\omega_T)\,\in\,\Omega_1\times\ldots\times\Omega_1}[u_T^{\omega_1,\ldots,\omega_T} - \psi\; v_T^{\omega_1,\ldots,\omega_T}]}{[|\Omega_1| \times \ldots \times |\Omega_T|]}. \tag{5.8}$$

In equation (5.8), ψ is the slope of the utility function below the goal G. It is a penalty factor for being short of this desired final objective.

5.3.2 Scenario Generation

There is a significant number of generally accepted methods for generating scenarios, from bootstrapping historical data to sampling from continuous distributions to obtain a pre-sampled problem, which is then solved as a substitute for the original problem.

The specification of the vector autoregressive model should be chosen carefully. Although some inter-temporal relationships between the returns might be weakly significant based on historical data as the previous section points out, that does not imply that these relationships are also useful for generating scenarios for a financial optimization model with a long term horizon. To avoid any problems with unstable and spurious predictability of returns, some authors avoid using lagged variables for explaining the returns of stocks or other asset classes in the vector autoregressive model and reserve it for obvious categories where the time series clearly exhibit some memory and serial correlation. For instance Boender in his Asset-Liability Management ("ALM") simulation system for Dutch pension funds (Boender 1997) only uses a first-order autoregressive process for modeling the returns on deposits and the variations in wage levels, and treats separately the returns on stock, bond and real estate returns included in the simulation.

5.3.3 Theoretical Considerations

It is now commonly accepted within the stochastic programming/operations research community that decomposition algorithms are not so useful when used for solving LPs in extensive form because standard LP optimizers have grown in power. Thus, when dealing with a stochastic program where the uncertain parameters are assumed to have continous distributions, a pre-sampling strategy achieved by plain Monte Carlo sampling should work fine. However, a theoretical issue that needs to be carefully addressed is to what extent a pre-sampling strategy leads to a solution similar to the one of the original problem. We know from theory that as the sample size goes to infinity, the solution of the pre-sample stochastic program epi-converges to the solution of the original problem. For finite sample sizes, we resort to bounds on the objective to establish how close the objective of the solved problem is to the objective of the original problem.

There are also other theoretical issues raised by our framework of which we need to be aware. Firstly, suppose we prefix a certain number of reallocation points. Unless we investigate numerically the pertinence of our prefixed distribution of the reallocation points, we may restrict ourselves to sub-optimal allocation policies (with respect to the performance of other asset allocation policies with the same number of stages but a different temporal distribution). Secondly, if we avoid prefixing these reallocation points but decide to optimize under the sole constraint of having a maximum number of reallocation points (whose time distribution is left a priori unconstrained), we should investigate if we can find a "rebalancing rule" (formulated for instance as an optimal stopping problem) that would trigger a reallocation decision. Thirdly, suppose we have no constraint at all on both the number and timing of reallocation points, we should compare the performances of our restricted setting's solution (where we have a prefixed, in number and temporal distribution, of reallocation stages) and the performance of the unconstrained solution (performance which will necessarily be at least as good as ours as it is unconstrained).[6]

5.3.4 DECIS Implementation

We use GAMS[7] and DECIS to solve our problem. DECIS (Infanger 1997) is a system for solving large-scale stochastic programs that can use Benders decomposition and Monte Carlo simulation with importance sampling or control variates as variance reduction techniques. Hence DECIS includes a variety of solution strategies and can solve problems with numerous stochastic parameters. For solving master and subproblems, DECIS interfaces with MINOS (Murtagh and Saunders 1983) or CPLEX (CPLEX Optimization 1989).

As we are considering long-term horizon, we set up a 4-stage piecewise linear model. We then use a two-stage decomposition approach to solve the problem and a pre-sampling strategy. Hence we sample a certain number of times from our VAR process and then proceed to solve this pre-sampled stochastic program). We do not address here the potential issue previously noted on the quality of the solution of the pre-sampled problem relative to the solution of the original problem. Further references on this issue can be found in the work by Infanger (Infanger 1999).

[6]In the general case, these questions of optimal timing of reallocation points are complicated from an analytical standpoint and we do not address them in this thesis, nor do we report any numerical trial runs on these issues.

[7]The General Algebraic Modeling System (GAMS) is specifically designed for modeling linear, nonlinear and mixed-integer optimization problems. Further references can be found at http://www.gams.com.

Number of Scenarios	GAMS+CPLEX	DECIS+GAMS+CPLEX	Time Reduction
40,000	574.8	234.6	59.2%
60,000	1,046.3	386.8	63.0%
80,000	2,013.0	472.9	76.5%
100,000	3,493.9	706.3	79.8%
120,000	6,555.5	1,659.6	74.7%

Table 5.3: Analysis of Computation Time (sec.)

5.3.5 Advantages of using DECIS and Computational Time

It is important to understand what advantages there are to using DECIS and a decomposition method in solving the large-scale LP that is the result of our pre-sampling strategy. For a general discussion of the advantages of decomposition techniques, we refer the reader to the treatment by Ruszczyński and Shapiro (Ruszczyński and Shapiro 2003). There are two issues at hand here: (i) memory requirements (that is directly related to the scale of the LP we're trying to solve) and (ii) computational time.

Of the first issue, i.e. memory usage, it is worth noticing that in a decomposition approach, the "basis" is decoupled. Consequently, the memory requirements for any basis factorization are reduced. The approach is similar to an iterative process for solving linear equations such as a block Gauss-Siedel, as opposed to a direct method (Strang 1988). For some problems, the memory requirements to solve the problem directly may exceed the memory available. By decomposing the larger linear program into smaller subproblems, the subsequent memory needed in the optimization process is reduced.

On the second issue of computational time, it can be noted that if we focus here on a 3-year time horizon, there is little need to worry about computational time as long as the solution time is not more than a few hours. However, exactly the same programming framework may be applied to daily data for much shorter horizons (e.g. three days instead of three years). In this case, computational time is important. Also, it is interesting in its own sake to compare computational times with or without using DECIS. If we assume that computational time is more or less linear with the number of elementary operations performed in the optimization process, then computational time can be used as a first-order proxy for DECIS contribution to reducing the problem complexity. Table 5.3 summarizes the computational time results.

Model (Sample size in yrs)	Stock	Bond	Cash
VAR (20)	18.56%	8.41%	1.29%
GBM (20)	14.16%	8.74%	3.11%
VAR (30)	15.52%	8.37%	1.25%
GBM (30)	12.93%	8.44%	3.30%
VAR (40)	14.29%	7.53%	0.99%
GBM (40)	13.46%	8.03%	1.92%
VAR (50)	16.36%	6.77%	0.86%
GBM (50)	15.92%	7.69%	1.17%

Table 5.4: Mean absolute deviations for 1-yr ahead forecasted vs. realized returns. The historical sample size (in years) is indicated in parentheses.

5.4 Results

5.4.1 Out-of-Sample Forecasting

We compare the out-of-sample forecasts we obtain by using VAR or GBM modeling of returns. Table 5.4 shows forecasts for both GBM and VAR models of returns[8] in the case of yearly data.[9]

These results show that the VAR forecasts are slightly better than GBM forecasts for the bond returns and significantly better for cash returns. However, the one-year ahead forecasts are better with GBM model for stock returns.[10] Also, these results show empirically that there is a historical window size that minimizes the mean absolute deviations between the one-year ahead forecasts and the realized returns.

5.4.2 Results Variations with respect to Final Goal

Our model's end result is the optimal recommendation for wealth allocation between asset classes at the time the model is run. We first show results in the case of a multivariate GBM fitting of the full sample of yearly returns and the effects of varying the end goal value. Table 5.5 displays our results for a pre-sampled scenario tree comprised of 70x(60x50) = 210,000 scenarios. This produced a large-scale LP comprised of 3063 rows, 6187 columns and 21370 non-zeros elements. Table 5.5 shows how the optimal allocation varies with respect to the final goal.

[8]The VAR model is initialized with historical sample's last year returns.

[9]The same results can be made available for monthly data.

[10]This was to be expected as our VAR fitting is not statistically significant for stock returns and GBM model is better for this asset class.

Final Goal	Stock	Bond	Cash	Objective
105	82.5%	17.5%		33.9
110	70.3%	16.5%	13.2%	22.0
115	66.6%	14.1%	19.3%	7.5
120	71.9%	20.3%	7.8%	-11.1
125	79.4%	20.6%		-34.1
130	86.4%	13.6%		-61.2

Table 5.5: Allocation results for GBM model with varying goal.

As Table 5.5 shows, the objective value decreases as the goal is increased, as was expected.[11] Also we can notice that cash is not used if the final goal is either too low or too high. An interpretation for this result is to say that if we are significantly above the goal, we can take more risk without fearing to be penalized, and if we are significantly below the goal, we need to take on more risk to achieve the goal, which translates in both cases into avoiding cash and increasing the stock allocation. Another interpretation for these results is that if we are either significantly above or under the goal, the problem almost reduces to maximizing an expected value over a linear function (which is equivalent to dealing with a risk-neutral investor) and hence we would choose an asset mix that maximizes the expected value without concern for the increased risk. Hence, far from the goal, we should expect the stock allocation to be increased, as is the case with our results.

5.4.3 Comparisons between GBM and VAR results

Table 5.6 and Table 5.8 show the variations of the results with respect to whether we model asset returns under a GBM framework or a VAR model with different initial conditions ("IC").[12] The initial conditions represent the state of the system at the time the optimization is conducted. So, initial conditions are the last returns available from the data sample at the time of optimization. However, for research purposes and for the sake of stability, it can be argued that initial conditions should be chosen as the average returns of the historical sample used in the analysis.

Tables 5.7 and 5.9 detail the values of the initial conditions used for our analysis. In Table 5.7, the first set of initial conditions (IC1) corresponds to each asset class initial lagged value set equal to the historical sample's mean return. The following initial conditions listed

[11]These results were obtained for an initial wealth of 100.

[12]This set of results was done on monthly data from a historical window composed of the last 120 months available from the sample.

Model	Stock	Bond	Cash	Objective
GBM	24.6%	49.3%	26.1%	-6.9
IC1	31.7%	52.4%	15.9%	-6.8
IC2	30.1%	17.1%	52.8%	-7.6
IC3	28.6%	71.4%		-5.9
IC4	35.3%	62.9%	1.8%	-6.1
IC5	25.5%	32.2%	42.3%	-7.6
IC6	28.7%	51.0%	20.3%	-3.7
IC7	24.4%	32.7%	42.8%	-9.6

Table 5.6: Allocation results when VAR initial conditions are centered around sample averages.

Initial Conditions	Stock	Bond	Cash
IC1	0.96%	0.54%	0.34%
IC2	5.06%	0.54%	0.34%
IC3	-3.14%	0.54%	0.34%
IC4	0.96%	2.63%	0.34%
IC5	0.96%	-1.55%	0.34%
IC6	0.96%	0.54%	0.47%
IC7	0.96%	0.54%	0.21%

Table 5.7: Values of Initial Conditions

Final Goal	Stock	Bond	Cash	Objective Value
GBM	24.6%	49.3%	26.1%	- 6.9
IC8	23.2%		76.8%	-12.1
IC9	26.8%		73.2%	-11.9
IC10	19.4%	14.8%	65.8%	-12.1

Table 5.8: Allocation results for VAR initial conditions centered around sample's last returns.

Initial Conditions	Stock	Bond	Cash
IC8	5.21%	1.03%	0.08%
IC8	9.31%	1.03%	0.08%
IC10	1.11%	1.03%	0.08%

Table 5.9: Values for Initial Conditions

(IC2 to IC7) correspond to the same values with the exception that each asset class initial return is varied (one after another) by +/- one standard deviation.

In Table 5.9, the first set of initial conditions (IC8) corresponds to each asset class initial lagged value set equal to the returns of the historical sample's last period. IC9 and IC10 respectively correspond to increasing the stock returns initial conditions by +/- one standard deviation.

We can see from these results that there is a much greater difference in optimal allocation results than in optimal values of the approximated stochastic programs. For instance, Table 5.8 shows that the difference between GBM and VAR IC1 optimal values is much less than 1.5% whereas the stock allocations vary by more than 22%. This is in line with the observation by Dupačová (Dupačová 1999) that:

> . . . in general, it is much easier to estimate the precision of the obtained optimal value than of optimal solutions.

We can also observe that, if the sum of the allocations to bond and cash asset classes is relatively stable, the allocation between the two asset classes is not and varies greatly. For instance by looking at the VAR results for IC1, IC2 and IC3, the sum of bond and cash allocations is between 68 to 72%. However, it is almost as if the bond and cash allocations respectively at 52.4% and 15.9% in the case of IC1 were swapped for IC2, for which they respectively become 17.1% and 52.8%.

The results also suggest that by using initial conditions for the VAR simulations that are exactly or close to the historical sample averages, we find less of a discrepancy between the GBM and VAR cases, with optimal values much closer than when using initial conditions equal to the sample's last year returns. In all cases, the distribution between bond and cash is highly sensitive to the initial conditions used for simulating the VAR process.

5.4.4 Analysis of Nonconvexity

An important observation in our analysis is the fact that the allocation results are not convex with respect to the initial conditions.

For instance, IC1 represents initial conditions equal to average values of the historical sample returns. We could expect the optimal allocation produced by this set of initial values to be somewhat between the optimal allocations produced by the sets of initial conditions IC2 and IC3, as IC2 and IC3 are only different from IC1 in the value of the initial stock lag (set at +/- one standard deviation from the mean). But we notice that it is not the case for the stock allocation as IC1 produces a recommendation of 31.7%, which is outside the [28.6%-30.1%] range given by IC3-IC2.

	Stock	Bond
Returns in Good Economy	1.20%	1.13%
Returns in Bad Economy	1.06%	1.07%
Optimal Allocation	28.6 %	71.4 %

Table 5.10: First Problem

	Stock	Bond
Returns in Good Economy	1.10%	1.12%
Returns in Bad Economy	0.96%	1.06%
Optimal Allocation	0%	100%

Table 5.11: Second Problem

This somewhat counterintuitive result can be better understood if we consider, without loss of generality, a simpler example for a one period model with two asset classes.

Simple Example Let's consider the following one-period problem. We have two asset classes (e.g. stock and bond) and a two-piece linear concave utility function as previously described with a goal of 115, a slope of 1 above the goal and a slope of 10 below the goal. The objective is to compute the optimal allocation starting with an initial wealth of 100 that maximizes the expected value.

We look at three different situations corresponding to three different sets of values for the asset class returns. The first problem described by Table 5.4.4 yields an optimal objective value of -41.13. The second problem described by Table 5.4.4 has lower returns (with the same returns spread in a good state of the economy vs. in a bad state of the economy) and yields an optimal objective value of -60. The third problem (Table 5.4.4) has returns that are a convex combination of the two preceding ones (P2 + 0.75 (P3 - P2)) and has an optimal objective value of -48.95.

In such a simple setting we could expect to get an optimal allocation for P3 that is somewhat between those of P1 and P2. The returns of P3 are constructed this way and

	Stock	Bond
Returns in Good Economy	1.18%	1.128%
Returns in Bad Economy	1.04%	1.068%
Optimal Allocation	47.4%	52.6%

Table 5.12: Third Problem

in fact the objective values seem to verify this convexity property (i.e. P3's optimal value -48.95 is between P1's optimal value of -41.13 and P2's optimal value of -60). However, we clearly see in this example that P3's optimal stock allocation of 47.37% is clearly outside the range of P2-P1, which is 0%-28.57%. Notice that the variations of the returns of both stock and bond asset classes in this simple setting can be easily obtained from a VAR Lag 1 framework.

5.5 Conclusion

We used a multi-stage stochastic programming framework for analyzing the effect of modeling serial correlation by means of a vector autoregressive process on the computed optimal asset allocations. We have shown that for certain asset classes considered, i.e. the bond and cash asset classes in our setting, a VAR model improves the out-of-sample forecasting of returns and is more appropriate than a GBM fitting (the out-of-sample forecasting being roughly equivalent for the GBM and VAR models of stock returns). Hence using a VAR model for the bond and cash returns and a GBM model for stock returns seems the most appropriate combination for improving the accuracy of forecasts.

Our results show that the results of the VAR model can be very sensitive to the initial conditions used. This effect is particularly significant on the allocation itself (especially between bond and cash), rather than the objective value. As our convexity analysis shows, there is an inherent instability of the allocation results when modeling returns via a vector autoregressive process because of its dependence on initial conditions. As the allocation results are not necessarily convex with respect to the initial conditions, it makes it more difficult for us to find simple confidence intervals for our results by using convexity arguments.

Chapter 6

Bayesian Stochastic Programming

It is important to bear in mind that the value of using stochastic programming for asset allocation purposes is very much tied to our ability to forecast future returns and the consistency of our future scenarios with what we know of economic realities. This is true of any stochastic program in general: finding the appropriate means for modeling uncertainty is often critical and drives the results. The results are useful only insofar as the parameters of the stochastic program have been correctly estimated.

When using stochastic programming for financial optimization applications, an interesting question is to what extent our modeling of the stochastic returns is going to change as we evolve within our programming horizon. We can take different points of view on this issue. A very simplistic one would be to assume that returns are stationary and that we would not want update our beliefs about asset returns as we progress into the planning horizon. That is, we can take the point of view that all our estimates should be based upon the historical window of returns we had available at the origin (if this is how we have estimated the processes describing the returns). This is shown in Figure 6.1. Another approach is to say that, as we progress into the planning horizon, we will update the parameters based upon the additional realized returns we observe. So for instance in this approach, at time t_1, we will update the stochastic processes based upon an expanded historical window that includes the original historical window and the additional returns observed in the first period. This is shown in Figure 6.2. Another modeling strategy is to use a ”rolling historical window”. With this methodology, the parameters of the stochastic process are reevaluated using a new historical window that has been “rolled forward”. This is shown in Figure 6.3. This approach is often favored by financial modelers who view old information as irrelevant. Often, the latest historical sample will be cut to reflect a belief of “local” or “recent stability” and a distribution of financial returns will be estimated based upon this assumption of

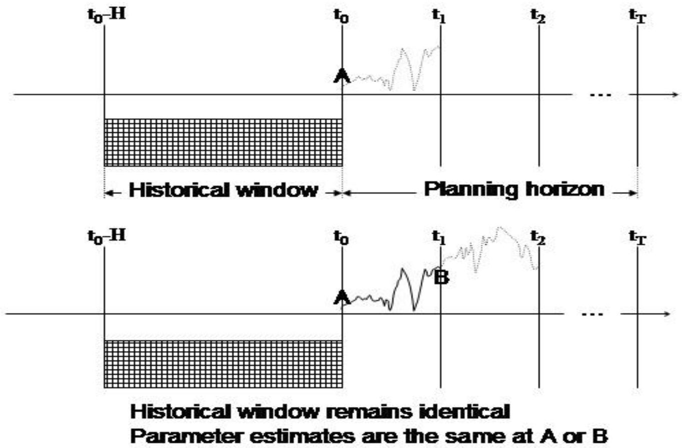


Figure 6.1: Frequentist Approach: Fixed Historical Window.

"local stationarity".

The three previous methods reflect a "frequentist" approach of the world whereby we assume that, over a certain period of time, returns can be treated as if they were drawn from a distribution with "true" parameters and we only choose different historical periods to try to estimate these parameters.

Another radically different approach, called Bayesian, questions this assumption. The Bayesian approach considers the underlying parameters used to model the stochastic processes as random themselves. The modeler has certain prior beliefs on the random description of these parameters and updates these beliefs, to obtain posterior beliefs, as new information is made available. This is shown in 6.4.

Hence, as our stochastic program unfolds over the planning horizon, we would want to reflect this behavior and have a consistent representation of our future actions. The differences between these approaches may be minimal if the original historical window is of significant size relative to the planning horizon. For instance, suppose we are dealing with a model of monthly returns for which we have twenty years of historical data and assume we are only considering a planning horizon three months ahead (with reallocations one- and two-month ahead). We would expect that, whatever method used, the updates made after the first month, at t_1, and after the second month, at t_2, will be very minor. However,

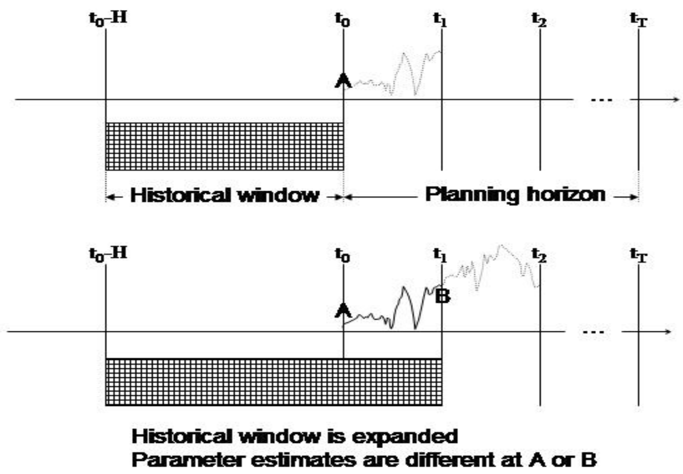


Figure 6.2: Frequentist Approach: Expanding Historical Window.

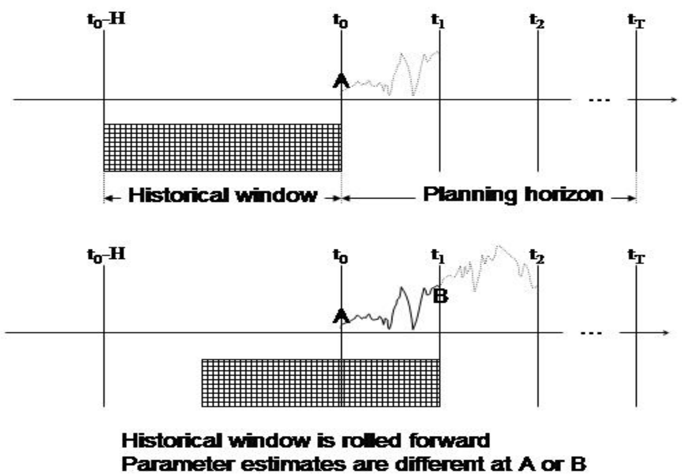


Figure 6.3: Frequentist Approach: Rolling Historical Window.

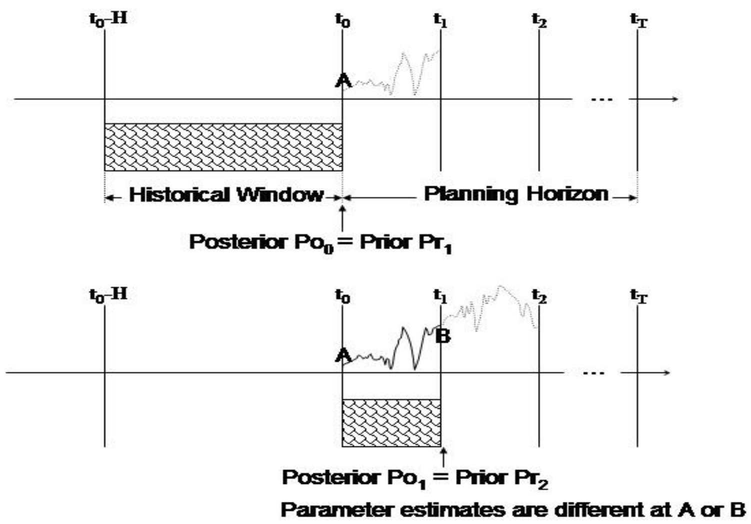


Figure 6.4: Bayesian Approach: Priors are constantly updated at every stage by taking into account the realized returns just observed in the last stage.

in the situation where we plan for a long-term horizon, using limited historical data, the results produced by these methods may differ significantly.

This chapter investigates whether using a Bayesian VAR or a simple VAR approach matters for our multistage program previously introduced. It also shows the particular usefulness of the Bayesian approach in a specific application, computing the allocation for a fund of funds. The chapter summarizes the important concepts of Bayesian analysis and the different priors traditionally used. It then compares the allocation results to those of the previous chapter. Last, it presents a financial application, i.e. computing an optimal fund allocation for a fund of funds.

6.1 The Bayesian Approach

In contrast to classical statistics that assume the existence of true parameters θ, Bayesian statistics regard θ itself as a random variable.

For instance, we may assume that a given set of observations $\mathbf{y} = (y_1, y_2, ..., y_T)'$ are drawn from a Gaussian distribution with parameters $\theta = (\mu, \sigma)'$. We would then compute our estimator $\hat{\theta}$ based upon the maximum likelihood principle. Royall provides a complete treatment of the likelihood paradigm (Royall 1997). The estimator $\hat{\theta}$ would be found as maximizing the following expression:

$$f(\mathbf{y};\theta) = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}}\, exp\left[\frac{-(y_t-\mu)^2}{2\sigma^2}\right]. \tag{6.1}$$

In contrast to this classical approach, the Bayesian view as described by Hamilton is (Hamilton 1994):

> ...θ itself is regarded as a random variable. All inference about θ takes the form of statements of probability... The view is that the analyst will always have some uncertainty about θ, and the goal of statistical analysis is to describe the uncertainty in terms of a probability distribution.

The sample likelihood 6.1 is viewed as the density of $\mathbf{y}$ conditional on the value of the random variable θ, denoted $f(\mathbf{y}|\theta)$. The prior density that represents our a priori beliefs on θ is given by $f(\theta)$. And the joint density of $\mathbf{y}$ and θ is given by:

$$f(\mathbf{y},\theta) = f(\mathbf{y}|\theta)\, f(\theta). \tag{6.2}$$

Once the data $\mathbf{y}$ has been observed, we update our prior beliefs on θ by computing θ's posterior density:

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y};\theta)}{f(\mathbf{y})}. \tag{6.3}$$

Since $f(\mathbf{y}) = \int_{-\infty}^{\infty} f(\mathbf{y},\theta)\, d\theta$ and by way of equation 6.2, we get:

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\, f(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{y},\theta)\, d\theta}, \tag{6.4}$$

which is Bayes' law.

6.2 Bayesian Vector Autoregression

VAR models of higher order have limitations that are due to (i) the large number of coefficients that need to be estimated as well as (ii) high correlation in the variables that leads to degraded precision in the parameter estimates. On (i), for n variables with p lags, the total number of coefficients that need to be estimated is $n \times (n \times p + 1)$. So for 5 variables with 2 lags, we need to estimate 55 coefficients. On (ii), we need to bear in mind that, according to the model, the independent variables $y_{i,t}$ are regressed against their lagged values $y_{i,t-1}, y_{i,t-2}, ..., y_{i,t-p}$, and that this structure tends to produce high correlations that lead to degraded precision in the parameter estimates. Hence large samples of observations involving time series variables that cover many years are needed to estimate the VAR model, and these are not always available. Hence, even with a lag p moderately large, the coefficient values, estimated by unrestricted OLS[1] methods, are often not very well determined in a finite set of data. In particular, Litterman shows that this problem is accentuated when dealing with economic series that exhibit trends or persistent local levels (Litterman 1980, Litterman 1986). To address these issues, Litterman, Doan and Sims suggest an alternative method for estimating the coefficients in these cases that rests on the use of Bayesian prior information (Doan, Litterman, and Sims 1984).

As Robertson and Tallman put it (Robertson and Tallman 1999):

> The idea is to treat the coefficients as random quantities around given mean values, with the tightness of the distributions about these prior means determined via a set of hyperparameters. The OLS coefficient estimator is then modified to incorporate the inexact prior information contained in these distributions. The main technical issues involve specifying the form of the prior distributions and determining the form of the estimators.

6.2.1 Minnesota Prior

Models of prior information distinguish between coefficients that should a priori be significant and coefficients that should not (and a priori should be zero). Each coefficient has a prior mean and variance. The Minnesota prior is one such model where the coefficients associated with the first-lagged dependent variables in each equation of the VAR are given a prior mean of 1. All other coefficients are assigned a prior mean of 0. So in equation i, the Minnesota prior takes the form:

$$\beta_{ii1} \sim N(1, \sigma^2_{ii1}) \quad for \quad k = 1, \tag{6.5}$$

$$\beta_{ijk} \sim N(0, \sigma^2_{ijk}) \quad for \quad i \neq j \quad or \quad k > 1. \tag{6.6}$$

To deal with the fact that a VAR model contains a large number of parameters, Doan, Litterman and Sims suggested using a few hyperparameters to generate a formula that would specify the standard deviation of the prior imposed on variable j in equation i at lag

[1]OLS: Ordinary Least Squares.

k (Doan, Litterman, and Sims 1984), namely:

$$\sigma_{ijk} = \theta\,\omega(i,j)\,k^{-\phi}\,\frac{\hat{\sigma}_{uj}}{\hat{\sigma}_{ui}} \tag{6.7}$$

There are three hyperparameters mentioned in equation (6.7) and a scaling factor. θ, the first hyperparameter, is the "overall tightness", reflecting the standard deviation of the prior on the first lag of the dependent variable. The second hyperparameter, the weight function $\omega(i,j)$, specifies the tightness of the prior for the variable j in equation i relative to the tightness of the own lags of variable i in equation i. Notice that this matrix of weights is assumed independent of the lag. The third hyperparameter, the term $k^{-\phi}$ is called a lag decay function with $0 \le \phi \le 1$ reflecting the decay rate. Finally, the ratio $\frac{\hat{\sigma}_{uj}}{\hat{\sigma}_{ui}}$ is a scaling factor that adjusts for varying magnitudes of the variables across equations i and j: $\hat{\sigma}_{ui}$ is the estimated standard error from a univariate autoregression involving variable i and similarly for $\hat{\sigma}_{uj}$ with respect to equation j.

A typical weighting matrix would be:

$$\mathbf{W} = \begin{pmatrix} 1 & 0.5 & \cdots & 0.5 \\ 0.5 & 1 & & 0.5 \\ \vdots & & \ddots & \vdots \\ 0.5 & 0.5 & \cdots & 1 \end{pmatrix}. \tag{6.8}$$

In each equation, this weighting matrix imposes the prior mean of zero for coefficients on other variables more tightly than it imposes the prior mean of 1 for the first lag of each dependent variable.

6.2.2 Variants of the Minnesota prior

Variants of the Minnesota prior usually try to alter the fact that a standard Minnesota prior will treat all variables in the VAR model in the same way, except the lagged dependent variables. A possible and natural modification is to alter the weight matrix and hence the prior variances. For instance, LeSage and Pan have constructed a weight matrix based on spatial contiguity to emphasize variables from neighboring states in a multi-state agricultural output forecasting model (LeSage and Pan 1995). Also LeSage and Magura employed inter-industry input-output weights to place more emphasis on related industries in a multi-industry employment forecasting model (LeSage and Magura 1991).

The general concept at work here is that by leaving the prior means as they are in the Minnesota formulation, but by adjusting the standard deviations of the prior, we can place more or less emphasis on the sample data versus the prior itself. By increasing the prior variances, the sample data will have a greater weight in determining the posterior means. Conversely, by tightening the prior variances, it will be harder for the sample data to shift the posterior means away from the prior means.

6.2.3 Random-Walk Averaging Prior

Another more recent approach to altering the equal treatment character of the Minnesota prior is a "random-walk averaging prior" suggested by LeSage and Krivelyova (LeSage 1999). As noted, the Minnesota prior treats all variables in the VAR model (except the first lag of the dependent variable) in an identical fashion. The prior proposed by LeSage and Krivelyova (LeSage 1999) involves both prior means and variances motivated by the distinction between important and unimportant variables in each equation of the VAR model. It is a generalization of the Minnesota prior.

In this setting, a weight matrix $\mathbf{W_0}$ is set up that is supposed to reflect the important or unimportant variables in each equation. The weight matrix contains values of unity in positions associated with important variables in each equation of the VAR model and values of 0 for unimportant variables. For example, in the example of matrix $\mathbf{W_0}$, the important variables in the third equation are variables 2 and 4. Notice that in this example, only variable 4 is considered to have an important autoregressive influence.:

$$\mathbf{W_0} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix} \tag{6.9}$$

The matrix $\mathbf{W_0}$ is then normalized to produce the matrix $\mathbf{C_0}$:

$$\mathbf{W_0} = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.33 & 0 & 0.33 & 0.33 \\ 0 & 0.5 & 0 & 0.5 \\ 0.33 & 0.33 & 0 & 0.33 \end{pmatrix} \tag{6.10}$$

This yields the following equation i in the VAR model:

$$y_{it} = \alpha_i + \sum_{j=1}^{n} c_{ij} y_{j,t-1} + u_{it} \tag{6.11}$$

and, in the case of our example, the VAR system of equations:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \begin{bmatrix} 0.5y_{2t} + 0.5y_{4t} \\ 0.33y_{1,t-1} + 0.33y_{3,t-1} + 0.33y_{4,t-1} \\ 0.5y_{2t} + 0.5y_{4t} \\ 0.33y_{1,t-1} + 0.33y_{2,t-1} + 0.33y_{4,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}. \tag{6.12}$$

In other words, this suggests having a prior mean for the coefficients on variables associated with the first lags of important variables equal to $\frac{1}{c_i}$, where c_i is the number of important variables in each equation i of the model. So, in our example of equation (6.12), the prior means associated with the lagged important variables $y_{2,t-1}$ and $y_{4,t-1}$ in the first VAR equation are 0.5. The prior means for the unimportant variables y_{1t} and y_{3t} in the same equation are 0.

This prior formulation allows us to downweight the lagged dependent variable using a zero prior mean to discount the autoregressive influence of past values of this variable and thus is far less restricted than the Minnesota prior. The Minnesota prior can be seen as the particular case of a simple random-walk $y_{it} = \alpha_i + y_{i,t-1} + u_{it}$, where the intercept term reflects the drift and is estimated using a diffuse prior. The random-walk averaging prior is centered on a random-walk model that averages over important variables in each equation of the model and allows for drift as well. As in the case of the Minnesota prior, the drift parameters α_i are estimated using a diffuse prior. Also consistent with the Minnesota prior, this generalization uses zero as a prior mean for coefficients on all lags other than first lags. It is also important to note that all time series used in this model need to be scaled or transformed to have similar magnitudes. However, this is not an issue when time series data can be expressed as percentage changes, as is the case of most financial applications that focus on returns.

The prior variances in the random-walk averaging prior can vary but respect the same guiding principles as the Minnesota prior. The prior variances differ according to whether the coefficients considered are associated with important or unimportant variables. LeSage (LeSage 1999) states the following guidelines for the prior variances:

- Parameters associated with unimportant variables should be assigned a smaller prior variance, so the zero prior means are imposed more 'tightly' or with more certainty.

- First lags of important variables are given a small prior variance, so the prior means force averaging over the first lags of important variables.

- Parameters associated with unimportant variables at lags greater than one will be given a prior variance that becomes smaller as the lag length increases to reflect the belief that influence decays with time.

- Parameters associated with lags other than first lags of important variables will have a larger prior variance, so the prior means of zero are imposed 'loosely'. This is motivated by the fact that we do not really have a great deal of confidence in the zero prior mean specification for longer lags of important variables. We think they should exert some influence, making the prior mean of zero somewhat inappropriate. As for unimportant variables, lag decay is still imposed on longer lags of important variables by decreasing prior variance with increasing lag length.

As for the Minnesota prior, LeSage reiterates the two main reasons why prior means for important variables at lags greater than one are set at zero (LeSage 1999):

> First, it is difficult to specify a reasonable alternative prior mean for these variables that would have universal applicability in a large number of VAR model applications... The second motivation for relying on inappropriate zero prior means for longer lags of the important variables is that overparametrization and collinearity problems that plague the VAR model are best overcome by relying on a parsimonious representation. Zero prior means for the majority of the large number of coefficients in the VAR model are consistent with this goal of parsimony and have been demonstrated to produce improved forecast accuracy in a wide variety of applications.

6.2.4 Estimating the Coefficients of a BVAR model

Consider the regression model:

$$y_t = \mathbf{x_t'}\,\beta + u_t, \tag{6.13}$$

where $u_t \sim i.i.d.\,N(0, \sigma^2)$, and $\mathbf{x_t}$ is a $(k \times 1)$ vector of exogenous explanatory variables and β is a $(k \times 1)$ vector of coefficients. Assuming we have T observations in our sample of observations, let:

$$\mathbf{y}_{(T\times 1)} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \qquad \mathbf{X}_{(T\times k)} = \begin{bmatrix} \mathbf{x_1'} \\ \mathbf{x_2'} \\ \vdots \\ \mathbf{x_T'} \end{bmatrix}.$$

If we treat β as random but σ^2 as known, we get the likelihood:

$$\begin{aligned} f(\mathbf{y}|\beta,\, \mathbf{X};\sigma^2) &= \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_t - \mathbf{x_t'}\,\beta)^2}{2\sigma^2}\right] \\ &= \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left[\frac{-(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2\sigma^2}\right]. \end{aligned} \tag{6.14}$$

If we assume that prior information is represented by $\beta \sim \mathcal{N}(m, \sigma^2\mathbf{M})$, where $\mathbf{M}$ is a matrix that scales our prior beliefs on how β is dispersed around our prior mean m, we can write[2]:

$$f(\beta;\sigma^2) = \frac{1}{(2\pi\sigma^2)^{k/2}}\, |\mathbf{M}|^{-1/2} \exp\left[\frac{-(\beta-\mathbf{m})'\mathbf{M}^{-1}(\beta-\mathbf{m})}{2\sigma^2}\right]. \tag{6.15}$$

The Minnesota prior (or other variants previously described) can be embedded in this framework of assuming $\beta \sim \mathcal{N}(m, \sigma^2\mathbf{M})$, provided an adjustment is made on the constant coefficient (which is traditionally assumed to have a diffuse prior) and a column of 1 is added to the $\mathbf{X}$ matrix of regressors. In each VAR equation, a prior mean for the constant coefficient can be estimated by performing a univariate regression first. In addition, a "loose" prior variance can be attributed to the constant coefficient via the matrix $\mathbf{M}$, by having artificially large values on both the first row and column (so that there is also little prior covariance restrictions between the constant coefficient and the other regression coefficients)[3].

As knowledge about the exogenous variables $\mathbf{X}$ is presumed to have no effect on the prior distribution, so that equation 6.15 also describes $f(\beta|\mathbf{X};\sigma^2)$. Using Bayes' law:

$$f(\beta|\mathbf{y},\mathbf{X};\sigma^2)\; f(\mathbf{y}|\mathbf{X};\sigma^2) = f(\mathbf{y}|\beta,\, \mathbf{X};\sigma^2)\; f(\beta|\mathbf{X};\sigma^2),$$

[2]We also assume $\mathbf{M}$ is invertible.
[3]We still need to make sure the matrix $\mathbf{M}$ stays invertible.

and using equations 6.14 and 6.15, this yields[4]:

$$f(\beta|\mathbf{y},\mathbf{X};\sigma^2) = \frac{1}{(2\pi\sigma^2)^{k/2}} \left|\mathbf{M}^{-1} + \mathbf{X}'\mathbf{X}\right|^{1/2} \exp\left[\frac{-(\beta - \mathbf{m}^*)'(\mathbf{M}^{-1} + \mathbf{X}'\mathbf{X})(\beta - \mathbf{m}^*)}{2\sigma^2}\right] \tag{6.16}$$

$$f(\mathbf{y}|\mathbf{X};\sigma^2) = \frac{1}{(2\pi\sigma^2)^{T/2}} \left|\mathbf{I}_T + \mathbf{XMX}'\right|^{-1/2} \exp\left[\frac{-(\mathbf{y} - \mathbf{Xm})'(\mathbf{XMX}')^{-1}(\mathbf{y} - \mathbf{Xm})}{2\sigma^2}\right] \tag{6.17}$$

$$\mathbf{m}^* = (\mathbf{M}^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{M}^{-1}\mathbf{m} + \mathbf{X}'\mathbf{y}). \tag{6.18}$$

This means that the conditional distribution of β given the observed data $\mathbf{y}$ is:

$$f(\beta|\mathbf{y},\mathbf{X};\sigma^2) \sim \mathcal{N}(\mathbf{m}^*,\, \sigma^2(\mathbf{M}^{-1} + \mathbf{X}'\mathbf{X})^{-1})$$

and the marginal distribution of $\mathbf{y}$ given $\mathbf{x}$ is:

$$f(\mathbf{y}|\mathbf{X};\sigma^2) \sim \mathcal{N}(\mathbf{Xm},\, \sigma^2(\mathbf{I}_T + \mathbf{XMX}')).$$

6.3 Comparison of BVAR vs. VAR forecasts

We perform the same analysis as in the previous chapter on monthly returns for the 10-year Treasury notes and for the 30-day Treasury bills. We vary the historical sample size from 12 months to 60 months and graph the improvement of the BVAR model forecast over the VAR forecast (in percent). These results reflect the average forecasting improvements of BVAR over VAR calculated over the full sample of returns. For instance, using a 12-month historical sample size, we averaged the forecasting results over 655[5] forecasts realized with a rolling historical sample. For each forecast, we compared the absolute deviation between the out-of-sample realized returns and the forecasted returns for both VAR and BVAR models. We then compared the percentage improvement of the BVAR forecast over the VAR forecast. If this improvement is negative, it simply means that the BVAR forecast did not perform as well as the VAR forecast. The BVAR forecast uses a standard Minnesota prior as noted before. Figures 6.5 and 6.6 show that overall the BVAR model seems to perform better over the VAR forecasts for small historical samples of 12 and 24 months.

[4]Detailed proof is in the work by Hamilton (Hamilton 1994).

[5]The full sample size is comprised of 696 monthly returns, the forecasting horizon is up to 30-month ahead and the historical sample size used is 12-month. Thus, by rolling forward our historical sample, we perform $696 - 12 - 30 + 1 = 655$ out-of-sample forecasts.

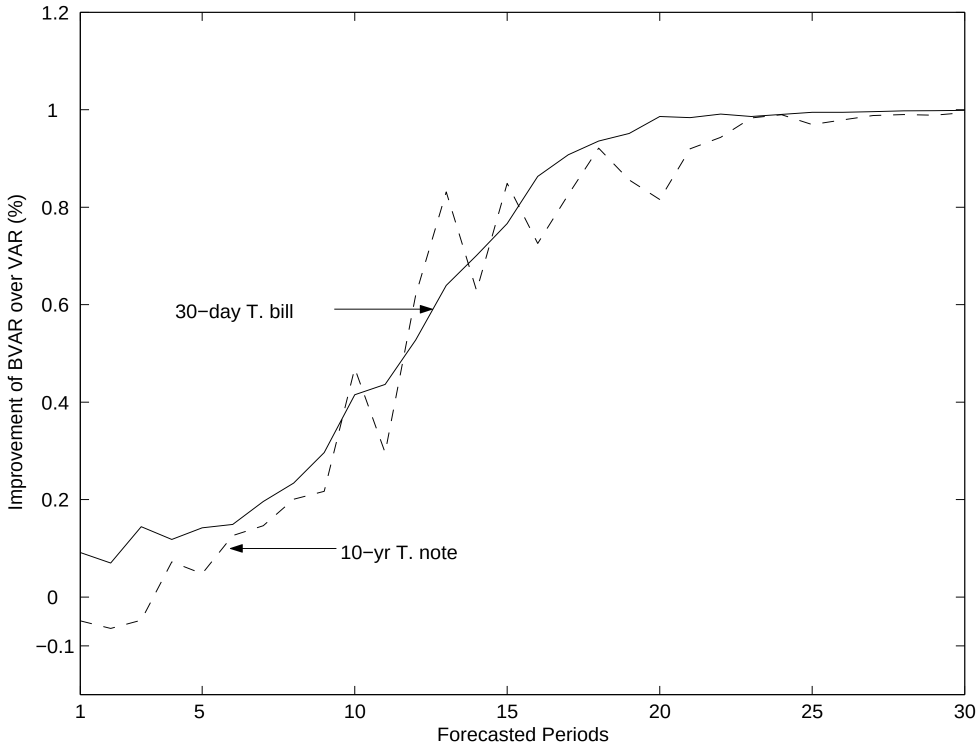


Figure 6.5: Comparing BVAR vs. VAR forecasts (12-month historical sample). Values of the Minnesota prior are set with a symmetric weight matrix containing off-diagonal values equal to 0.5, an overall tightness of 0.1 and an overall decay of 1. As the historical sample of 12-month is rolled forward within our full sample of 696 monthly returns, these results are averaged over 655 out-of-sample forecasts using a forecasting period of up to 30-month ahead.

Figures 6.7, 6.8 and 6.9 show that this improvement of BVAR over VAR disappears as the sample sizes are increased to 36, 48 and 60 months.

6.4 Application to Funds of Funds

A particular application of concern here is the problem of computing asset allocations for funds of funds. A fund of funds is a pool of money from investors in various financial markets that invests in “hedge funds”. The term “hedge fund” is misleading. At first, such funds were designed to provide real hedging services for banks or large financial institutions. Today, though the term ”hedge” remains (and some hedge funds’ core business is to develop hedging strategies for their clients), the term primarily designates investment funds with

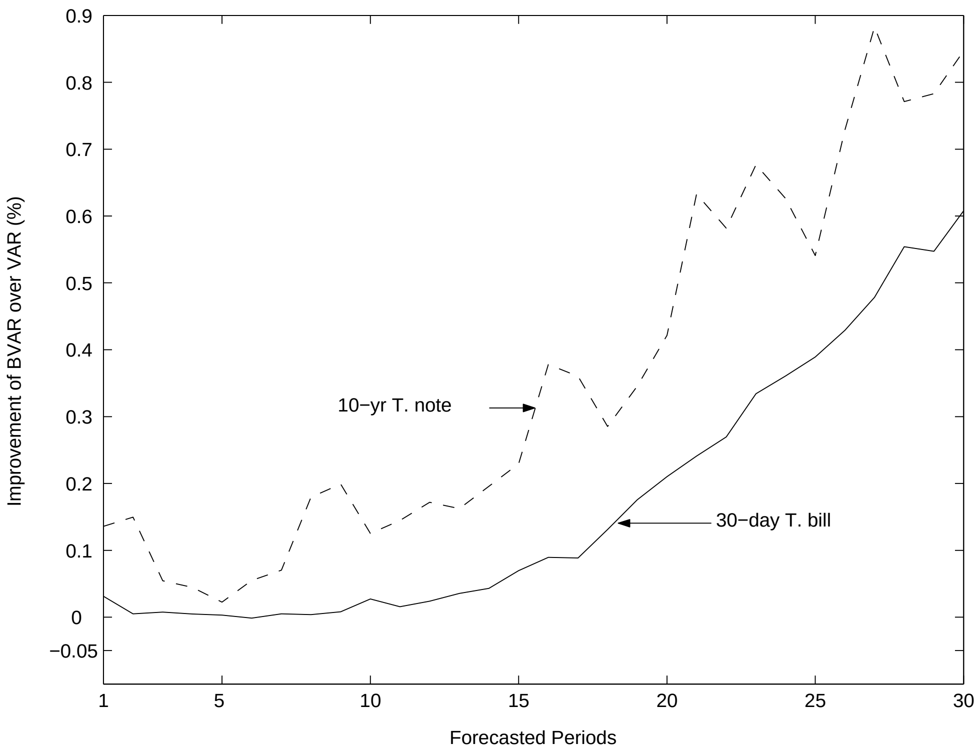


Figure 6.6: Comparing BVAR vs. VAR forecasts (24-month historical sample). Values of the Minnesota prior are set with a symmetric weight matrix containing off-diagonal values equal to 0.5, an overall tightness of 0.1 and an overall decay of 1. As the historical sample of 24-month is rolled forward within our full sample of 696 monthly returns, these results are averaged over 643 out-of-sample forecasts using a forecasting period of up to 30-month ahead.

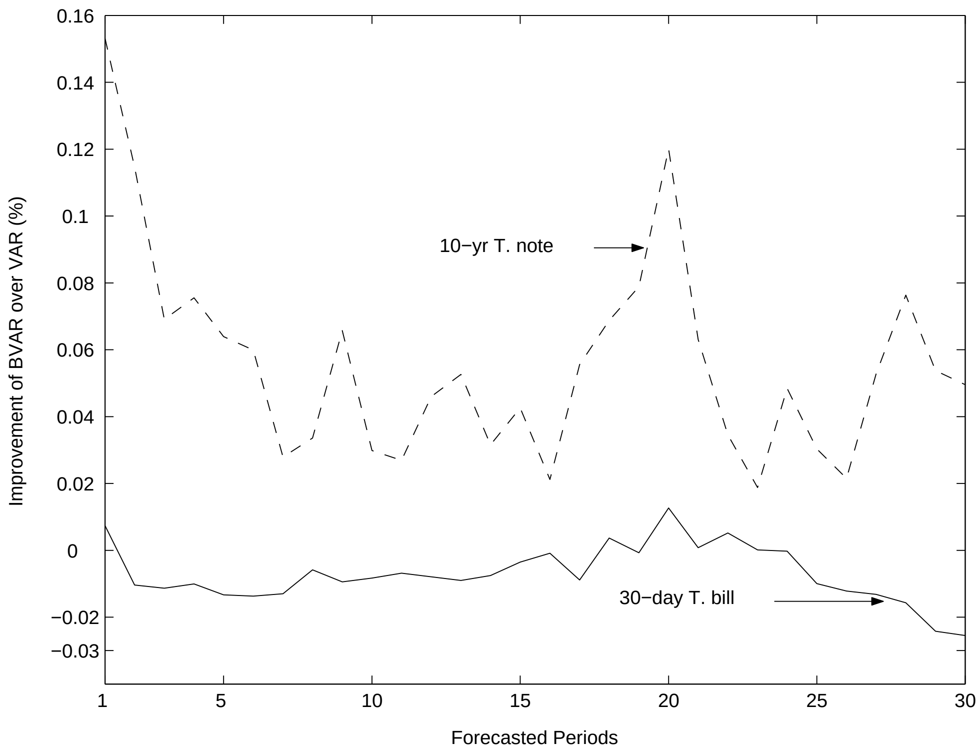


Figure 6.7: Comparing BVAR vs. VAR forecasts (36-month historical sample). Values of the Minnesota prior are set with a symmetric weight matrix containing off-diagonal values equal to 0.5, an overall tightness of 0.1 and an overall decay of 1. As the historical sample of 36-month is rolled forward within our full sample of 696 monthly returns, these results are averaged over 631 out-of-sample forecasts using a forecasting period of up to 30-month ahead.

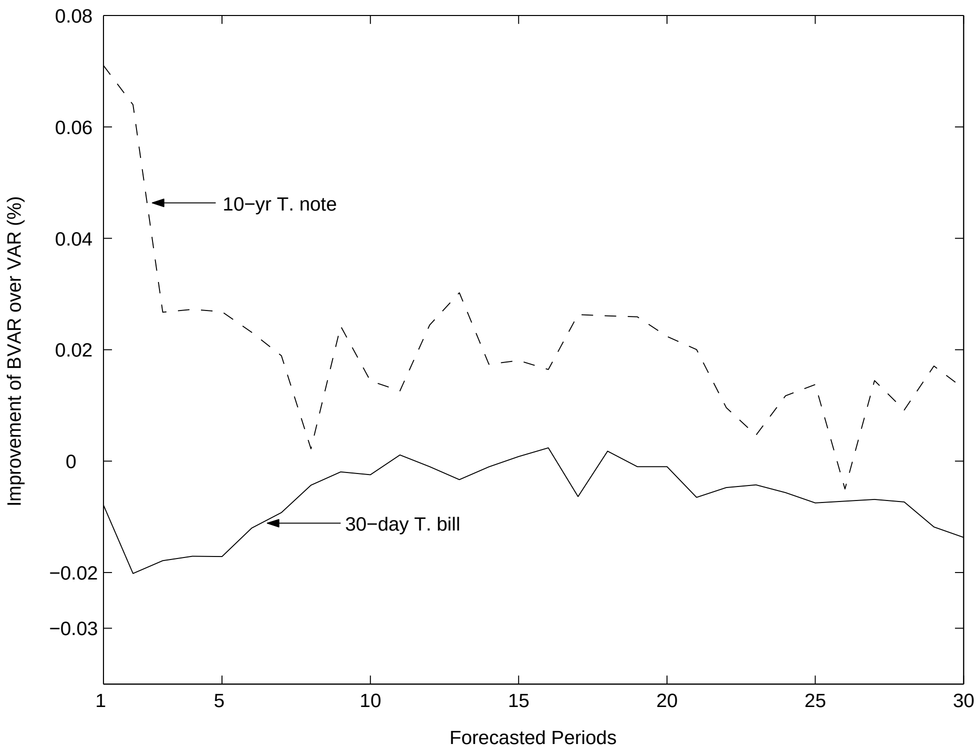


Figure 6.8: Comparing BVAR vs. VAR forecasts (48-month historical sample). Values of the Minnesota prior are set with a symmetric weight matrix containing off-diagonal values equal to 0.5, an overall tightness of 0.1 and an overall decay of 1. As the historical sample of 48-month is rolled forward within our full sample of 696 monthly returns, these results are averaged over 619 out-of-sample forecasts using a forecasting period of up to 30-month ahead.

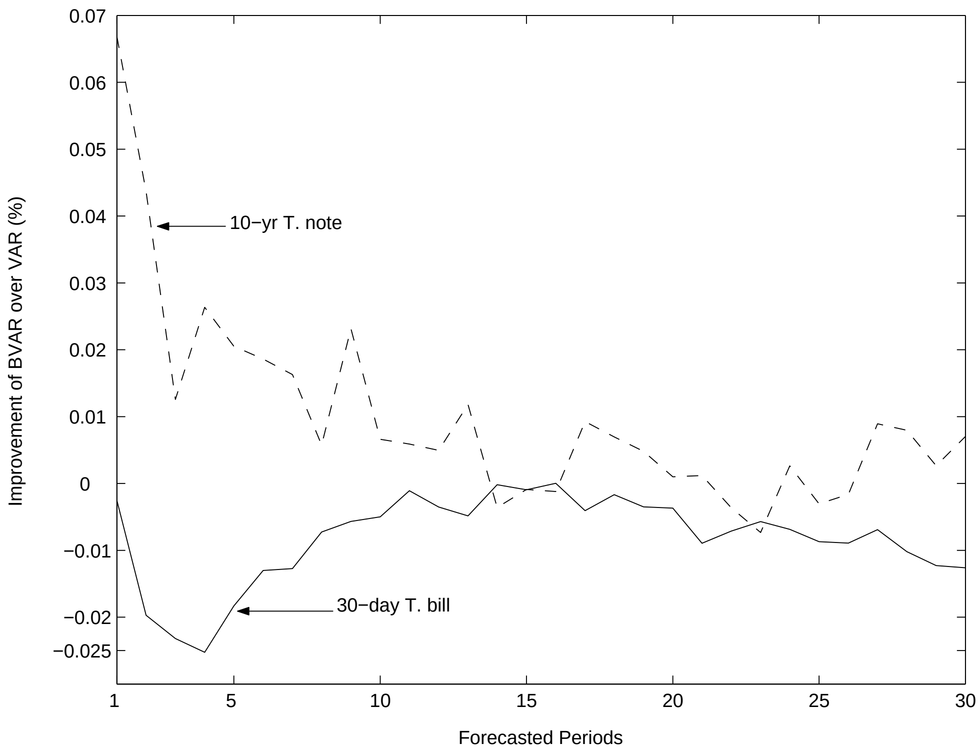


Figure 6.9: Comparing BVAR vs. VAR forecasts (60-month historical sample). Values of the Minnesota prior are set with a symmetric weight matrix containing off-diagonal values equal to 0.5, an overall tightness of 0.1 and an overall decay of 1. As the historical sample of 60-month is rolled forward within our full sample of 696 monthly returns, these results are averaged over 607 out-of-sample forecasts using a forecasting period of up to 30-month ahead.

a few accredited investors (whether institutions or wealthy private investors) that seek to maximize “absolute” returns.[6] The hedge fund industry has grown explosively since 1980. If by the late 1980s, the number of funds had increased to about 100, there were more than 1,200 hedge funds in 1997 with more than $200 billion[7] of assets under management. By May 2003, there were more than 6,000 hedge funds worldwide with more than $600 billion of assets under management (Ziemba 2003).

For an outsider, assessing the performance of hedge funds is difficult due to looser regulatory requirements and minimal disclosure policies. The manager of a fund of funds, who is wondering how to allocate resources to each hedge fund, is facing a difficult allocation problem. Forecasting the performance of each hedge fund is a difficult endeavor. The track record of each fund is often short (typically a few years of quarterly results). Reconstructing a hedge fund’s performance from its holdings is not easy as often hedge funds take positions that are small enough to escape filing requirements with the Securities and Exchange Commission. So, all in all, the allocation of the fund of funds may have to be made on a great deal of qualitative judgment rather than on a thorough quantitative assessment. Moreover, early work on hedge fund performances indicates there is little persistence in hedge fund performance (Ziemba 2003).

Our work suggests using our multistage stochastic program with a Bayesian VAR modeling of returns for this allocation decision. We argue for the following: firstly, investing in hedge funds may generate significant transaction costs that can be easily included in our model. Secondly, because of the shortage of significant performance time series, and the need for the manager of a fund of funds to assess different intangibles, using a BVAR model with adjusted prior assumptions seems the most appropriate tool available for forecasting future performance. A lot of funds should have correlated returns: they have the same type of trading strategy, they invest in the same markets, etc. So we can capture some of this knowledge of common influences in the prior’s covariance matrix. Lastly, if we are trying to optimize over an horizon of a few years (assume for instance three years), it may be very important that we preserve some Bayesian consistency in our modeling of the fund of funds reallocation behavior. Hence, for all the reasons previously noted, we would expect a BVAR model of hedge fund returns to be an appropriate forecasting tool.

[6] The term “absolute” returns is used to emphasize the fact that returns are maximized independently of any “benchmark” returns, as it traditionally the case for many mutual funds.

[7] Any reference to $ (”dollar”) is by default US $.

Chapter 7

Conclusion and Further Research

This thesis has addressed two research questions and provided a framework for an important and current financial problem, the problem of allocating funds between funds.

The first research question we have addressed is the issue of predictability and serial correlation of equity returns. This question has been for many years and continues to be a controversial subject among academics and professional financial analysts. Our contributions is to have proposed a novel methodology for detecting local bursts of serial correlation and, by analyzing historical data and conditional returns, to have shown new evidence of market predictability for equity indices over short-term horizons of a few days. Our results suggest both momentum and reversal effects for the variations of these broad-based equity indices. Further work could be done by translating our sampling scheme and performing, for instance, the same analysis with daily opening prices.

Our second area of research was analyzing the variations of optimal asset allocations, within a multi-stage stochastic programming framework, with respect to the statistical modeling of returns. Restricting ourselves to a multi-stage stochastic programming framework, we have compared the optimal asset allocations obtained by switching from a geometric Brownian motion ("GBM") model to a vector autoregressive ("VAR") model of asset class returns. In the process, we have showed clear evidence of serial correlation for Treasury bonds' and bills' returns. Moreover, in the VAR case, we have showed that the allocation results vary significantly depending on the initial conditions. For both the GBM and the VAR models, we have showed that the allocation results are very sensitive to the historical samples used for calibrating the models. We have also introduced a third statistical model of asset class returns, a Bayesian vector autoregressive ("BVAR") approach, and have shown empirically that this model provides better and more stable out-of-sample forecasts of returns, especially when the different statistical models are calibrated using small samples of historical data.

Our previous results suggests that the multi-stage stochastic programming framework developed here, when coupled with a BVAR modeling of asset class returns, is particularly appropriate for assisting allocation decisions with limited information and significant transaction costs, as is the case for funds of funds. Further work could be done involving out-of-sample comparisons of the performances of the different allocations derived from the different statistical models of asset class returns.

Appendix A

A Brief History of Mathematical Finance

This appendix provides a review of some historical developments of mathematical finance in general and the different models of financial returns. A second section briefly addresses some epistemological considerations.

A.1 A Brief History of Mathematical Finance

A.1.1 Introduction of Random Walk

Bachelier, in his doctoral thesis, was the first researcher to attempt to qualify the stochastic nature of stock market movements. One of the remarkable feats of Bachelier's work is that he is the first researcher to describe stock market returns as following a Gaussian distribution. As Bachelier states[1]:

> The determination of these fluctuations (of stock market prices[2]) depends on an infinite number of factors; it is, therefore to aspire to mathematical predictions of it. Contradictory opinions concerning these changes diverge so much that at the same instant buyers believe in a price increase and sellers in a price decrease.

Thus, Bachelier is the first to consider an infinite number of factors that summed up together (and without dwelling here on the technical conditions required) will produce a Gaussian distribution (as could have been derived from the Central Limit Theorem).

[1] This is from p. 17 of Cootner's translation listed in the bibliography along with Bachelier's original work (Bachelier 1900).

[2] This is our addition.

Another important point in Bachelier's thesis is the fact that he considers that there is an infinity of factors (of similar magnitude) that determine stock price movements. Implicit in this statement is that it is completely useless to attempt to predict those movements. In short and loosely speaking, stock price movements are the results of an infinite sum of small steps (random walk) and hence follow a Gaussian distribution and are unpredictable. In other words, Bachelier is the first to apply the concept of random walk in a financial context and to introduce Brownian motion to finance.

A.1.2 Concept of Market Efficiency

It is very important, we believe, to have a clear idea of this concept, as it is often used in a very murky way by both practicioners and academics alike. The idea of "market efficiency" is that the market prices fully reflect the available information on the underlying security and the economic reality of the entity they represent. As Walter states (Walter 2003):

> If the relevant information necessary for a transaction is correctly transmitted in the price, and considered as an exchange mechanism, the market is then called "efficient".

With this definition, we see a first aspect of market efficiency: financial markets and the prices they provide need to correctly reflect the prices of the underlying economic realities. But there is also a second aspect important to market efficiency (aspect which is not always alluded to in the murky definition of the concept), i.e. the role of financial markets as exchanges for allocating risks. In other words, markets are deemed "efficient" if different investors with different risk profiles can find the risk-returns tradeoffs most appropriate for them. This second aspect of efficiency deals with the allocational efficiency of the markets.

A.1.3 Empirical Testings

Subsequent to Bachelier's early work, there was no direct test of his Gaussian model of stock prices until after the Crash of 1929. This empirical work amounted essentially to statistical testing of financial data along two dimensions: (i) verifying that returns, if they really are unpredictable, are indeed time independent and that there is no serial correlation; (ii) characterizing the amplitudes of variations of prices and returns (if Bachelier's work was assumed to be true, this amounted to fitting a Gaussian distribution on existing data and estimating the mean and volatility[3] parameters).

[3] As by tradition, "volatility" is synonymous with "standard deviation".

Different tests of serial independence were carried out that concurred with Bachelier's assertion that markets are random. We mention here the studies by Cowles , Working, Cowles and Jones, and Kendall (Cowles 1933, Working 1934, Cowles and Jones 1937, Kendall 1953). The work by Osborne showed that geometric Brownian motion was more appropriate for the limiting distribution than Brownian motion (Osborne 1959). Building upon known properties of particle movements in statistical mechanics, Osborne inferred a Gaussian density in the first differences in the logarithms of quoted prices. The statistical investigations continued and as Walter states:

> One after another, following Roberts' suggestions, Larson, Working, Houthakker, Alexander, Cootner, Moore, Granger and Morgenstern bring strong support to the random walk thesis, and confirmed that successive price changes can be considered comparable in first approximation to a Brownian motion (Larson 1960, Working 1960, Houthaker 1961, Alexander 1961, Cootner 1962, Moore 1962, Granger and Morgenstern 1964).

A.1.4 Wedge between Academics and Practicioners

The statistical validation of random walk models led to an open conflict between academics who saw there the proof that markets were unpredictable and practicioners, such as chartists, whose very jobs seemed seriously threatened by the academic claims. As Alexander writes (Alexander 1961):

> There is a remarkable contradiction between the concepts of behavior of speculative market prices held by professional stock market analysts on the one hand and by academic statisticians and economists on the other. The professional analysts operate in the belief that there exist certain trend generating facts, knowable today, that will guide a speculator to profit if only he can read them correctly.

Partisans of the existence of trends believed that predictability was possible either because of "fundamental" reasons (by looking at underlying business basics and their forecasts) or because of "technical" reasons (they would chart the history of prices and believed the price trajectories followed certain patterns supposed to reflect general rules of the market).

A.1.5 Debate between Active and Passive Management

The Gaussian modelling of stock price movements had a major impact on the asset management industry. It raised the question of whether active management was really possible or whether it was some kind of charlatanism. Again, in a brief sweep at the issues, the reasoning was the following: (i) if markets are efficient, they reflect instantaneously a lot of exogenous and unpredictable information, hence their price movements are unpredictable; (ii) if prices are unpredictable, at least in direction, there is little reason to believe an active manager is really adding value doing stock picking.

There was a flurry of studies on the performance of portfolio managers to see if they were performing better than the market. This was a natural endeavor if one believed markets were unpredictable. We provide here a few references such as (in chronological order) the studies by Treynor, Sharpe and Jensen (Treynor 1965, Sharpe 1966, Jensen 1968, Jensen 1969), all written between 1965 and 1969.

What these studies underlined was that, net of management fees, active managers were on average doing no better than the market overall. It suddenly seemed that active management was more a matter of luck than real talent. Some argued that investors were just as well off investing in a general market index. This led several banks (e.g. Wells Fargo) to pioneer the development and marketing of index funds to their clients as early as 1971[4].

A.1.6 Debate between Economic Predictability and Random Returns

A major argument used by so-called "fundamentalists" is that the economic reality of a business is usually changing at a slow pace. It should therefore be reasonable to expect some predictability (whether medium-term or long-term) in the associated financial returns. For instance, suppose one values a stock using a discounted cash flow ("DCF") analysis of its future dividends. It is often the case that a firm's dividend policy is predictable a few months or years in advance, usually displaying some serial correlation. Similarly, usually the operational accounting metrics used display some serial correlation. Hence it is not a priori unreasonable that there could be serial correlation in the company valuation itself if we believe there should be a certain correspondence between a company's operational results and its stock price (all the more if we make the simplifying assumption that the company has no debt).

This debate is again to be understood within the concept of market efficiency and the translation of economic realities (e.g. the state of a company's business) into financial information and prices. As Fama states (Fama 1965):

> ... In essence, there is as yet no general model of price formation in the stock market which explains price levels and distributions of price changes in terms of

[4]Bernstein provides an account of this development (Bernstein 1992).

> behavior of more basic economic variables. Developing and testing such a model would contribute greatly toward establishing sound theoretical foundations in this area.

Almost forty years later, it seems the question remains open though different elements of answer have been proposed.

A.1.7 Reconciliation of Different Views

Fama was one of the first ones to reconcile the seemingly mutually exclusive views of having both efficient and predictable markets. One way for doing so is by considering market imperfections (e.g. transaction costs). By and large, Fama's view is that markets can be considered efficient and that there may be serial dependence in stock price returns but that this dependence cannot be exploited for excess gains over a standard buy-and-hold model of the market index. Fama asserts (Fama 1965):

> ... the independence assumption is an adequate description of reality as long as the actual degree of dependence in the series of price changes is not sufficient to allow the past history of the series to be used to predict the future in a way which makes expected profits greater than they would be under a naive buy-and-hold model.

A.1.8 Difficulties in Testing the Efficient Market Concept

Taxonomy of Information Sets This classification is due to Roberts (Roberts 1967) who distinguishes among:

- **Weak-Form Efficiency:** The information set includes only the history of prices or returns themselves.
- **Semistrong-Form Efficiency:** The information set includes all information known to all market participants (publicly available information).
- **Strong-Form Efficiency:** The information set includes all information known to any market participant (private information).

Testing a Joint Hypothesis Testing those concepts and different levels of efficiency is very difficult in essence. Let us assume first that we are in a financial system that provides equal public information to all participants[5]. Hence according to the classification above, we could test for the semistrong-form of efficiency. This situation corresponds intuitively to the orginal setting of the Bachelier-Osborne model[6] where the information shocks are homogeneous in nature or, loosely speaking, of the same order of magnitude[7]. This representation of the new information arrival translates into a Gaussian distribution for price changes. Testing for "efficiency" in this context boils down to testing for the Gaussian distribution of price changes. If this statistical test is rejected, it is then the whole aforementioned logic which is rejected but not necessarily the "efficiency" itself of the markets at hand. It could very well be, as Mandelbrot will argue (Mandelbrot 1962, Mandelbrot 1971), that the Bachelier-Osborne assumption of new information shocks of equal importance is wrong. Mandelbrot develops the case that shocks should be described by a "wild randomness" (which translates into Paretian-stable, fat-tailed distribution) rather than by a "mild randomness" (as was assumed until then in the Gaussian-efficiency model). From a scientific standpoint, we can only test the statistical inferences of these models. If a particular statistical distribution is rejected (whether Gaussian or fat-tailed), strictly speaking we can only reject our joint hypothesis. But we cannot decide between rejecting the "market efficiency" concept (which would translate into considering that markets do not fully and accurately reflect new information) or rejecting our underlying assumption on the arrivals and nature of new information shocks (which would translate into saying that new information does not emerge the way we thought it did).

A.1.9 The More Recent Developments

Lévy's stable distributions. Mandelbrot is unequivocal in his rejection of a Gaussian description of price changes and proposes using Lévy's stable distributions instead of the standard Gaussian distribution (Mandelbrot 1962). Mandelbrot further refined this idea convinced that financial markets were subject to fractures and not just smooth transitions (Mandelbrot 1997). Mandelbrot introduces the concept of "wild randomness" (or Paretian randomness) to "mild randomness" (or Gaussian randomness). As noted, Mandelbrot had proposed Paretian (or Lévy's) distributions to model stock prices as early as 1962 but the

[5]This is an unrealistic assumption. Even, if strictly speaking, this may be the case as information is freely available, behavioral finance models show that investors have limited attention spans and capacities to process information. Getting the information is not just the issue, heeding it, being able to process it and acting upon it is just as important.

[6]To the best of our knowledge, this joint-name attribution was first used by Fama (Fama 1965) and generally refers to the Gaussian model of stock price movements as originally introduced by Bachelier (Bachelier 1900) and further developed by Osborne (Osborne 1959).

[7]A condition that is required for the application of the Central Limit Theorem.

use of these distributions did not spread quickly within the financial community as they did not provide closed-form solutions to traditional financial problems (e.g. the determination of a mean-variance portfolio). However, it has become increasingly clear (all the more since the 1987 crash[8]) that financial returns are too fat-tailed to be lognormal. Quoting a recent study by Green and Figlewski on the subject (Green and Figlewski 1999), further referenced by Sornette (Sornette 2003):

> There are more realizations in the extreme tails (and the extreme values themselves are more extreme) than a lognormal distribution allows for. In other words, the standard valuation models are based on assumptions about the returns process that are not empirically supported for actual financial markets.

Development of Behavioral Finance. In the last score of years, significant studies have been done of the Stock Market that emphasize the significant role behavioral finance considerations can play. For instance, the so-called "herding" phenomenon has been documented that reflects the conservative attitude of most fund managers and the fact that they are often very averse to taking contrarian positions to their peers'. In other words, for pension managers concerned with preserving their jobs, there is more risk in being right by themselves than in being wrong with everybody else. Or in other words, if a pension manager is wrong with many of his peers, he/she will be less likely to be criticized. This creates a "herding" phenomenon among some money managers that can sometimes deeply influence the market in that suddenly a significant number of managers start selling or buying the same type of equity (a particular industrial sector, a certain size of companies, etc.). Further references can be found in the works by Hong et al. or Bikhchandani and Sharma (Hong, Kudik, and Solomon 2000, Bikhchandani and Sharma 2001). Generally speaking, a wide range of social dynamics can also account for stock price movements as documented by Shiller (Shiller 1984). All these endogenous phenomena underline the fact that stock price movements are not just the product of exogenous shocks but also of endogenous fluctuations of opinions, group pressures and other factors. Financial markets have never been historically immune to fads and fashions[9] and lately, the "Internet boom" of the late 1990s (and subsequent "Internet bust" of the early 2000s) shows how excessive the markets can be due to the emergence of new technologies and its corollary of massive diffusion of new and unchecked opinions.

[8] An interesting article to read on the subject is Business Week's article "The Efficient Market was a Good Idea - and then Came the Crash," dated February 22^{nd}, 1998, pp. 38-39.

[9] Further reference is needed on the story of Law.

Implications on ALM Models In our view, there are two venues of improvement for dynamic asset allocation models: (α) improving the statistical description of financial returns and (β) doing a better job of incorporating macroeconomic considerations to try to distinguish market fads from fundamental trends. On the first issue, the financial returns used in dynamic asset allocation models need to reflect the acknowledgement that financial returns may be better described by fat-tailed distributions and the emergence of fractal descriptions of financial markets (that have the remarkable property that they are immune to scale changes). In this case, the focus is likely to be on the accuracy of our numerical solutions. On the second issue, we need to take into account behavioral finance considerations, all the more if they allow us to detect the formation of irrational trends (e.g. market bubbles) to benefit in the short term and take directional positions in the markets (however controversial this issue is).

Outstanding Issues Both the inclusion of fat-tailed distributions or the use of behavioral finance considerations have their own drawbacks. The main weakness for dealing with fat-tailed distributions is that the extreme events that usually drive the calibrations of the tails of the distributions (certainly on the down side) are hard to predict and can vary greatly in nature. Sometimes the market swings wildly because of the apparition of great uncertainty that no expert is really able to assess. This is the case, we believe, in situations involving large geopolitical events (e.g. the terrorist attack of 9/11/01 on the World Trade Center). The far-reaching consequences of such an event are very hard to estimate immediately. It may be the case that in such extreme situation multi-agent behavioral finance models are the appropriate way to model the markets' response. It is likely that, in such a situation of uncertainty, there are "anchoring" and "imitation" effects between different groups of investors that become prevalent, at least in the short run. In any event, estimating the frequency of these major events from historical data may not make any sense. The philosophical question of whether history repeats itself remains open. Some major market corrections were the product of complex chain reactions and systemic breakdowns (e.g. the market correction of August and September 1998 that followed the demise of LTCM). These systemic breakdowns are better addressed now than they were in the past. However, nothing guarantees that we have anticipated all possible systemic market breakdowns, which are unlikely to happen twice in the same form. The second issue involves the complexity of behavioral finance models and their game-theoretic nature. Some models break down financial agents into different groups and argue that large swings in markets can be the product of those different groups. One relatively clear instance of this phenomenon was the Internet boom-and-bust of the late 1990s and early 2000s. In this instance, we believe (and will further address at a later stage) that there was an asymmetric equilibrium between different groups of investors (modelling and proving this is the case is not an easy task).

A.2 Epistemological Considerations in Finance

Epistemology is usually defined as the branch of science that studies the nature of knowledge, its presuppositions and foundations, and its extent and validity. There is a fundamental epistemological issue with financial theory arising from the fact that financial market dynamics are the consequences of the actions of many agents, each with their own views and beliefs.

It is worth mentioning here a comparison between physics and finance. As is pointed out by Roehner (Roehner 2002) when he describes the relatively new field of econophysics (defined as the "investigation of economic problems by physicists"), there is been little emphasis put on empirical work in economics or finance compared to other disciplines such as physics. As Roehner points out, "several Nobel prizes have been awarded for work which remains completely theoretical ... one would in vain search for any statistical test in the works of [Samuelson, Debreu or Allais][10]. This striking contrast emphasizes the fact that observation and experimental evidence have a completely different status in physics and economics."

A.2.1 Two-Body Problems and Multi-Agent Systems

As Roehner further points out (Roehner 2002), "most of the problems that physics and biology were able to solve in the nineteenth and early twentieth century were of the two-body type." Whether it is the Sun-Mars interaction studied by Kepler (in the 17th century), the two heat reservoir problem studied by Carnot in thermodynamics (in the 19th century), the proton-electron model by Bohr or the Sun-Mercury refined trajectories calculated by Schwarzschild (in the 20th century), these were all works that assumed a limited level of complexity (the interactions between two bodies).

As Roehner further suggests, much of the early theories of economics or finance were not tested by empirical evidence because "no real economic systems matched the two-body assumptions even in an approximate way."

[10] Our editing.

Roehner further states that the level of complexity in economics or finance is of the order of n-non identical body problems. The financial markets' complexity can be compared to that of a **meteorological system with n-non identical bodies of air and water masses interacting**.

In our view, finance is an even more complex system than physical systems with n interacting bodies as financial markets can be best represented as multi-agent systems (not multi-body systems), with each agent having a mind of his/her own. Hence the behavior of each component of the overall system can be highly complex and unpredictable. It cannot be assumed to follow a couple of simple laws (such as short and long range interactions, etc.) Each agent's behavior is best described by a game theoretic framework where beliefs about the market in general and about other agents' behavior need to be taken into account.

This idea was popularized by Soros (Soros 1987) who notes that *"the participants' thinking introduces an element of uncertainty into the subject matter... Situations which have thinking participants may be impervious to the methods of natural science but they are susceptible to the methods of alchemy. The thinking of participants, exactly because it is not governed by reality, is easily influenced by theories. In the field of natural phenomena, scientific method is effective only when its theories are valid; but in social, political and economic matters, theories can be effective without being valid."*

Bibliography

Alexander, S., 1961, "Price Movements in Speculative Markets: Trends or Random Walks?," *Industrial Management Review*, 2, 7–26.

Aït-Sahalia, Y., and P. Mykland, 2004, "Estimators of Diffusions with Randomly Spaced Discrete Observations: A General Theory," *The Annals of Statistics*, 32, 2186–2222.

Arrow, K., 1986, "Rationality of Self and Others in an Economic System," *Journal of Business*, 59, 385–400.

Bachelier, L., 1900, "Théorie de la Spéculation," *Annales de l'Ecole Normale Supérieure*, 17, 21–86, English translation by A.J. Boness can be found in 'The Random Character of Stock Market Prices,' edited by P.H. Cootner, at MIT Press (1967), pp. 17-78.

Bell, D., H. Raiffa, and A. Tversky, 1988, *Decision Making - Descriptive, Normative, and Prescriptive Interactions*, Cambridge University Press, Cambridge, U.K.

Bellman, R., 1957, *Dynamic Programming*, Princeton University Press, Princeton, N.J.

Bernstein, P., 1992, *Capital Ideas, The Improbable Origin of Modern Wall Street*, The Free Press, New York.

Bikhchandani, S., and S. Sharma, 2001, "Herd Behavior in Financial Markets," *IMF Staff Papers*, 47.

Birge, J., and F. Louveaux, 1997, *Introduction to Stochastic Programming*, Springer-Verlag New York.

Blake, D., B. Lehmann, and A. Timmermann, 1999, "Asset Allocation Dynamics and Pension Fund Performance," *Journal of Business*, 72, 429–461.

Boender, G., 1997, "A hybrid simulation/optimization scenario model for asset/liability management," *European Journal of Operational Research*, 99, 126–135.

Brandimarte, P., 2002, *Numerical Methods in Finance: A MATLAB-based Introduction*, John Wiley & Sons.

Brigo, D., and F. Mercurio, 2000, "Discrete Time vs Continuous Time Stock-price Dynamics and implications for Option Pricing," Reduced version in Finance & Stochastics 4, pp. 147-159, February 2000.

Brinson, G., L. Hood, and G. Beebower, 1986, "Determinants of Portfolio Performance," *Financial Analysts Journal*, 42, 39–44.

Brinson, G., L. Hood, and G. Beebower, 1991, "Determinants of Portfolio Performance II: An Update," *Financial Analysts Journal*, 47, 40–48.

Campbell, J., and L. Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* . Clarendon Lectures in Economics, Oxford University Press.

Cariño, D., and A. Turner, 1998, "Multiperiod Asset Allocation with Derivative Assets," in W.T. Ziemba, and J.M. Mulvey (ed.), *Worlwide Asset and Liability Modeling* . pp. 182–204, Cambridge University Press, Cambridge, UK.

Chatfield, C., 1996, *The Analysis of Time Series* . Text in Statistical Sciences, Chapman & Hall/CRC, 5th edn.

Collomb, B., 1971, "Goal-Interval Approaches to Intertemporal Analysis and Decentralization in Management," Ph.D. thesis, University of Texas at Austin.

Cootner, P., 1962, "Stock Prices: Random vs. Systematic Changes," *Industrial Management Review*, 3, 24–45.

Cowles, A., 1933, "Can Stock Market Forecasters Forecast?," *Econometrica*, 1, 309–324.

Cowles, A., and H. Jones, 1937, "Some A Posteriori Probabilities in Stock Market Action," *Econometrica*, 5.

CPLEX Optimization, 1989, *Using the CPLEX Callable Library*Incline Village, NV 89451.

Dantzig, G., and G. Infanger, 1993, "Multi-stage stochastic linear programs for portfolio optimization," *Annals of Operations Research*, 45, 59–76.

Dempster, M., I. Evstigneev, and K. Schenk-Hoppé, 2003, "Exponential Growth of Fixed Mix Assets in Stationary Markets," *Finance and Stochastics*, 7, 263–276.

Doan, T., R. Litterman, and C. Sims, 1984, "Forecasting and conditional projections using realistic prior distributions," *Econometric Reviews*, 3, 1–100.

Duffie, D., 2001, *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, N.J., 3rd edn.

Dupačová, J., 1999, "Portfolio Optimization via Stochastic Programming," *Mathematical Methods of Operations Research*, 50, 245–270.

Efron, B., and R. Tibshirani, 1993, *An Introduction to the Bootstrap* . No. 57 in Monographs on Statistics and Applied Probability, Chapman & Hall/CRC.

Fama, E., 1965, "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34–105.

Fleten, S., K. Høyland, and S. Wallace, 2002, "The Performance of Stochastic Dynamic and Fixed Mix Portfolio Models," *European Journal of Operations Research*, 140, 37–49.

Friedman, M., 1957, *A Theory of the Consumption Function*, Princeton University Press, Princeton, N.J.

Friedman, M., and L. Savage, 1948, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56, 279–304.

Gill, P., W. Murray, and M. Wright, 1986, *Practical Optimization*, Academic Press, London, UK.

Granger, C., and O. Morgenstern, 1964, "Spectral Analysis of New York Stock Market," *Kyklos*, 17, 162–188.

Green, T., and S. Figlewski, 1999, "Market Risk and Model Risk for a Financial Institution Writing Options," *Journal of Finance*, 54, 1465–1499.

Hamilton, J., 1994, *Time Series Analysis* . chap. 11, pp. 291–350, Princeton University Press.

He, H., 1989, "Convergence from Discrete to Continuous Time Financial Models," Finance Working Paper 190.

Hensel, C., D. Ezra, and J. Ilkiow, 1991, "The Importance of the Asset Allocation Decision," *Financial Analysts Journal*, 47, 65–72.

Hong, H., J. Kudik, and A. Solomon, 2000, "Security Analysts' Career Concerns and Herding of Earnings Forecasts," *RAND Journal of Economics*, 31, 121–44.

Houthaker, H., 1961, "Systematic and Random Elements in Short Term Price Movements," *American Economic Review*, 51, 164–172.

Infanger, G., 1993, *Planning under Uncertainty*, Boyd & Fraser Publishing Company.

Infanger, G., 1997, *DECIS User's Guide*.

Infanger, G., 1999, "Managing Risk Using Multi-Stage Stochastic Optimization," Technical Report SOL 99-2, Systems Optimization Laboratory, Stanford, CA.

Infanger, G., 2002, "MS&E 348 Class Notes," Stanford University, CA.

Jensen, M., 1968, "The Performance of Mutual Funds in the Period 1945-1964," *Journal of Finance*, 23, 387–419.

Jensen, M., 1969, "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios," *Journal of Business*, 42, 167–247.

Johansen, A., and D. Sornette, 2001, "Large Stock Market Price Drawdowns Are Outliers," *Journal of Risk*, 4, 69–110.

Judge, G., R. Hill, W. Griffiths, U. Lutkepohl, and T. Lee, 1988, *The Theory and Practice of Econometrics*, John Wiley & Sons, New York.

Kahneman, D., and A. Tversky, 1973, "On the Psychology of Prediction," *Psychological Review*, 80, 237–51.

Kahneman, D., and A. Tversky, 1979, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 363–91.

Kahneman, D., and A. Tversky, 1984, "Choices, Values and Frames," *American Psychologist*, 39, 341–50.

Kendall, M., 1953, "The Analysis of Economic Time Series - Part 1: Prices," *Journal of the Royal Statistical Society*, 96, 11–25, Series A.

Kreps, D., 1990, *A Course in Microeconomic Theory* . chap. 18, pp. 661–719, Prentice Hall.

Larson, A., 1960, "Measure of a Random Process in Future Prices," *Food Research Institute Studies*, 1.

LeSage, J., 1999, *Applied Econometrics using MATLAB*Department of Econometrics, University of Toledo.

LeSage, J., and M. Magura, 1991, “Using interindustry input-output relations as a Bayesian prior in employment forecasting models,” *International Journal of Forecasting*, 7, 231–238.

LeSage, J., and Z. Pan, 1995, “Using Spatial Contiguity as Bayesian Prior Information in Regional Forecasting Models,” *International Regional Science Review*, 18, 33–53.

Litterman, R., 1980, “A Bayesian Procedure for Forecasting With Vector Autoregression,” Working Paper.

Litterman, R., 1986, “Forecasting with Bayesian Vector Autoregressions - Five Years of Experience.,” *Journal of Business and Economic Statistics*, 4, 25–38.

Luenberger, D., 1998, *Investment Science*, Oxford University Press.

Malkiel, B., 2003, *A Random Walk Down Wall Street*, W.W. Norton & Company, New York, 8th edn.

Mandelbrot, B., 1962, “Sur certains prix spéculatifs: faits empiriques et modèle basé sur les processus stables additifs non gaussiens de Paul Lévy,” Compte-rendu pp. 3968-3970, Académie des Sciences, Séance du 4 juin 1962.

Mandelbrot, B., 1971, “When Can Price Be Arbitraged Efficiently? A Limit to the Validity of the Random Walk and Martingale Models,” *Review of Economics and Statistics*, 53, 225–236.

Mandelbrot, B., 1997, *Fractals and Scaling in Finance*, Springer.

Markowitz, H., 1952a, “Portfolio Selection,” *Journal of Finance*, 7, 77–91.

Markowitz, H., 1952b, “The Utility of Wealth,” *Journal of Political Economy*, 60, 151–58.

Markowitz, H., 1956, “The Optimization of a Quadratic Function Subject to Linear Constraints,” *Naval Research Logistics Quarterly*, 3, 111–133.

Markowitz, H., 1987a, *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell, New York.

Markowitz, H., 1987b, *Portfolio Selection*, Wiley, New York.

Merton, R., 1969, “Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case,” *Review of Economics and Statistics*, 51, 247–57.

Merton, R., 1971, "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, 3, 373–413.

Merton, R., 1990, *Continuous-Time Finance*, Oxford, UK.

Moore, A., 1962, "A Statistical Analysis of Common Stock Prices," Ph.D. thesis, University of Chicago, Graduate School of Business.

Mossin, J., 1968, "Optimal Multiperiod Portfolio Policies," *Journal of Business*, 41, 215–229.

Mulvey, J., and H. Vladimirou, 1989, "Stochastic Network Optimization Models for Investment Planning," *Annals of Operations Research*, 20, 187–217.

Murtagh, B., and M. Saunders, 1983, *MINOS User's Guide*Department of Operations Research - SOL, Stanford, CA, SOL 83-20.

Osborne, M., 1959, "Brownian Motion in the Stock Market," *Operations Research*, 7, 145–173.

Roberts, H., 1967, "Statistical versus Clinical Prediction of the Stock Market," Unpublished Manuscript.

Robertson, J., and E. Tallman, 1999, "Vector Autoregressions: Forecasting and Reality," *Economic Review of Federal Reserve Bank of Atlanta*.

Roehner, B., 2002, *Patterns of Speculation: A Study in Observational Econophysics*, Cambridge University Press.

Royall, R., 1997, *Statistical Evidence, A Likelihood Paradigm* . , vol. 71 of *Monographs on Statistics and Applied Probability*, Chapman Hall CRC.

Ruszczyński, A., and A. Shapiro, 2003, *Decomposition Methods*, vol. 10 of *Handbooks in OR&MS*, . chap. 3, Elsevier Science B.V.

Samuelson, P., 1969, "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *The Review of Economics and Statistics*, 51, 239–246.

Savage, L., 1954, *Foundations of Statistics*, Wiley, New York, 2nd ed., Dover, N.Y., 1972.

Sharpe, W., 1966, "Mutual Fund Performance," *Journal of Business*, 39, 119–138.

Shefrin, H., and M. Statman, 1985, "The Disposition to Sell Winners Too Early and Ride Losers Too Long," *The Journal of Finance*, XL, 777–790.

Shiller, R., 1984, "Stock Prices and Social Dynamics," *The Brookings Papers on Economic Activity*, 2, 457–510.

Sims, C., 1980, "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Sornette, D., 2003, *Why Stock Markets Crash: Critical Events in Complex Financial Systems*, Princeton University Press, Princeton, New Jersey.

Soros, G., 1987, *The Alchemy of Finance: Reading the Mind of the Market*, John Wiley & Sons, Inc.

Strang, G., 1988, *Linear Algebra and its Applications* . chap. 7, pp. 380–86, Brooks/Cole Thomson Learning, 3rd edn.

Thaler, R., 1980, "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization*, 1, 39–60.

Thaler, R., 1981, "Some Empirical Evidence on Dynamic Inconsistency," *Economic Letters*, 8, 201–7.

Thaler, R., 1985, "Mental Accounting and Consumer Choice," *Marketing Science*, 4, 199–214.

Thaler, R., 1992, *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, Princeton University Press, Princeton, N.J.

Treynor, J., 1965, "How to Rate Management of Investment Funds," *Harvard Business Review*, 43, 63–75.

Tversky, A., and D. Kahneman, 1991, "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106, 1039–61.

von Neumann, J., and O. Morgenstern, 1944, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, N.J.

Walter, C., 2003, "The Efficiency Market Hypothesis, the Gaussian Assumption, and the Investment Management Industry," Working Paper.

Working, H., 1934, "A Random Difference Series for Use in the Analysis of Time Series," *Journal of the American Statistical Association*, pp. 11–24.

Working, H., 1960, “Notes on the Correlation of First Differences of Averages in a Random Chain,” *Econometrica*, 28, 916–918.

Ziemba, 2003, Monograph from The Research Foundation of AIMR.