Multistage Mean-Variance Portfolio Selection in Cointegrated Vector Autoregressive Systems

by

Melanie Beth Rudoy

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2009

Author .....
Department of Electrical Engineering and Computer Science
January 27, 2009

Certified by .....
Dr. Charles E. Rohrs
Research Scientist, Digital Signal Processing Group
Thesis Supervisor

Accepted by .....
Professor T. P. Orlando
Chairman, Department Committee on Graduate Students

Multistage Mean-Variance Portfolio Selection in Cointegrated Vector Autoregressive Systems

by

Melanie Beth Rudoy

Submitted to the Department of Electrical Engineering and Computer Science
on January 27, 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The problem of portfolio choice is an example of sequential decision making under uncertainty. Investors must consider their attitudes towards risk and reward in face of an unknown future, in order to make complex financial choices. Often, mathematical models of investor preferences and asset return dynamics aid in this process, resulting in a wide range of portfolio choice paradigms, one of which is considered in this thesis. Specifically, it is assumed that the investor operates so as to maximize his expected terminal wealth, subject to a risk (variance) constraint, in what is known as mean-variance optimal (MVO) portfolio selection, and that the log-prices of the assets evolve according a simple linear system known as a cointegrated vector autoregressive (VAR) process. While MVO portfolio choice remains the most popular formulation for single-stage asset allocation problems in both academia and industry, computational difficulties traditionally limit its use in a dynamic, multistage setting. Cointegration models are popular among industry practitioners as they encode the belief that the log-prices of many groups of assets are not WSS, yet move together in a coordinated fashion. Such systems exhibit temporary states of disequilibrium or relative asset mis-pricings that can be exploited for profit.

Here, a set of multiperiod trading strategies are developed and studied. Both static and dynamic frameworks are considered, in which rebalancing is prohibited or allowed, respectively. Throughout this work, the relationship between the resulting portfolio weight vectors and the geometry of a cointegrated VAR process is demonstrated. In the static case, the performance of the MVO solution is analyzed in terms of the use of leverage, the correlation structure of the inter-stage portfolio returns, and the investment time horizon. In the dynamic setting, the use of inter-temporal hedging enables the investor to further exploit the negative correlation among the inter-stage returns. However, the stochastic parameters of the per-stage asset return distributions prohibit the development of a closed-form solution to the dynamic MVO problem, necessitating the use of Monte Carlo methods. To address the computational limitations of this numerical approximation, a set of four approximate dynamic schemes are considered. Each relaxation is suboptimal, yet admits a tractable solution. The relative performance of these strategies, demonstrated through simulations involving synthetic and real data, depends again on the investment time horizon, the use of leverage and the statistical properties of the inter-stage portfolio returns.

Thesis Supervisor: Dr. Charles E. Rohrs
Title: Research Scientist, Digital Signal Processing Group

Acknowledgments

No words can fully capture the love and gratitude I have for my best friend and husband, Daniel Rudoy. From the beginning of our relationship, he has always encouraged me to work harder and dream bigger, whether it be academically, professionally, or personally. He is always looking out toward the next horizon, always asking "What's next?" rather than "Are we done yet?". I am fortunate to have a husband who is also a colleague, who takes the time to listen to my ideas, understand my work, and proofread my papers. I can't imagine going on this journey without you, and I know the best is yet to come. I love you.

To my advisor, Charlie Rohrs: A million thank yous! I feel extremely fortunate to have worked with an advisor who cares so deeply about my personal development. For Charlie, the goal is not to push papers out the door, but rather to develop deep understanding and insight. From preparing for quals to discussing research, Charlie has taught me the fundamental skills of formulating and solving problems, of identifying what I know and what I don't know, and of reconciling results with intuition. I will always remember our research collaborations, joint teaching experiences, and of course, the delicious football parties.

I am greatly indebted to my committee. Thank you to Professor Al Oppenheim for encouraging me to always question assumptions, to never blindly accept what is commonly stated in practice as truth. Thank you to Professor John Wyatt for pushing me to think critically about my own work, to understand both its strengths and its weaknesses. Thank you to Professor Andrew Lo for the technical suggestions and ideas, and for giving an electrical engineer the confidence to know her work is legitimate from within the finance world.

In addition to serving on my committee, Professor Oppenheim also played an extremely important role in guiding me through the early stages of the PhD program. Through his graduate counseling group, I had the privilege of learning the ins and outs of grad school both from his perspective, and that of upperclassmen Laura Zager and Zahi Karam and a long list of MIT faculty and administrators. Needless to say, embarking on a five year graduate program is a daunting task, and sometimes just knowing that there is someone in your corner cheering you on is all you need to make it through the day. And no, I didn't

forget about the six words – *"One step. Next step. Repeat. Done."*

One of the most important parts of graduate school is being part of a peer group. I am extremely fortunate to have had the opportunity to be a member of such a wonderful research lab, the Digital Signal Processing Group (DSPG) within the Research Laboratory of Electronics (RLE). Thank you to all of the students, past and present – Petros Boufounos, Maya Said, Sourav Dey, Tom Baran, Zahi Karam, Al Kharbouch, Dennis Wei, Joe Sikora, Shay Maymon, John Paul Kitchens, Archana Venkataraman, Ross Bland, Matt Willsey, Joonsung Lee, and the newest member, Jeremy Leow. It is rare to find a group of people who are not only excited about their own work, but genuinely care about the work of those around them. It is a common sight here to walk into the lab on a random afternoon and see three people crowded around a white board, brainstorming some new idea or deriving some result. Thank you to my office mate during the first four years, Tom Baran. Tom, you are a true engineer at heart – I have always been inspired by your love of playing and tinkering. Thank you to Sourav for believing early on that the topic of portfolio optimization and cointegration would ultimately develop into a thesis – there are very few people who *see* linear algebra as clearly as you do! Last, but certainly not least, I need to say a huge thank you to the man who makes the group function day to day, Eric Strattman. Eric is the reason the lab runs smoothly and efficiently, and I greatly appreciate how his efforts have facilitated my grad school experience.

There are many people outside of MIT who helped shape and guide this thesis. The first is Jay Damask, who handed me my first book on cointegration during the summer of 2006, and said *learn this, implement this, let's make some money from this*! That internship literally gave birth to this thesis, and for that I am eternally grateful. Second is the algorithmic trading group at GMO, where I worked during the summer of 2008. Thank you to Mark Mueller, Chris Darnell, Nehal Patel, and Nick Choo, from whom I learned a great deal about multiperiod trading models and DP. I look forward to starting the next chapter of my life working among such a talented and impressive group of people.

Over the course of my professional and academic career, I have been fortunate to have had the honor of learning from an amazing group of teachers and mentors. Thank you

to Professor Max Mintz for inspiring me with your boundless energy and genuine love for teaching. Thank you to Professor Thad Welch for showing me how much fun DSP can be. Thank you to Arnold and the entire PEDS team for believing in me and giving me the confidence to pursue the PhD dream.

I know this thesis would not have been possible without the love and support of my family. Thank you to my parents, Shirley and Mark Leader, for understanding why I wanted to quit my job, sell my house and car, move up to Boston, and enter into a graduate program with no set roadmap, no end in sight. While I know they wish that I lived closer so we could spend more time together, I appreciate the fact that they have always encouraged me to go out into the world and make my own path. My work ethic and determination to succeed comes from my mother, who always did what she needed to do to support her two kids. Thank you to Dan's parents, Gregory Rudoy and Alexandra Dashevskaya, who have always provided me with an immeasurable amount of support and encouragement (and goodies from the Russian store). I will never forget how it felt to see all four of them in the audience at my defense, smiling and nodding their heads as I rambled on. Thank you to Craig Shames for being a great and supportive brother.

Someone once said "Friends are the most important ingredient in this recipe of life." I want to thank all of my friends who have stood by me over the past five years, despite my awful phone habits (or lack thereof!). In particular, I know I would not have survived grad school without Emily Fox and Erin Aylward – I have always admired your determination and confidence. To my running buddy, Leah, I know that my experience training for those races with you helped me to develop the mental toughness needed to make it through the "last lap" of grad school. To Judy, for inspiring me with your fighting spirit and courage to rise above any challenge. To Aysha, for always believing in me and for following your dreams. To Heidi, for being one of the first people to put the PhD idea in my head and for making everything look so simple. To Kristy and Galina, for setting an example in taking risks and pursuing dreams. To Anne, for pushing me to finish. To Ravi and Jaime, for your support and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is concerned with the application of signal processing and stochastic control theory to financial decision making. Specifically, the problem of multistage portfolio selection within a restricted universe of financial assets is considered. Since the 1950s, many variations of this theme have been explored in the finance literature. This has led to the classification of portfolio selection problems according to an extensive list of properties, including investor preferences, time horizons, and statistical models for the investment opportunities. One important, yet little studied, case is when the prices (or log-prices) of the underlying assets are assumed to evolve according to a particular type of linear system, known as a cointegrated vector autoregressive (VAR) process. Here, the random process modeling each component process is nonstationary[1]; however, it is possible to find linear combinations of the signals that produce asymptotically wide-sense stationary (AWSS) random processes. Constraining the price dynamics to follow such a model induces a particular statistical structure on the per-stage asset returns, knowledge of which can be exploited to achieve a higher mean portfolio return over the cumulative investment period.

The organization of this chapter is as follows. Section 1.1 explains why the cointegrated VAR process was chosen as the basis for this thesis, and is followed by a brief description of the intended audience for this work in Section 1.1.2. The utility of this research in light of the ongoing global financial crisis is discussed in Section 1.1.3. The role of quantitative models within the finance world is discussed in Section 1.1.4, and the inherent limitations of the methods and models used in thesis are outlined in Section 1.1.5. Finally, in Section 1.2, a detailed survey of the overall thesis is given.

[1]Throughout this thesis, the term *nonstationary* refers to a specific type of nonstationarity, in which the process variance grows linearly over time due to the presence of one or more integrators (poles at unity). Such processes are referred to as marginally unstable systems, and are described in Section 3.1.

1.1 Thesis Background and Motivation

1.1.1 Why Cointegration

The model at the forefront of this thesis is the cointegrated vector autoregressive (VAR) process. Pick up your favorite econometrics textbook, and without doubt, there will be at least one entry under "cointegration" in the index [19, 27, 60]. These linear systems are commonly used by both academics and industry practitioners to describe groups of time series in which each underlying component process is well-modeled by a nonstationary process that contains one or more integrators, yet there exists a linear combination of the signals that produces an asymptotically wide-sense stationary (AWSS) random process. In the time domain, such signals often appear to move together in a coordinated fashion, a property which is commonly present in financial time series, such as stocks from the same economic sector, or government bonds covering varying times to maturity on the yield curve. The cointegrated VAR process is a simple model that encodes the belief that while the prices (or log-prices) of many financial instruments are nonstationary, the corresponding returns (or log-returns) admit asymptotically stationary distributions, with non-negligible inter-asset *and* inter-temporal correlations.

While detection and estimation techniques for this class of models are well studied, surprisingly little has been written about how to trade such systems in a multiperiod environment. As discussed in greater detail in Section 3.5.1, the majority of the existing literature focuses on statistical arbitrage techniques that exploit the *mean-reverting* property of the AWSS linear combination of the underlying series. In this thesis, a set of techniques for trading a system of cointegrated assets is developed, that is optimal in the "mean-variance" sense, a criterion that is precisely defined in Section 2.3.1. The resulting trading strategies tend to be *market neutral*, as it is possible to earn profit both from increasing and decreasing asset prices. Trading systems built around cointegrated securities exploit the co-movement of prices, and are therefore robust in both bull and bear markets. In addition, during times of crisis, asset returns tend to become more correlated and the corresponding prices tend to exhibit an increased degree of co-movement, potentially increasing the applicability of cointegration models.

The study of linear systems within the field of electrical engineering has a long and fruitful history, and this thesis offers a new look at cointegrated systems through a detailed linear systems approach. In particular, the use of state-space analysis methods for the study of cointegrated VAR processes enables one to gain a deeper understanding of the underlying structure, or geometry, of the system. The resulting portfolio weight vectors can also be interpreted within the context of this geometry, providing a deeper level of understanding and intuition for each strategy. For example, the *direction* of the portfolio weight vectors relative to the underlying subspaces of a cointegrated VAR process indicate whether the *short-term error-correcting* forces or the *long-term common-trend* forces more heavily influence the investment decision. In addition, the *length* of the portfolio weight vector indicates how the notion of *leverage* is utilized in order to achieve a desired level of risk. Throughout this thesis, the geometry of asset allocation is considered, both for a fixed time interval and as the length of the investment time horizon increases.

1.1.2 Intended Audience

The intended audience for the asset allocation schemes presented in this thesis are professional traders that work within proprietary trading groups or hedge funds. Despite the relatively simple underlying framework, a fair degree of mathematical sophistication, network infrastructure, and computing power are required to run the strategies period to period[2]. The average retirement investor typically relies on a long-term *buy and hold* strategy [56], and may, at most, rebalance the portfolio holdings once or twice a year. On the other hand, Wall Street traders often build algorithmic trading strategies, where computerized models make a multitude of trading decisions per day, executing them with no human intervention. These systems often fall into the class of *statistical arbitrage* methods, in which relatively small gains can be made over short time horizons by identifying temporary asset mis-pricings within the financial marketplace. However, implementation disclaimer aside, the general themes developed in this thesis and lessons learned regarding the fundamental tradeoff between risk and return apply at some level to all investors, regardless of whether or not they utilize the algorithms in practice.

[2]The length of time within one period is left up to the trading system designer to determine. One period could correspond to a year, a day, a second, etc.

1.1.3 The Global Financial Crisis

It is a precipitous task to write a thesis on the topic of portfolio selection, from the perspective of an electrical engineer, in the middle of a global financial crisis. The collapse of the sub-prime mortgage market led to the decline of many Wall Street firms, including Bear Stearns, Lehman Brothers and Merrill Lynch. In early September, the United States government was called upon to bail out mortgage lenders Fannie Mae and Freddie Mac, and shortly followed suit with a rescue package for the insurance giant AIG. On October 3, the Congress passed the Emergency Economic Stabilization Act of 2008, which authorized the Treasury department to spend up to 700 billion dollars to purchase "distressed" assets directly from the banks, which includes complex mortgage-backed securities. But the crisis is not limited to the housing financial industry. The value of worldwide equity and commodity markets also declined, as banks around the world scrambled to get bad loans off their books. The commercial paper market, the market in which corporations finance their operating expenses, completely dried up, driving stock prices even further down, and the banks stopped lending money to each other. The full effect of the events of September and October 2008 remains to be seen.

In writing this thesis, the backdrop of a global crisis has a few benefits, and naturally, a few disadvantages. On the one hand, the crisis brings to light the need for expanded research from government, industry, and academia into *all areas* of finance, including corporate finance, financial markets, securitization, financial regulation, and investment management. While this thesis falls under the investment management umbrella, it only scratches the surface of complex issues such as risk management, the effect of leverage (borrowing), and the impact of government regulation (such as a ban on short-sales). No single model or framework can capture all aspects of portfolio selection, or hope to be equally relevant to all investors, independent of their goals and risk preferences. However, each piece of research brings to bear its own insights and intuition, and only through the aggregation of these ideas can we begin to better understand the financial world around us.

While the current crisis highlights the need for financial research, it also brings into question the legitimacy of complex mathematical models of financial systems. Many trading strategies fail in the middle of a crisis. As a result, banks or hedge funds that do not ap-

propriately manage their risk can go bankrupt overnight as they find themselves unable to cover huge accumulated losses. While some degree of disbelief is healthy, these models do play an important role in helping managers to understand financial markets and products, a theme that is explored in greater detail in Section 1.1.4.

Throughout much of this thesis, the reader is asked to operate within an idealized financial market, far from the realities of the global crisis, one in which there is an infinite supply of liquidity and trades are executed immediately at the desired price, with no transaction costs. The reader must trust that the asset return model is correct and remains valid over the entire investment horizon. However, such an idealized world does not exist, and therefore this thesis also addresses *some practical* restrictions and issues, such as the use of a budget-constraint, the presence of a no short-sale constraint, and the impact of model parameter estimation error. One is not so naive to think that one can create a model that perfectly describes every aspect of a true trading system. As this thesis demonstrates, there is still much knowledge to be gained from the study of *simple models* that capture even a small subset of *real-world behavior*.

1.1.4 The Role of Financial Models

Quantitative models can be found in all areas of modern finance theory. There are models that quantify the relationship between the expected returns among a set of assets and their relative risk levels, such as the Capital Asset Pricing Model (CAPM) or the Arbitrage Pricing Theory (APT). There are models used to price financial products, including stocks, bonds, options, and complex instruments such as mortgage-backed securities. And of particular interest for this thesis, there are models used to guide asset allocation decisions, such as the Markowitz Mean-Variance criterion or the set of expected utility-theoretic paradigms. While this list is in no sense complete, it begins to give a sense of the wide range of topics covered under the heading of mathematical finance.

Econometricians and financial engineers seek models that can capture some measure of observed market behavior, such as inter-asset return correlations or long-term common trends between macro-economics variables. In contrast to the natural sciences where one can devise repeatable experiments to test a hypothesis, in the field of quantitative finance

only one sample path of a random process is available. One must identify interesting behavior from historical data alone and devise a model that explains not only the current observation set, but also future data. No one expects a financial model to be able to perfectly predict tomorrow's stock prices, or to be able to perfectly explain the relationship between stock market returns and government bond yields. A model is just a model, and is meant to guide an investor's actions, not dictate them.

It is important to realize the fundamental limits of any financial model. A detailed analysis of the limitations of the model used in this thesis is presented in Section 1.1.5. Perhaps some on Wall Street have still not learned the lesson that ultimately brought down the hedge fund Long-Term Capital Management (LTCM) in 1998, after Russia defaulted on its government bonds. The following description of one of LTCM's partners, Lawrence Hilibrand, from Roger Lowenstein's book *When Genius Failed*, best summarizes the type of blind faith in financial models that can at times lead to devastating consequences,

> If the firm could have been distilled into a single person, it would have been Hilibrand. While veteran traders tend to be cynical and insecure, the result of years of wrong guesses and narrow escapes, Hilibrand was cool and maddeningly self-confident. An incredibly hard worker, he was the pure arbitraguer; he believed in the models, stuck to his prices, was untroubled by doubt. Rosenfeld hated to hedge by selling a falling asset, as theory prescribed; Hilibrand *beleived* and simply followed the form [39].

If the current financial crisis teaches us anything, it is that all financial engineers *should be* troubled by doubt, and *should not* always blindly follow form.

1.1.5 Model and Methodology Limitations

In early November, the following statement appeared in the New York Times, "Todays economic turmoil, it seems, is an implicit indictment of the arcane field of financial engineering.... the larger failure, they say, was human in how the risk models were applied, understood and managed" [38]. In order to prevent future failures of this magnitude, financial engineers must learn the delicate skill of identifying not only the strengths, but also the weaknesses of their models. In this section, the inherent limitations of the models used in

this thesis are given.

First, throughout this thesis, an extremely simple measure of risk is used to describe the investor's preferences, namely the variance of the terminal portfolio return. Future returns are random and are thus not known exactly, and the variance statistic captures the *dispersion* of the set of possible outcomes about the mean or expected outcome. Under the assumption that returns are Normally distributed, the variance completely characterizes the *shape* of the return distribution. However, it is generally accepted that returns are heavy-tailed, meaning that rare events such as extreme positive and negative returns are more likely to occur than predicted by the Gaussian distribution. In addition, the use of variance as a measure of risk implies that the investor cares equally about upside and downside deviations from the mean, which in most cases is not true. Furthermore, the use of variance to define risk does not account for the possibility that the investor could go bankrupt over the course of the investment horizon. While there are many other definitions of risk that do address these concerns, some of which are discussed in Section 2.2, the use of variance is selected due to the fact that it is simple, intuitive, and computationally tractable.

Second, it is assumed that the stochastic input to the cointgerated vector autoregressive system used to describe the evolution of the asset log-prices over time is driven by an independent and identically distributed (i.i.d.) Gaussian random process. This assumption, in turn, induces per-stage asset returns that are jointly Normal and dependent. As stated above, the Gaussian postulate does not capture the heavy-tailed nature of past historical returns, but it does provide computational tractability. In many places throughout this thesis, the given results can be easily extended to consider alternative innovation distributions. There are only three places where the Gaussian assumption is explicitly used. The first is in the maximum likelihood estimator used to determine the parameters of a cointegrated vector autoregressive process from sample data, as described in Section 3.4.2. Second, in Chapter 5, the analytic form of the Gaussian probability distribution function is used to compute a set of importance weights needed to compute the optimal dynamic portfolio weight vector. And third, in Section 6.2, the Gaussian assumption is invoked in order to compute certain expectations in closed-form using Gaussian product moment factoring.

Third, the task of identifying assets in which the log-prices are cointegrated is in itself a non-trivial undertaking. It is often considered an *artform* to identify cointegrated securities or to determine the time scale over which to sample the data. For example, a pair of assets may be well-modeled by a cointegrated VAR process when sampled according to a stochastic arrival process, but not when sampled on a uniform discrete-time grid. In addition, the parameters of a cointegrated VAR process may in practice only remain constant over short time horizons, corresponding to a handful of stages, after which time the models need to be retrained and the trading stratgegy reinitialized. As an alternative to identifying naturally occuring cointegrated securities, one can construct tracking portfolios that are cointegrated with a target financial instrument, such as an individual stock or index fund [2]. The resulting system is cointegrated by design, and the trading strategies developed in this thesis may be directly applied.

1.2 Thesis Organization

Chapter 2 provides the reader with the necessary background related to the theory of portfolio selection. The concepts of simple and log-returns are defined, and a group of common models for the time evolution of returns is presented. A set of portfolio risk measures are surveyed, including those based on the dispersion of a return distribution and the curvature of the investor preference function (i.e., utility function). Single-stage and multistage asset allocation problems are formally defined, and discussed within both the mean-variance and expected utility frameworks.

Chapter 3, in conjunction with Appendix A, provides the reader with an overview of cointegrated vector autoregressive processes, taking a detailed linear systems approach. Both error-correcting and state-space forms for a cointegrated VAR process are discussed, and their equivalence is established. An alternative form of the total system response (i.e., Granger Representation) is given which clearly decomposes the process into stationary and non-stationary components, each of which exists within non-orthogonal subspaces. In addition, the maximum likelihood estimation procedure for a cointegration VAR model is detailed, and the existing literature on cointegration in portfolio theory and econometrics is surveyed.

In Chapter 4, the problem of static asset allocation, when inter-stage rebalancing is prohibited, is explored. In particular, two strategies are studied corresponding to the so-called "Beta" solution, popular among industry practitioners, and the mean-variance optimal (MVO) solution derived here. The performance of these asset allocation rules both with and without a budget constraint is measured, and the statistics of the per-stage portfolio returns are computed. In particular, the covariance, or correlation, between the inter-stage portfolio returns is quantified, and is interpreted within the context of the geometry of a cointegrated VAR system. The role of both per-asset and net portfolio leverage in achieving a given level of portfolio risk is also explored. In addition, the asymptotic properties of the MVO solution are derived, and the conditions are given under which the MVO solution converges to the Beta solution in the limit of an infinite trading horizon.

In Chapter 5, the problem of dynamic asset allocation, when inter-stage rebalancing is allowed, is explored. The original dynamic MVO problem is mapped into an auxilliary framework that enables the sequence of optimal portfolio policies to be computed using dynamic programming. The resulting solution cannot be fully computed in closed-form, and an efficient numerical approximation scheme based on Monte Carlo and importance sampling methods is described. The dynamic MVO asset allocation problem is also presented within the context of a linear quadratic regulator with random system matrices. In addition, the inclusion of a budget constraint, explicit risk-free asset, and a no short-sale constraint are discussed in detail. Lastly, the per-stage portfolio return statistics and asymptotic properties of the dynamic MVO solution are studied.

To address the numerical implementation issues surrounding the MVO dynamic portfolio choice solution, Chapter 6 details a set of four approximate dynamic trading strategies. Each scheme relaxes one or more of the assumptions of the original problem in order to derive a suboptimal, yet tractable, solution. The first strategy, known as the *separable embedding certainty equivalence* approximation scheme, replaces the stochastic parameters of the per-stage return distributions with their time t_0 conditional expectations. The second strategy, known as the *sequential rescaling* approach, imposes the assumption that the solution to an $(N+1)$-stage problem is found by modifying only the scale (i.e., degree of net

leverage), not the direction (i.e., relative asset proportions) of the N-stage optimal solution. The third approximation strategy, known as the *optimal linear* scheme, parametrizes the portfolio policy at each stage using a linear function of the log-prices. Finally, the fourth scheme, known as the *semi-myopic* approach, solves the multistage problem as a series of consecutive single-stage problems. All the dynamic asset allocation schemes considered are compared through an empirical study of their corresponding risk-reward characteristics using synthetic data. Again, the relative performance of each scheme is explained in terms of the use of leverage and the inter-stage return correlations.

Chapter 2

Portfolio Theory

This chapter presents an overview of the theory of portfolio selection, which studies the allocation of capital among a set of investment instruments. Investors often have significantly different objectives, time horizons, risk tolerances and views on investment opportunities, all of which must be taken into account in a systematic manner. For example, some investors seek low-risk, short-term capital preservation, while others, with a high risk tolerance, seek long-term growth. Portfolio theory provides a mathematical framework in which to encode these goals and beliefs.

The organization of this chapter is as follows. In Section 2.1, the fundamental notion of an asset return is defined, and a set of commonly employed models for the evolution of returns over time is presented. A survey of portfolio risk measures is given in Section 2.2, including those based on the dispersion of a return distribution and the curvature of a utility function, which is a mathematical mapping used to encode an investor's attitudes towards risk and reward. In Section 2.3, the single-stage asset allocation problem is defined, and both the Markowitz mean-variance and expected utility frameworks are presented. Lastly, in Section 2.4, the mulitstage portfolio choice problem is defined, and presented from the mean-variance and utility-theoretic points of view.

2.1 Asset Returns

The first step in establishing a mathematical framework for portfolio selection is to select asset prices or returns as the basic unit of measure in order to describe the value of a tradable security over time. While this distinction may seem trivial due to the simple relationship between them, as defined in Eq. 2.1 below, the subsequent impact on the

choice of statistical models is significant. Whereas returns may be positive or negative, asset prices are constrained to be nonnegative, implying that a two-sided distribution (e.g., Gaussian) should not be used to model them. In addition, it is widely accepted that returns, *not prices*, exhibit the properties of stationarity[1] and ergodicity[2] [19]. It is for these reasons that throughout this thesis, the basic portfolio choice problem is formulated as a function of the underlying asset returns. Specifically, the asset *log-returns*, not simple returns, are used, as discussed in Section 2.1.1. Having selected asset returns as the basic unit of measure, the next step is to select a model that describes their evolution over time. A set of common models are presented in Section 2.1.2, while discussion of the specific model used in this thesis, the cointegrated vector autoregressive process, is deferred to Chapter 3.

2.1.1 Types of Returns

Let $p_k \in \mathbb{R}^+$ denote the price of a single asset at time t_k, and let $R_k \in \mathbb{R}$ denote the corresponding return over the period from $(t_{k-1}, t_k]$, defined as:

$$R_k = \frac{p_k - p_{k-1}}{p_{k-1}} = \frac{p_k}{p_{k-1}} - 1, \tag{2.1}$$

which represents the percent change in value of the asset. Here the subscript k indicates that the value of the return is known at time t_k. This type of return is often referred to as a *simple return*. A second type of return, the *log-return*, is defined as the change in the asset's log-price over the length of the investment period, as:

$$r_k = \log(1 + R_k) = \log(p_k) - \log(p_{k-1}). \tag{2.2}$$

Log-returns are also known as *continuously-compounded returns*, since the quantity $\log(1 + R_k)$ represents the equivalent continuously-compounded rate, r_k^c, corresponding to the simple rate R_k. For example, when R_k is 10% per year, then r_k^c is computed as:

$$1 + R_k = 1 + 0.1 = e^{r_k^c} \quad \rightarrow \quad r_k^c = \log(1.1) = 0.0953.$$

[1] For a complete discussion of stationarity, see Section 3.1.

[2] The property of *mean ergodicity*, which loosely states that a time average can be replaced with an ensemble average, is required so that the parameters of a return model can be estimated from a single sample path of historical data.

When R_k is sufficiently near zero, so that the Taylor series approximation given by:

$$\log(1 + R_k) \approx r_k, \tag{2.3}$$

is valid, the log-return is a good proxy for the simple return. The relationship holds with equality when the simple rate R_k is continuously compounded. Note that the *monotonic* relationship between the simple and log-returns, given by $R_k = e^{r_k} - 1$, implies that optimizing in one domain can be equivalent to optimizing in the other. For example, if a portfolio choice problem is constructed so as to maximize the simple return of an investment over a horizon of length T, then the optimal portfolio also maximizes the equivalent log-return over a period of the same length.

In a multiperiod setting, the total simple return across a set of N investment periods, denoted by R_T, is computed as the product of the per-stage simple returns, as follows:

$$1 + R_T = \prod_{k=1}^{N} (1 + R_k) = \prod_{k=1}^{N} \frac{p_k}{p_{k-1}} = \frac{p_N}{p_0}. \tag{2.4}$$

One advantage of using log-returns over simple returns is that the multiperiod log-return is equal to the sum of the per-stage returns, rather than their product, and is given by:

$$\begin{aligned} r_T = \log(1 + R_T) &= \log\left(\prod_{k=1}^{N} (1 + R_k)\right) = \sum_{k=1}^{N} \log\left(\frac{p_k}{p_{k-1}}\right), \\ &= \sum_{k=1}^{N} (\log(p_k) - \log(p_{k-1})) = \log(p_N) - \log(p_0). \end{aligned} \tag{2.5}$$

The additive accumulation of the log-return is beneficial in multiperiod portfolio selection problems, so that efficient computational techniques, such as dynamic programming, can be readily applied.

In addition to the important distinction between multiplicative and additive multistage returns, simple and log-returns have different properties when computing the return of a *portfolio of assets*. While the portfolio simple return is computed as a linear combination of the simple returns of the constituent assets, it is not possible to compute the log-return of a portfolio in the same manner, due to the fact that the log of a sum is not equal to

the sum of logs. In order to circumvent this issue, the approximation of Eq. 2.3 is used[3]. Thus, the single-stage individual asset returns are given by the change in the log-prices, and the per-stage portfolio returns are given by a weighted sum of the asset log-returns. Furthermore, the portfolio returns are assumed to add across stages in accordance with the properties of log-returns.

2.1.2 Statistical Models of Returns

Numerous models exist to describe the evolution of asset returns over time. The choice of model depends on the set of properties one is trying to describe, such as cross-asset and temporal return correlations, mean-reverting behavior, or common stochastic and growth trends. A model may describe the behavior of a single asset, or may jointly define the evolution of a set of dependent assets. As a general rule, the properties of a system can best be understood by choosing the simplest model that captures the set of desired behaviors. This section surveys a set of commonly used models, popular in both the literature and among practitioners.

One simple model is to assume that the per-stage asset returns are best represented as white noise, i.e. are independent and identically distributed (i.i.d.). The returns may be modeled by a Normal distribution, or a heavy-tailed distribution, such as a Pareto or Cauchy distribution. However, one must be careful in a multiperiod setting to understand the impact of the single-stage return model on the corresponding multiperiod return. For example, if the single-stage simple returns are assumed to be Normally distributed, then the multistage returns are no longer Gaussian. On the other hand, if the single-stage log-returns are assumed to be Gaussian, the equivalent multistage returns, now formed as the sum of the per-stage log-returns, are also Gaussian. This, in turn, implies that the single-stage and multistage simple returns follow a *shifted log-normal* distribution as the support of the probability density function is over $[-1, \infty]$ rather than $[0, \infty]$ [19].

In order to capture serial autocorrelation of the returns, a linear Markov process may be used. A common choice is to assume that the log-return, $r[k]$, is well-modeled by a

[3]If the assumption is not valid over the time-scale initially chosen for the problem, one can increase the rate at which the log-price process is sampled until the assumption *is* valid.

first-order autoregressive (AR) process, given by:

$$r_k = \alpha r[k-1] + \sigma z[k],$$

where z is a zero-mean, unit-variance white noise sequence. The resulting process is stationary so long as $|\alpha| < 1$, with zero mean and variance $\sigma_r^2 = \frac{\sigma^2}{1-\alpha^2}$. The corresponding autocorrelation function decays exponentially as $E\left[r[k]r[k-m]\right] = \sigma_z^2 \alpha^{|m|}$.

In some cases, it may be desirable to jointly model the return dynamics of a set of p assets, such as when examining multiple stocks from the same industry. One simple, yet powerful, model is a vector autoregressive (VAR) process, which is able to capture not only the serial correlation of the return processes in time, but also the inter-asset correlations or cross-sectional interactions. Letting $\mathbf{r}_k \in \mathbb{R}^p$ denote a vector of stacked asset returns at time k, an $L^{\text{th}}-$order VAR process is defined as:

$$\mathbf{r}_k = \sum_{i=1}^{L} \mathbf{A}_i \mathbf{r}_{k-i} + \mathbf{z}_k, \tag{2.6}$$

where the matrices $\mathbf{A}_i$ encode the coupling relationships between $\mathbf{r}_k$ and its lags, and $\mathbf{z}_k \in \mathbb{R}^p$ is typically assumed to be an i.i.d. zero-mean Gaussian process with covariance matrix $\mathbf{\Psi}$. An overview of these processes is presented in Appendix A.

2.2 Risk Measures

The second step in establishing a mathematical framework for portfolio selection is to define a measure for investment risk. Fundamentally, portfolio choice is the study of risk and reward. For assets within the class of "risk-free securities," the value of the asset at any future date is known exactly, and is computed according to a deterministic schedule. While the concept of a risk-free asset is an idealization, for all intents and purposes U.S. Treasuries, particularly T-bills, can be thought of as providing a guaranteed rate of return with no default risk. For all other assets, the future value is not known with certainty, placing these investments within the class of "risky securities." In general, many investors make financial decisions so as to maximize their reward, i.e. portfolio return, while placing some type of constraint on the associated risk. However, not all investors agree how risk should be

defined in a precise mathematical manner. Here, two common approaches are summarized. First, in Section 2.2.1, risk is defined by the dispersion of the return distribution. Second, in Section 2.2.2, risk is defined in terms of the curvature of the investor preference function (utility function) of wealth. Given these two risk paradigms, Section 2.2.3 concludes with a discussion of portfolio budget constraints, which limit the amount of risk an investor may assume.

2.2.1 Return Distribution Dispersion

Perhaps the most popular proxy for risk used in portfolio choice problems is return variance. In [42], Markowitz simply states that variance, or standard deviation, is the "natural measure" of the dispersion of the return distribution. Indeed, when it is assumed that returns are Normally distributed, the variance exactly defines this dispersion. In addition, the variance of a portfolio of multiple assets can easily be computed by knowing the variances and correlations among the constituent components. Given this definition of risk, Markowitz hypothesized that investors act to maximize the expected return of the portfolio, subject to a variance constraint, or equivalently, to minimize the variance of the portfolio subject to a constraint on the desired portfolio return. A detailed discussion of Markowitz's mean-variance portfolio choice framework is presented in Section 2.3.1.

Later, in [43], Markowitz proposed the use of the semi-variance (variance of deviations below the mean) as a measure of return distribution dispersion that focuses on "downside-risk." However, when the return distribution is symmetric, both variance and semi-variance measures produce identical results. A generalization of the semi-variance idea, known as the Lower Partial Moment [11, 25], considers the dispersion of returns relative to a benchmark other than the mean, such as the risk-free rate or the zero return. In practice, the choice of dispersion statistic should take into account both the shape of the return distribution and some baseline performance metric.

2.2.2 Utility-Theoretic Risk

In addition to Markowitz's variance risk criteria, an investor's attitudes towards risk and reward may be encoded in a more general mathematical mapping known as a *utility function*. Under the assumption that more wealth is preferred to less wealth, all utility functions

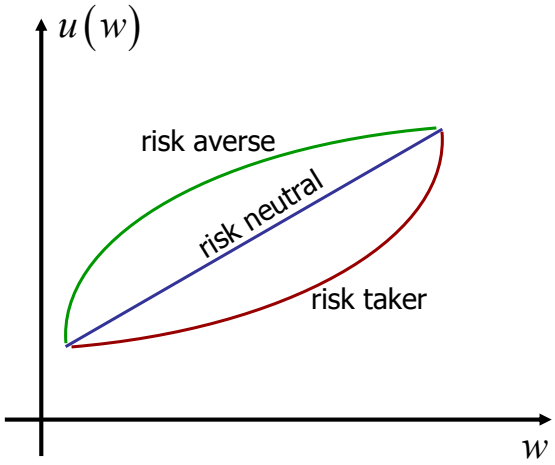


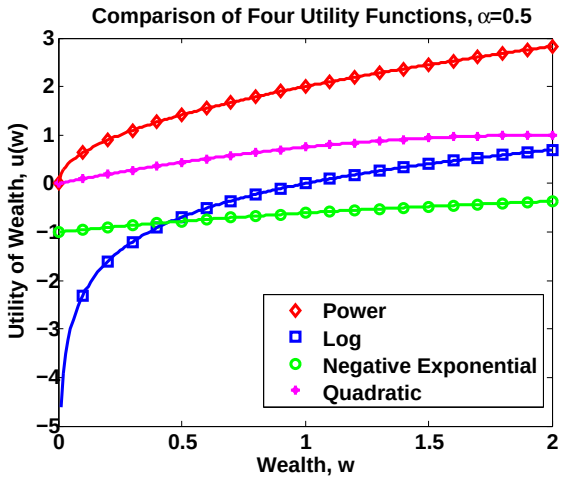
Figure 2-1. The investor's preferences towards risk may be inferred from the curvature of the utility function. A concave function indicates that the investor is a *risk averter*, a convex shape implies he is a *risk taker*, and a linear shape corresponds to a *risk neural* investor.

are monotonically increasing, but the shape of the function depends on the investor's risk preferences. A concave utility function indicates that the investor is a *risk averter*, a convex shape implies he is a *risk taker*, and a linear shape corresponds to a *risk neural* investor, as depicted in Figure 2-1.

The class of risk-averse utility functions can be further subdivided into two classes, based on the work of Pratt [49] and Arrow [4]. Let $R_A(w)$ and $R_R(w)$ denote the coefficients of absolute and relative risk aversion, respectively, defined as follows:

$$R_A(w) = -\frac{u''(w)}{u'(w)} \qquad R_R(w) = -\frac{u''(w)}{u'(w)}w = R_a(w)w.$$

The first sub-class contains all utility functions with constant absolute risk aversion (CARA). The canonical example of this class of risk averse functions is the negative exponential function, shown in the third panel of Fig. 2-2. The second sub-class contains all utility functions with constant relative risk aversion (CRRA), which includes the power and log utility functions shown in the top two panels of Fig. 2-2. Also of note is the quadratic utility function, shown in the fourth panel of Fig. 2-2, which is neither CRRA nor CARA, but is characterized by the fact that the marginal (incremental) utility, $u'(w)$, is linear in wealth. The plot in Fig. 2-2 illustrates the different curvatures of these utility functions for the case where the parameter α is 0.5.



Power Utility	Log Utility
$u(w) = \frac{w^{1-\alpha}}{1-\alpha}, \quad \alpha > 0, \ \alpha \neq 1$ $u'(w) = w^{-\alpha}$ $u''(w) = -\alpha w^{-\alpha-1}$ $R_A(w) = -\frac{u''(w)}{u'(w)} = \alpha w^{-1}$ $R_R(w) = -\frac{u''(w)}{u'(w)} w = \alpha$	$u(w) = \lim_{\alpha \to 1} \frac{w^{1-\alpha}}{1-\alpha} = \log(w),$ $u'(w) = w^{-1}$ $u''(w) = -w^{-2}$ $R_A(w) = -\frac{u''(w)}{u'(w)} = w^{-1}$ $R_R(w) = -\frac{u''(w)}{u'(w)} w = 1$
Negative Exponential Utility	**Quadratic Utility**
$u(w) = -\exp(-\alpha w), \quad \alpha > 0$ $u'(w) = \alpha \exp(-\alpha w)$ $u''(w) = -\alpha^2 \exp(-\alpha w)$ $R_A(w) = -\frac{u''(w)}{u'(w)} = \alpha$ $R_R(w) = -\frac{u''(w)}{u'(w)} w = \alpha w$	$u(w) = w - \frac{\alpha}{2} w^2, \quad w < \frac{1}{\alpha}$ $u'(w) = 1 - \alpha w$ $u''(w) = -\alpha$ $R_A(w) = -\frac{u''(w)}{u'(w)} = \frac{\alpha}{1-\alpha w}$ $R_R(w) = -\frac{u''(w)}{u'(w)} w = \frac{\alpha w}{1-\alpha w}$

Figure 2-2. Comparison of four risk averse utility functions.

2.2.3 Budget Constraint

The last component considered here for defining portfolio risk is the concept of a budget constraint. Such constraints are used to limit the degree of total market exposure assumed by an investor, by requiring that the total value of the portfolio equals the available wealth. This implies that the the investor *may not* arbitrarily borrow additional funds *for free* in

order to magnify (leverage) the effective realized returns. For example, suppose an investor has an initial wealth of \$100, and identifies an asset that earns a guaranteed 5% rate of return over a single period. If the investor were able to borrow an additional \$100 from a bank, he would have made a profit \$10, instead of \$5, making the effective realized return 10% instead of 5%. In the real world, of course, the investor would also need to pay interest on the money he borrowed, at an appropriate rate of interest, but it is clear from the example that leverage provides a powerful tool for increasing expected returns. The danger of leverage, however, is that it also increases the investor's exposure to market risk, making it possible to lose more money than the investor is worth. The ability to borrow capital can be combined with a budget constraint by explicitly including a risk-free asset in the model, so that the capital received from borrowing is accounted for in the computation of available wealth.

Formally, let $\mathbf{w} \in \mathbb{R}^p$ denote a portfolio weight vector of p assets, where w_i denotes the percentage of initial wealth invested in the i^{th} asset. The typical budget constraint is expressed as $\sum_{i=1}^{p} w_i = 1$, or $\mathbf{w}^T\mathbf{1} = 1$. In this setting, it is easy to enforce the budget constraint by applying an affine transformation to the portfolio weight vector, as derived next. Let $\mathbf{w}_{\text{c}}$ denote the constrained portfolio weight vector, and let $\mathbf{v} \in \mathbb{R}^{p-1}$, which are related as follows:

$$\mathbf{w}_{\text{c}} = \mathbf{c} + \mathbf{D}\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_{p-1} \\ 1 - \sum_{i=1}^{p-1} v_i \end{pmatrix}^T, \quad \mathbf{c} = \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}^T, \quad \mathbf{D} = \begin{pmatrix} \mathbb{I}_{p-1} \\ -\mathbf{1}^T \end{pmatrix}. \tag{2.7}$$

The optimal policy is then derived as a function of $\mathbf{v}$, which is subsequently mapped back into the higher dimensional space according to the rule $\mathbf{w}_{\text{c}}^* = \mathbf{c} + \mathbf{D}\mathbf{v}^*$. In a dynamic multistage setting, the investor is required to reinvest all of his *current wealth* at the beginning of each stage. In this case, the relationship $\mathbf{w}_{\text{c}} = \mathbf{c}(1 + r_k) + \mathbf{D}\mathbf{v}$ should be used, where r_k denotes the cumulative portfolio return at time t_k.

When short-selling is allowed (weights can be positive or negative), there are two possible ways in which budget constraints may be incorporated. The familiar budget constraint

$\mathbf{w}^T\mathbf{1} = 1$ is used when short-sales are viewed as a *source of income*. On the other hand, when short-sales are viewed as a *source of spending*, the Lintner budget constraint is used [37]. Here the ℓ_1 norm of the portfolio weight vector must equal unity:

$$||\mathbf{w}||_1 = \sum_{i=1}^{p} |w_i| = 1.$$

Omission of either budget constraint is equivalent to introducing a risk-free asset with rate $r_f = 0$. However, when a risk-free asset is explicitly included in the model, a budget constraint must be utilized in order to prevent the percent allocated to the risk-free asset from growing arbitrarily large.

2.3 Single-Stage Portfolio Choice

Having presented an overview of asset return models in Section 2.1 and definitions for portfolio risk in Section 2.2, two single-stage portfolio choice frameworks can now be described. In the first framework, based on the pioneering work of Markowitz, the optimal portfolio is defined as the one which maximizes the expected portfolio return, subject to a constraint on the associated portfolio return variance. In the second framework, the optimal portfolio is defined as the one that maximizes the expected value of the investor's utility of wealth.

2.3.1 Mean-Variance Optimization

As discussed in Section 2.2.1, the Markowitz formulation assumes that investors act to optimize the tradeoff between portfolio return risk (variance) and reward (mean), within a restricted universe (finite set) of tradable assets. The asset allocation scheme is chosen in order to maximize the expected return of the portfolio, subject to a variance constraint, or equivalently, to minimize the variance of the portfolio subject to a constraint on the desired portfolio return. It is also common to include a budget constraint, as discussed in Section 2.2.3. Such a formulation leads to the production of an *efficient frontier*, which is a functional representation of the expected return realized for each level of portfolio standard deviation, as described here.

Let $\mathbf{r} \in \mathbb{R}^p$ denote a random vector of returns corresponding to each of the p assets available to the investor, with known mean, $\boldsymbol{\mu}$, and covariance matrix, $\boldsymbol{\Sigma}$. Given this information,

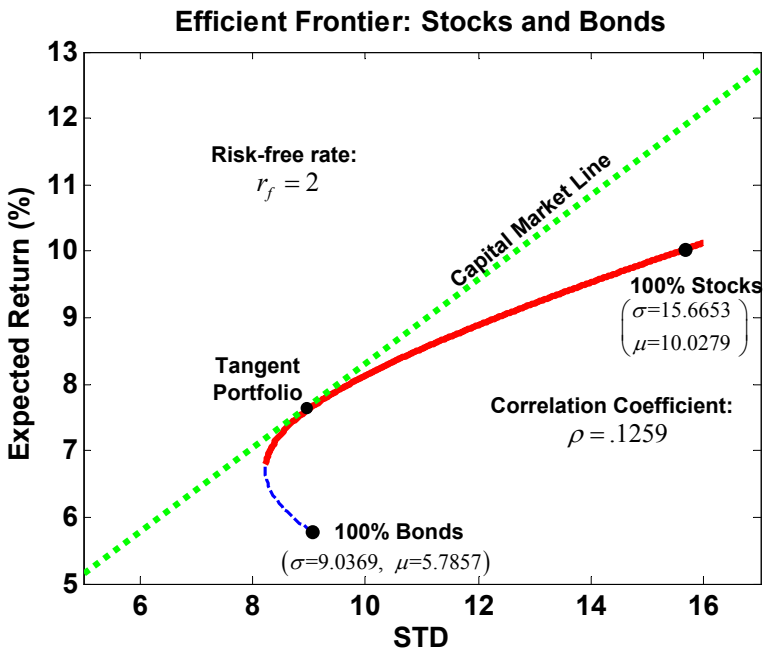


Figure 2-3. Mean-Variance Efficient Frontier and Capital Market Line for a financial universe consisting of stocks ($S\&P$ 500 Composite Index) and bonds (10 Year U.S. Government Treasury Bond) and a risk-free asset with rate $r_f = 2\%$. Returns are given in annual percentages.

the Markowitz mean-variance investor seeks the optimal portfolio $\mathbf{w}^*$, so that:

$$\begin{aligned} \mathbf{w}^* = \arg\max_{\mathbf{w}} \quad & \mathbf{w}^T \boldsymbol{\mu} \\ \text{subject to} \quad & \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} = \sigma_0^2 \\ & \mathbf{w}^T \mathbf{1} = 1, \end{aligned}$$

where the variance σ_0^2 defines the allowable "risk-budget", set by the investor. It is important to note that the Markowitz framework does not assume that the return distributions are jointly Gaussian; it simply states that the investor makes decisions based only on the first and second moments of the return distribution. In cases where these returns are Normally distributed, the first and second moments completely characterize the full probability distribution of the return. The mean-variance asset allocation scheme is illustrated using a portfolio comprised of two risky assets in the following example.

Example 2.1.

Suppose an investor must decide how to allocate his wealth between two risky assets, corresponding to the $S\&P$ 500 Composite index and the 10 Year U.S. Treasury Bond[4], with

[4]Here, the U.S. Treasury bond is being used as a risky asset, as compared with the discussion in Section

single-period returns r_S and r_B, respectively. Through the analysis of past daily returns from 1925 to 2000 obtained via the CRSP database (Center for Research in Security Prices), the first and second order statistics for the annual return over the next year were estimated to be:

$$\boldsymbol{\mu} = \begin{pmatrix} E[r_S] \\ E[r_B] \end{pmatrix} = \begin{pmatrix} 10.03 \\ 5.79 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \text{var}[r_A] & \text{cov}[r_A, r_B] \\ \text{cov}[r_A, r_B] & \text{var}[r_B] \end{pmatrix} = \begin{pmatrix} 245.40 & 17.83 \\ 17.83 & 81.67 \end{pmatrix}.$$

As evidenced by the non-diagonal structure of the covariance matrix $\boldsymbol{\Sigma}$, the two assets are correlated, with a correlation coefficient of $\rho = 0.13$. The set of efficient (optimal) portfolios that result for this example as the allowable risk parameter varies is shown by the red curve (solid line) in Figure 2-3. The blue curve (dashed line) highlights portfolios that do not maximize return for a given level of risk, but rather minimize risk for a given level of return.

When a risk-free asset is also available, such as a 30-day U.S. government T-bill, the investor can achieve a higher expected return for the same level of risk by employing a strategy of investing in both the market portfolio and the risk-free asset, as described by James Tobin in [58]. This "Two Fund Separation Theorem" leads to the creation of the *capital market line* (CML), the line tangent to the efficient frontier that intercepts the y-axis at the risk-free rate, r_f. For the example given here, a representative CML is shown in green (dotted line) in Figure 2-3, and the tangent portfolio is highlighted. If the investor is allowed to be leveraged, he may borrow money from the risk-free asset in order to achieve the set of operating points on the CML to the right of the tangent portfolio. The slope of the CML is referred to as the Sharpe ratio [55], defined as:

$$S = \frac{E[r_p] - r_f}{\sigma_p}, \tag{2.8}$$

where r_p denotes the portfolio return and σ_p denotes the corresponding portfolio return standard deviation. The Sharpe ratio measures the amount of return a strategy can earn in excess of the risk-free asset, per unit risk. Note that removal of the budget constraint implies the investor can borrow for free (i.e., $r_f = 0$), resulting in an efficient frontier that is

2.2, where the U.S. Treasury T-Bill was cited as a source of a risk-free asset. The difference here is that due to the long time horizon of 10 years, the face value of the bond within the fixed income market fluctuates day to day, and is therefore considered risky.

a straight line through the origin. Without a budget constraint, only the variance constraint determines the *size (i.e., scale)* of the portfolio. ■

Another interesting point of view regarding the Markowitz mean-variance solution is that it also satisfies Roy's Safety-First portfolio choice criteria [51], in which the investor acts so as to minimize the probability a disaster occurs. Here disaster is defined as the event that the total return is less than a pre-defined threshold, d, and both the return mean, m, and standard deviation, σ, are known to the investor. According to Chebyshev's inequality, the probability that the realized return, r, is less than the desired threshold d is bounded above by:

$$P(r \leq d) = P(r - m \leq d - m) \leq P(|m - r| \geq m - d) \leq \frac{\sigma^2}{(m-d)^2}.$$

Minimizing $P(r \leq d)$ is equivalent to maximizing $\frac{m-d}{\sigma}$, which in turn, is equivalent to Markowitz's mean-variance criteria.

2.3.2 Expected Utility Maximization

Recall from Section 2.2.2 that in addition to using return variance in order to define risk, it is also possible to encode an investor's attitudes towards risk and reward in a mathematical mapping known as a utility function. In this case, the investment decision is chosen to maximize the expected utility of the investor's portfolio return at the end of the investment horizon. This framework has its foundations in the early work of von Neumann and Morgenstern [62] on game theory, and was first applied to portfolio selection in 1958 by Tobin [58].

Formally, an investor with a total return preference function u, seeks the optimal portfolio $\mathbf{w}^*$, so that:

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \quad E\left[u\left(\mathbf{w}^T\mathbf{r}\right)\right],$$

where $\mathbf{r}$ again denotes a random vector of individual asset returns. Depending on the choice of utility function, a closed-form solution may exist or numerical methods may be required.

There are three scenarios concerning utility functions and return distributions in which the resulting portfolio weights are equivalent to the Markowitz mean-variance formulation

[20], as listed here:

- Quadratic utility, with no additional assumption on asset returns,
- Exponential utility (from CARA family), with normally distributed asset returns,
- Power utility (from CRRA family), with log-normally distributed asset returns.

While not explored further here, the relationship between these utility-theoretic portfolio choice problems and the Markowitz mean-variance approach is exploited in Chapter 5 in order to solve the dynamic, multiperiod mean-variance problem.

2.4 Multistage Portfolio Choice

The foundations for a theory of multiperiod portfolio selection can be traced back to the 1962 work of Tobin [59], where he states that "...the portfolio sequence or strategy that promises the highest return for one date is not the same one that promises the highest return for another date," and rebalancing of the portfolio may be desired so long as the "...new portfolio must promise enough advantage in return to compensate for these [*transaction*] costs." In this section, multistage extensions to the mean-variance and expected utility maximization portfolio choice frameworks are described. In addition, brief discussions of portfolio policy classifications and transaction costs are presented in Sections 2.4.3 and 2.4.4, respectively.

2.4.1 Mean-Variance Optimization

In a multistage setting, there are two extensions to the mean-variance portfolio choice problem often considered in the literature. In the first, a constraint is placed on the per-stage variances of the portfolio, while in the second, a constraint is placed on the variance of the total, cumulative return. In cases where the per-stage returns are independent, these two formulations are equivalent. However, if the portfolio returns are serially correlated, the latter generalization can achieve higher expected returns for a given level of risk by exploiting inter-stage return dependence.

Let $\mathbf{w}_k$ denote the portfolio weight vector at time t_k, $\mathbf{r}_k$ denote the stacked log-returns

of a set of p assets at stage k, and the scalar r_k denote the stage k portfolio return. The total portfolio log-return, r_T, across N stages, is computed as:

$$r_T = \sum_{i=1}^{N} r_i = \sum_{i=0}^{N-1} \mathbf{w}_i^T \mathbf{r}_{i+1}.$$

The investor seeks the optimal sequence of portfolios, $\left\{\mathbf{w}_0^*, ..., \mathbf{w}_{N-1}^*\right\}$, by solving either Problem M_1 or Problem M_2, as described below.

Case 1: Additive Mean-Variance Framework One multistage extension to the Markowitz mean-variance framework that does not consider the inter-stage return correlation structure is given by Problem M_1, as follows:

$$\left.\begin{array}{rl} \left\{\mathbf{w}_0^*, ..., \mathbf{w}_{N-1}^*\right\} = \underset{\{\mathbf{w}_0,...,\mathbf{w}_{N-1}\}}{\arg\max} & E\left[\sum_{i=0}^{N-1} \mathbf{w}_i^T \mathbf{r}_{i+1}\right] \\ \text{subject to} & \sum_{i=0}^{N-1} \text{var}\left[\mathbf{w}_i^T \mathbf{r}_{i+1}\right] = \sigma_0^2 \end{array}\right\} M_1,$$

where σ_0^2 represents the total allowable "risk-budget". The portfolio weight vectors over all but the first stage, $\{\mathbf{w}_1, \cdots, \mathbf{w}_{N-1}\}$, are each state dependent random variables. The portfolio at time t_k is a *function* of the observed value of the asset log-prices at the beginning of each stage, i.e. $\mathbf{w}_k = f_k(\mathbf{x}_k)$. While Problem M_1 ignores non-zero correlations among the inter-stage returns, the corresponding objective function is additive in the number of stages, and therefore dynamic programming techniques can yield computationally efficient solutions in order to determine the optimal sequence of portfolio policy functions.

Case 2: Non-additive Mean-Variance Framework The second multistage extension to the Markowitz mean-variance framework does take into account the inter-stage return statistics, and is given by Problem M_2, as follows:

$$\left.\begin{array}{rl} \left\{\mathbf{w}_0^*, ..., \mathbf{w}_{N-1}^*\right\} = \underset{\{\mathbf{w}_0,...,\mathbf{w}_{N-1}\}}{\arg\max} & E\left[\sum_{i=0}^{N-1} \mathbf{w}_i^T \mathbf{r}_{i+1}\right] \\ \text{subject to} & \text{var}\left[\sum_{i=0}^{N-1} \mathbf{w}_i^T \mathbf{r}_{i+1}\right] = \sigma_0^2 \end{array}\right\} M_2,$$

where:

$$\text{var}_{t_0}\left[\sum_{i=0}^{N-1} \mathbf{w}_i^T \mathbf{r}_{i+1}\right] = \sum_{i=0}^{N-1} \text{var}_{t_0}\left[\mathbf{w}_i^T \mathbf{r}_{i+1}\right] + \sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} \text{cov}_{t_0}\left[\mathbf{w}_i^T \mathbf{r}_{i+1}, \mathbf{w}_j^T \mathbf{r}_{j+1}\right].$$

The trading strategy produced by Problem M_2 may be preferable over the trading strategy corresponding to Problem M_1 when the inter-stage portfolio return are negatively correlated. Such a correlation structure enables a greater degree of per-stage variance to be taken, while maintaining the same level of total variance. This enables the investor to utilize leverage to earn a higher expected total portfolio return, a theme that is explored in great detail throughout this thesis.

In contrast to Problem M_1, Problem M_2 does not admit an additive cost function due to the inclusion of covariance terms, and therefore dynamic programming techniques cannot be naively applied in order to learn the optimal sequence of portfolio policy functions. In addition, it does not fall within the class of expected utility maximization problems, due to the presence of the variance constraint. Specifically, the variance term may be expanded out as follows:

$$\text{var}_{t_0}\left[\sum_{k=0}^{N-1}\mathbf{w}_k^T\mathbf{r}_{k+1}\right] = E\left[\left(\sum_{k=0}^{N-1}\mathbf{w}_k^T\mathbf{r}_{k+1}\right)^2\right] - E\left[\sum_{k=0}^{N-1}\mathbf{w}_k^T\mathbf{r}_{k+1}\right]^2.$$

The squared expectation operator prevents this term from being expressed as the expected value of a function of terminal wealth. However, as discussed in Chapter 5, it is possible to map Problem M_2 into a related, auxiliary problem using a quadratic utility function in order to learn the optimal sequence of MVO portfolios [36].

2.4.2 Expected Utility Maximization

Extensions of the expected utility maximization framework to multiple stages can be traced to the 1968 work of Mossin [45], in which it is stated that "any sequence of portfolio decisions must be contingent upon the outcomes of previous periods and at the same time take into account information on future probability distributions." Mossin was the first to propose that investors act to maximize the expected utility of terminal wealth, expressed formally as:

$$\left\{\mathbf{w}_0^*, ..., \mathbf{w}_{N-1}^*\right\} = \underset{\{\mathbf{w}_0,...,\mathbf{w}_{N-1}\}}{\arg\max} \quad E\left[u\left(r_T\right)\right] = \underset{\{\mathbf{w}_0,...,\mathbf{w}_{N-1}\}}{\arg\max} \quad E\left[u\left(\sum_{i=0}^{N-1}\mathbf{w}_i^T\mathbf{r}_{i+1}\right)\right].$$

In this work, Mossin also proposed the use of dynamic programming methods to solve such optimization problems. In the following year, Samuelson and Merton expanded upon Mossin's work [44, 52], by including the expected utility of consumption in the portfolio choice framework, working in discrete and continuous-time, respectively.

2.4.3 Policy Classification

Once the multistage portfolio selection problem has been solved, the resulting policy can be characterized as belonging to one of three general classes. Under certain conditions, the solution to the multiperiod problem may reduce to solving a sequence of independent single-stage problems, in which case the investor is said to follow a *myopic* policy [13]. This situation typically arises when one assumes that the asset returns are independent and identically distributed (i.i.d.) in time. When the return distributions are allowed to vary at each stage, but are known in advance, the resulting solution is referred to as a *pre-commitment* policy, and the portfolio is rebalanced according to a deterministic schedule [7]. A truly dynamic portfolio selection problem arises only when the return distributions change over time in a stochastic manner, and the new information that becomes available at the beginning of each stage is taken into consideration before an investment decision is made. Policies generated in the presence of return distribution uncertainty typically have one component corresponding to the original *myopic* policy, and a second component that *hedges* against future investment uncertainty [47].

2.4.4 Transaction Costs

There are many forms of transaction costs that penalize the investor for executing a given portfolio policy. Possible sources of transaction costs include broker or exchange commissions and fees, bid/ask spreads, price fluctuations (timing cost), and the price impact of trading (liquidity cost) [64]. While this is an important and interesting topic from both academic and practical perspectives, transactions costs are not considered in this thesis.

Chapter 3

Cointegration

This chapter presents an overview of cointegrated vector autoregressive (VAR) processes using a state-space approach, expanding on the development of the VAR model found in Appendix A. A single definition of the *cointegration property* is given, which is stated independent of any specific random process describing the time evolution of the constituent signals. Given this definition, two equivalent sets of restrictions on the parameters of a VAR process are given, each of which is shown to produce a cointegrated system.

In contrast to other cointegration primers found in the literature, special attention is placed here on understanding the geometry of the resulting system, a theme that is continued in the subsequent development of mean-variance optimal portfolios. Of particular interest is the distribution of cointegrated data over time, which clusters around a reduced-dimension subspace of $\mathbb{R}^p$, the space spanned by the asset log-prices vectors. In addition, the subspaces associated with the short and long-term dynamics are identified and characterized in terms of the parameters of a cointegrated VAR system.

The organization of this chapter is as follows. In Section 3.1, a set of definitions regarding stability and integrated processes is presented. In Section 3.2, the error-correcting and state-space forms for a cointegrated VAR process are discussed, and their equivalence is established. Building on the state-space approach, the total system response is given in Section 3.3, resulting in an alternative form of the Granger Representation Theorem [29]. In Section 3.4, a formal estimation procedure for fitting a cointegration VAR model to a data set is detailed. Lastly, in Section 3.5, a brief survey of the existing literature on cointegration in portfolio theory and econometrics is given.

3.1 Preliminary Notions

Consider a scalar, discrete-time, linear time-invariant (LTI) system with the following autoregressive (AR) input-output relationship:

$$y[n] = \sum_{k=1}^{p} a_k y[n-k] + x[n], \tag{3.1}$$

for some $a_1, ..., a_p \in \mathbb{R}$. Letting $X(z)$ and $Y(z)$ denote the $\mathcal{Z}$-transforms of the signals $x[n]$ and $y[n]$, respectively, the system transfer function, $H(z)$, is given by:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}. \tag{3.2}$$

Assuming the system is causal (i.e., the output does not depend on future inputs), the unit impulse response is computed as follows:

$$h[n] = \mathcal{Z}^{-1}\left(H(z)\right) = \sum_{k=1}^{p} A_k \lambda_k^n s[n], \tag{3.3}$$

where $A_1, ..., A_p \in \mathbb{R}$, $s[n]$ is the unit-step function, and $\{\lambda_1, ..., \lambda_p\}$ denote the p roots of the polynomial in the denominator of Eq. 3.2, which are also equal to the *poles* of the system. An LTI system is *bounded-input, bounded-output (BIBO) stable* if the following condition holds:

$$\sum_{k=-\infty}^{\infty} |h[k]| < \infty.$$

For the specific system defined by Eq. 3.3, BIBO stability implies that $|\lambda_k| < 1,\ k = 1, ..., p$.

A random process $x[n]$ is *wide-sense stationary* (WSS) if both the mean and the variance of the process remain constant over time, and the autocorrelation function, $E[x[n]x[m]]$, only depends on the time difference, $|n-m|$, not the absolute times. A process that starts at time $n_0 = 0$ is *asymptotically WSS* (AWSS) if any initial condition response decays to zero as $n \to \infty$ *and* the process' statistics approach those of a WSS process [67]. A process that is WSS is also AWSS. For example, let $x[n]$ denote an independent and identically distributed (i.i.d.) Gaussian white noise process with zero mean and variance σ^2, and let $y[n]$ denote the output process produced by putting $x[n]$ through the system defined by Eq. 3.1 with

$p = 1$, $|a_1| < 1$, and arbitrary initial condition $y[0]$. Then, $x[n]$ is WSS, and $y[n]$ is AWSS, where the limiting stationary distribution has zero mean, variance $\frac{\sigma^2}{1-a_1^2}$, and autocorrelation function $E[x[n]x[n-k]] = a_1^{|k|}\sigma^2$. In summary, putting a WSS random process through a BIBO stable LTI system beginning at time $n_0 = 0$ produces an output process that is AWSS.

For the system defined by Eq. 3.1, if there is a single pole at unity, $\lambda_1 = 1$, while the remaining poles all satisfy $|\lambda_k| < 1$ for $k = 2, ..., p$, the system is said to be *marginally unstable*. When a WSS random process is passed through such a marginally unstable system, the output process is said to be *integrated of order one*, denoted as $I(1)$[1]. Integrated processes are not AWSS, as the process variance grows linearly with time. Such $I(1)$ processes have the property that the corresponding first-difference series, $\Delta x[n] = x[n] - x[n-1]$, is AWSS. A simple example of an $I(1)$ process is a *random walk*, which is the sum of i.i.d. random variables. Formally, the process evolves according to:

$$x[n] = x[n-1] + \epsilon[n] = \sum_{i=0}^{n} \epsilon[i],$$

with initial condition $x[0]$ and where the $\epsilon[n]$ are i.i.d. random variables.

Let $x_1[n]$ and $x_2[n]$ denote two scalar, discrete-time signals, each of which belongs to the class of $I(1)$ processes. For arbitrary $a_1, a_2 \in \mathbb{R}$ the linear combination $y[n] = a_1x_1[n] + a_2x_2[n]$ is also $I(1)$. However, in some cases, it is possible to choose a_1 and a_2 so that $y[n]$ is AWSS, thus removing the effect of a shared integrator, also referred to as a *common trend* or *random walk* component. When this occurs, x_1 and x_2 are said to be *co-integrated*. While the general notions of integrated and co-integrated time series apply to processes with an arbitrary number of integrators, attention here is restricted to the case where the underlying processes are all assumed to be $I(1)$, as many real-world financial time series have been empirically shown to fall within this class [12, 26].

[1]Even in a discrete-time setting, the name *integrator* is commonly used in place of the name *accumulator*.

3.2 Cointegrated Vector Autoregressive Processes

As defined in the previous section, a group of random processes are said to be cointegrated if each is well-modeled by an integrated process, yet there exists a linear combination of them that is AWSS. Here, a set of restrictions are imposed on the parameters of a vector autoregressive (VAR) process so that the underlying component signals are cointegrated. The conditions are given for both the error-correcting and state-space representations of a VAR process, in Sections 3.2.1 and 3.2.2, respectively. Furthermore, the equivalence of these restrictions is established.

3.2.1 Form 1: Error-Correction Model

Let $\mathbf{x}[n] \in \mathbb{R}^p$ denote a random vector in discrete time, indexed by n, which evolves according to an order k vector autoregressive, $VAR(k)$, process. As stated in Appendix A, the canonical form for such a process is given by:

$$\mathbf{x}[n] = \mathbf{\Pi}_1\mathbf{x}[n-1] + \mathbf{\Pi}_2\mathbf{x}[n-2] + \cdots + \mathbf{\Pi}_k\mathbf{x}[n-k] + \mathbf{\Phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n], \tag{3.4}$$

where:

- The $\mathbf{\Pi}_i \in \mathbb{R}^{p\times p}$ are matrices of coupling coefficients,
- $\mathbf{d} \in \mathbb{R}^{r\times 1}$ is a vector of deterministic inputs,
- $\mathbf{\Phi} \in \mathbb{R}^{p\times r}$ matrix of coefficients relating the deterministic terms to the elements of $\mathbf{x}$,
- $\boldsymbol{\epsilon} \in \mathbb{R}^{p\times 1}$ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{\Psi}$.

The input-output relationship may also be expressed in *error-correcting model* (ECM) form, as follows:

$$\Delta\mathbf{x}[n] = \mathbf{\Pi}\mathbf{x}[n-1] + \sum_{i=1}^{k-1}\mathbf{\Gamma}_i\Delta\mathbf{x}[n-i] + \mathbf{\Phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n] \tag{3.5}$$

where:

$$\mathbf{\Pi} \triangleq \sum_{i=1}^{k}\mathbf{\Pi}_i - \mathbb{I}_p, \tag{3.6}$$

$$\Delta\mathbf{x}[n-i] \triangleq \mathbf{x}[n-i] - \mathbf{x}[n-i-1]$$

$$\mathbf{\Gamma}_i \triangleq -\sum_{j=i+1}^{k}\mathbf{\Pi}_j \tag{3.7}$$

This form of a VAR process relates the current change in the output not only to the set of $k-1$ previous output changes, but also to the current log-price level (i.e., $\mathbf{x}[n-1]$) through the matrix $\mathbf{\Pi}$.

A VAR process exhibits the cointegration property if the matrix $\mathbf{\Pi}$ has reduced rank $r < p$, as stated in Restriction 3.1 below.

Restriction 3.1.

Let the vector random process $\mathbf{x}[n]$ *evolve according to a* $VAR(k)$ *model defined in the error-correcting model form of Eq. 3.5. Furthermore, assume that each of the component processes of* $\mathbf{x}[n]$ *is* $I(1)$*. The system is said to be* cointegrated *if the matrix* $\mathbf{\Pi}$*, defined by Eq. 3.6 has reduced rank* $0 < r < p$*, and can therefore be factored as the outer-product of two* $p \times r$ *matrices, as follows:*

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^T. \tag{3.8}$$

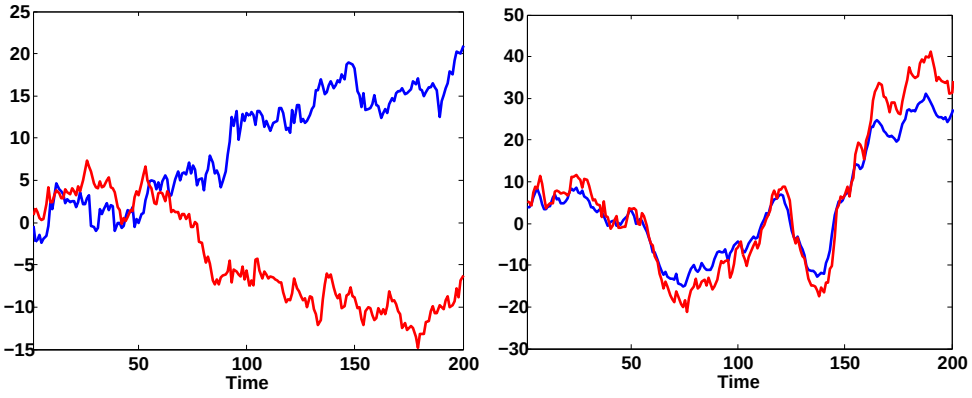
Given the reduced rank matrix $\mathbf{\Pi}$, the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are computed using the singular value decomposition of $\mathbf{\Pi}$, as described in Appendix 3.B.

In order to show that this restriction on a VAR system implies the component processes are cointegrated, substitute Eq. 3.8 into Eq. 3.5, as follows:

$$\Delta\mathbf{x}[n] = \boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{x}[n-1] + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta\mathbf{x}[n-i] + \mathbf{u}[n], \tag{3.9}$$

where the total system input $\mathbf{u}[n] = \boldsymbol{\phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n]$ is AWSS, implying that the deterministic component is zero or constant. Under the assumption that each component process is $I(1)$, the process on the left side, $\Delta\mathbf{x}[n]$, is AWSS, and therefore the process on the right-hand side must also be AWSS. The middle term, $\sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta\mathbf{x}[n-i]$, is AWSS as it is the sum of lagged first-differences, and the third term, $\mathbf{u}[n]$, is assumed to be AWSS. Therefore, the first term, $\boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{x}[n-1]$, must also be AWSS. This implies that $\boldsymbol{\beta}^T\mathbf{x}[n-1]$ is AWSS, and that each column of the matrix $\boldsymbol{\beta}$ is a cointegrating vector.

The following two examples illustrate the use Restriction 3.1.



(a) VAR(1) system from Ex. 3.1, comprised of two independent random walks that are not cointegrated.

(b) VAR(1) system from Ex. 3.2, comprised of two cointegrated time-series.

Figure 3-1. Sample paths of non-cointegrated and cointegrated VAR systems.

Example 3.1.

Consider the following VAR(1) model, with no deterministic inputs:

$$\mathbf{x}[n] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x}[n-1] + \boldsymbol{\epsilon}[n],$$

where $\boldsymbol{\epsilon}[n] \sim N(\mathbf{0}, \boldsymbol{\Psi})$ and $\boldsymbol{\Psi} = \mathbb{I}_2$, which denotes the 2×2 identity matrix. The matrix $\boldsymbol{\Pi} = \boldsymbol{\Pi}_1 - \mathbb{I}_2 = \mathbf{0}$ has rank 0, and cannot be factored according to Eq. 3.8. The underlying processes correspond to two independent random walks, as depicted in Fig. 3-1(a), and are not cointegrated. ■

Example 3.2.

Now consider a second $VAR(1)$ model, again with no deterministic inputs:

$$\mathbf{x}[n] = \begin{pmatrix} 1.18 & -0.14 \\ 0.51 & 0.62 \end{pmatrix} \mathbf{x}[n-1] + \boldsymbol{\epsilon}[n],$$

where $\boldsymbol{\epsilon}[n] \sim N\left(\mathbf{0}, 10^{-3}\mathbb{I}\right)$. Here, the matrix $\boldsymbol{\Pi}$ has rank 1, and factors according to:

$$\boldsymbol{\Pi} = \boldsymbol{\Pi}_1 - \mathbb{I}_2 = \begin{pmatrix} .18 & -0.14 \\ -0.49 & 0.62 \end{pmatrix} = \begin{pmatrix} -0.28 \\ -0.77 \end{pmatrix} \begin{pmatrix} -0.66 & 0.5 \end{pmatrix}.$$

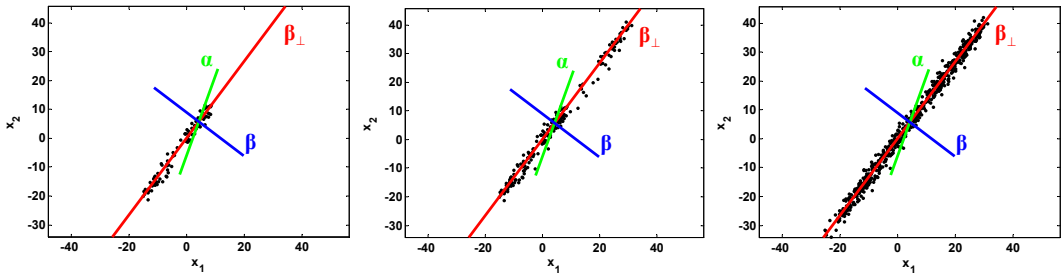
The ECM form is given by:

$$\Delta\mathbf{x}[n] = \boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{x}[n-1] + \boldsymbol{\epsilon}[n] = \begin{pmatrix} -0.28 \\ -0.77 \end{pmatrix} \begin{pmatrix} -0.66 & 0.5 \end{pmatrix} \mathbf{x}[n-1] + \boldsymbol{\epsilon}[n]. \tag{3.10}$$

Figure 3-1(b) depicts a sample path of this process of length 200. The two component signals share a long-term stochastic "common-trend" and move together in a coordinated fashion, yet deviate from each other by an amount that is well-modeled by a random process that is asymptotically WSS (AWSS). ■

After reading these two examples, readers may find it helpful to recall the analogy of the "drunk walking a dog" [46]. Just as a sample path of a random walk is often likened to a drunkard's walk, one may think of cointegration as what happens when the drunk takes a puppy along on the stroll. The two time series meander about in a seemingly random pattern, however, never deviate from each other "too much" due to the presence of the "leash", or rather *error-correction* forces that work to restore them to their long-run equilibrium behavior. The drunk and the dog share a single random walk (common trend), and each signal deviates from this shared trend by an amount well-modeled by a random process that is AWSS.

The system in Ex. 3.2 can also be used to study the geometry of a cointegrated system, revealed in a scatter plot of the component processes. Figure 3-2 displays three views of data generated from this model after $N = \{100, 200, 600\}$ steps, respectively, with arbitrary initial condition $\mathbf{x}[0] = \begin{pmatrix} 3.9 & 5.5 \end{pmatrix}^T$. As the figure indicates, the data appears to have finite variance in the direction of $\boldsymbol{\beta}$ and growing variance in the direction of $\boldsymbol{\beta}_\perp$. The space spanned by $\boldsymbol{\beta}_\perp$ represents the long-term or steady-state equilibrium relationship between the variables x_1 and x_2, and is often referred to as the *attractor* or *equilibrium* space. The scalar process $\boldsymbol{\beta}^T\mathbf{x}[n]$ measures *how far out of equilibrium* the process is at time n, measured by the distance of the vector $\mathbf{x}[n]$ from a hyperplane through the origin with normal vector $\boldsymbol{\beta}$. The space spanned by $\boldsymbol{\beta}$ is referred to as the *cointegrating* space, as any vector in this space is a cointegrating vector. The change to the process at each time step is determined by measuring the disequilibrium error, $\boldsymbol{\beta}^T\mathbf{x}[n]$, which is "corrected for" through the restoring "forces" in the $\boldsymbol{\alpha}$ direction, as described by Eq. 3.10. It is for this reason that the spaced spanned by $\boldsymbol{\alpha}$ is referred to as the *error-correcting* or *disequilibrium readjustment*



(a) $N = 100$ (b) $N = 200$ (c) $N = 600$

Figure 3-2. Scatter plot of cointegrated VAR system over time. The data has finite variance in the direction of $\boldsymbol{\beta}$ and growing variance in the direction of $\boldsymbol{\beta}_\perp$.

Subspace	Dimension	Common Names
$\boldsymbol{\alpha}$	r	*error-correcting space, disequilibrium readjustment space*
$\boldsymbol{\beta}$	r	*cointegrating space*
$\boldsymbol{\beta}_\perp$	p-r	*attractor space, equillibrium space*

Table 3.1. Subspaces of a cointegrated VAR system.

space. A summary of the subspaces defining the geometry of a cointegrated VAR system is given in Table 3.1.

3.2.2 Form 2: State-Space Model

As presented in Appendix A, a vector autoregressive process can be expressed in state-space form, as follows:

$$\underbrace{\begin{pmatrix} \mathbf{x}[n] \\ \mathbf{x}[n-1] \\ \vdots \\ \mathbf{x}[n-(k-1)] \end{pmatrix}}_{\mathbf{q}[n+1]} = \underbrace{\begin{pmatrix} \boldsymbol{\Pi}_1 & \cdots & \boldsymbol{\Pi}_{k-1} & \boldsymbol{\Pi}_k \\ \mathbb{I}_p & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbb{I}_p & \mathbf{0} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} \mathbf{x}[n-1] \\ \mathbf{x}[n-2] \\ \vdots \\ \mathbf{x}[n-k] \end{pmatrix}}_{\mathbf{q}[n]} + \underbrace{\begin{pmatrix} \mathbb{I}_p \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}}_{\mathbf{b}} \underbrace{\Big(\boldsymbol{\Phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n]\Big)}_{\mathbf{u}_n}$$

$$\underbrace{\mathbf{x}[n]}_{\mathbf{y}[n]} = \underbrace{\begin{pmatrix} \boldsymbol{\Pi}_1 & \cdots & \boldsymbol{\Pi}_{k-1} & \boldsymbol{\Pi}_k \end{pmatrix}}_{\mathbf{c}^T} \underbrace{\begin{pmatrix} \mathbf{x}[n-1] \\ \mathbf{x}[n-2] \\ \vdots \\ \mathbf{x}[n-k] \end{pmatrix}}_{\mathbf{q}_n} + \underbrace{\begin{pmatrix} \mathbb{I}_p \end{pmatrix}}_{\mathbf{d}[n]} \underbrace{\Big(\boldsymbol{\Phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n]\Big)}_{\mathbf{u}[n]} \quad (3.11)$$

where $\mathbf{A} \in \mathbb{R}^{pk \times pk}$, $\mathbf{b} \in \mathbb{R}^{pk \times p}$, $\mathbf{c}^T \in \mathbb{R}^{p \times pk}$, $\mathbf{d} \in \mathbb{R}^{p \times p}$, $\mathbf{u} \in \mathbb{R}^{p \times p}$, $\mathbf{q} \in \mathbb{R}^{pk \times 1}$, and $\mathbb{I}_p$ is the $p \times p$ identity matrix. Given this representation, the cointegration property can be expressed in terms of a simple restriction on the eigendecomposition of the state transition matrix, $\mathbf{A}$, as stated below.

Restriction 3.2.

Let the vector random process $\mathbf{x}[n]$ *evolve according to a VAR(k) model, represented in state-space form by Eq. 3.11. Furthermore, assume that the state transition matrix,* $\mathbf{A}$*, is diagonalizable, and admits an eigendecomposition of the form* $\mathbf{A} = \mathbf{M}\mathbf{D}\mathbf{M}^{-1}$*, where* $\mathbf{D}$ *is a diagonal matrix containing the eigenvalues of* $\mathbf{A}$*, and* $\mathbf{M}$ *is a matrix whose columns consist of the corresponding eigenvectors, which are all linearly independent, but not necessarily orthogonal. Then, the vector process is* cointegrated *if there is at least one eigenvalue satisfying* $|\lambda_i| < 1$*,* and *at least one eigenvalue satisfying* $\lambda_i = 1$*. Eigenvalues with* $|\lambda_i| > 1$ *and* $|\lambda_i| = 1$ *but* $\lambda_i \neq 1$ *are not considered.*

In order to establish that Restriction 3.2 implies the component processes are cointegrated, one must examine the total system response, often referred to as the *Granger Representation*. While the full derivation of the Granger Representation is deferred until Section 3.3, for now it is sufficient to state that this representation utilizes the eigenvalue assumption in order to express the process $\mathbf{x}[n]$ as a linear combination of integrated and AWSS random processes, denoted by $\mathbf{x}_1$ and $\mathbf{x}_2$ respectively, as follows:

$$\mathbf{x}[n] = \mathbf{C}_1\mathbf{x}_1[n] + \mathbf{C}_2\mathbf{x}_2[n], \tag{3.12}$$

where $\mathbf{C}_1, \mathbf{C}_2 \in \mathbb{R}^{p \times p}$. Furthermore, in Section 3.3.2 it is shown that:

$$\mathbf{C}_1 = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}_\perp^T \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}_\perp^T.$$

This implies that the scalar random process formed by taking the inner product of both sides of Eq. 3.12 with any vector $\mathbf{b} \in \text{span}\{\boldsymbol{\beta}\}$ is AWSS, since the term corresponding to the integrated component drops out, leaving:

$$\mathbf{b}^T\mathbf{x}[n] = \mathbf{b}^T\mathbf{C}_2\mathbf{x}_2[n].$$

Restriction 3.2 for a cointegrated process bears strong resemblance to the condition in Thm. A.1 for an AWSS VAR process. Only the existence of at least one eigenvalue at unity differentiates an AWSS VAR process from a cointegrated VAR process. In addition to containing at least one unit eigenvalue, it is also important that the process has at least one eigenvalue satisfying $|\lambda_i| < 1$. Recall from Ex. 3.1 that a system with $\mathbf{A} = \mathbf{\Pi}_1 = \mathbb{I}_2$, which contains 2 unit eigenvalues, generates two independent random walks, and is, therefore, not cointegrated.

The matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be expressed not only in terms of the SVD of $\mathbf{\Pi}$ as described in Appendix 3.B, but also in terms of the eigendecomposition of $\mathbf{A}$. Theorem 3.1 gives a set of basis vectors for each of the four subspaces of $\mathbf{\Pi}$ in terms of the eigenvectors of the state transition matrix $\mathbf{A}$, and Corollary 3.1 expresses the eigendecomposition of $\mathbf{A}$ in terms of the subspaces of $\mathbf{\Pi}$ for a VAR(1) system.

Theorem 3.1.

Consider a cointegrated VAR(k) model expressed in state-space form, defined by Eq. 3.11. Furthermore, consider the block form of the eigenvector and inverse eigenvector matrices $\mathbf{M}$ *and* $\mathbf{M}^{-1}$ *corresponding to the state transition matrix,* $\mathbf{A}$*, as introduced in Appendix A, and repeated here:*

$$\mathbf{M} = \begin{pmatrix} \mathbf{m}_{1,1} & \mathbf{m}_{1,2} & \cdots & \mathbf{m}_{1,pk} \\ \lambda_1^{-1}\mathbf{m}_{1,1} & \lambda_2^{-1}\mathbf{m}_{1,2} & \cdots & \lambda_{pk}^{-1}\mathbf{m}_{1,pk} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{-k+1}\mathbf{m}_{1,1} & \lambda_2^{-k+1}\mathbf{m}_{1,2} & \cdots & \lambda_{pk}^{-k+1}\mathbf{m}_{1,pk} \end{pmatrix},$$

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{n}_{1,1}^T & \mathbf{n}_{1,2}^T & \cdots & \mathbf{n}_{1,k}^T \\ \mathbf{n}_{2,1}^T & \mathbf{n}_{2,2}^T & \cdots & \mathbf{n}_{2,k}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{n}_{pk,1}^T & \mathbf{n}_{pk,2}^T & \cdots & \mathbf{n}_{pk,k}^T \end{pmatrix},$$

In addition, assume that the first $p - r$ *eigenvalues of* $\mathbf{A}$ *are equal to unity, while the rest satisfy* $|\lambda_i| < 1$ *for* $i = p - r + 1, ..., pk$*. Then, basis vectors for the four fundamental subspaces of* $\mathbf{\Pi}$ *can be defined in terms of the eigendecomposition of* $\mathbf{A}$*, as follows:*

***Subspace of* $\mathbf{\Pi}$**	***Name***	***Basis Vectors***
Column space	$\boldsymbol{\alpha}$	$\mathbf{m}_{1,p-r+1}, ..., \mathbf{m}_{1,pk}$
Left nullspace	$\boldsymbol{\alpha}_\perp$	$\mathbf{n}_{1,1}^T, ..., \mathbf{n}_{p-r,1}^T$
Row space	$\boldsymbol{\beta}$	$\sum_{j=1}^k \mathbf{n}_{p-r+1,j}^T, ..., \sum_{j=1}^k \mathbf{n}_{pk,j}^T$
Nullspace	$\boldsymbol{\beta}_\perp$	$\mathbf{m}_{1,1}, ..., \mathbf{m}_{1,p-r}$

Corollary 3.1.

For a cointegrated VAR(1) system, the eigendecomposition of $\mathbf{\Pi}$ *may be expressed as follows:*

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\beta}_\perp & \boldsymbol{\alpha} \end{pmatrix} \begin{pmatrix} \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix} \begin{pmatrix} \left(\boldsymbol{\alpha}_\perp^T \mathbf{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}_\perp^T \\ \left(\boldsymbol{\beta}^T \boldsymbol{\alpha}\right)^{-1} \boldsymbol{\beta}^T \end{pmatrix},$$

where the diagonal matrix $\mathbf{D}_2$ *contains all the non-unit eigenvalues.*

The proofs of Thm. 3.1 and Corollary 3.1 are given in Appexdix 3.A. As shown, there is a direct mapping between the subspaces of the matrix $\mathbf{\Pi}$ and the eigenvectors of the matrix $\mathbf{A}$, and therefore a restriction in one domain implies a restriction in the other. In particular, the VAR-ECM and state-space model cointegration restrictions of Sections 3.2.1 and 3.2.2, are equivalent, as established in Theorem 3.2.

Theorem 3.2.

Restrictions 3.1 and 3.2, which place conditions on the parameters of a vector autoregressive process so that the underlying time series are cointegrated, are equivalent.

The proof of Thm. 3.2 is given in Appendix 3.A.

3.3 System Response

Here, the moving-average representation for the general VAR process given in Section A.3, is extended to the case of a cointegrated VAR process. First, closed-form expressions for the zero-input and zero-state responses are derived and the resulting total system response is presented. This expression is an alternative form of the Granger Representation Theorem [30], which is then compared to existing versions found in the literature.

3.3.1 Zero Input Response

Recall the zero-input response (ZIR) for a general VAR process, given in Eq. A.22. Let the vector $\mathbf{g} = \mathbf{M}^{-1}\mathbf{q}[0]$ contain the expansion coefficients of the initial state vector $\mathbf{q}[0]$. In a cointegrated VAR system with exactly $p - r$ unit eigenvalues, the ZIR is equal to:

$$\begin{aligned} \mathbf{y}_{zir}[n] &= \sum_{i=1}^{p-r} g_i \mathbf{m}_{1,i} + \sum_{i=p-r+1}^{pk} g_i \lambda_i^{n+1} \mathbf{m}_{1,i}, \\ &= \boldsymbol{\beta}_\perp \mathbf{g} + \boldsymbol{\alpha} \mathbf{D}_2^{n+1} \mathbf{g}, \end{aligned} \tag{3.13}$$

where the columns of the matrices $\begin{pmatrix} \mathbf{m}_{1,1} & \cdots & \mathbf{m}_{1,p-r} \end{pmatrix}$ and $\begin{pmatrix} \mathbf{m}_{1,p-r+1} & \cdots & \mathbf{m}_{1,pk} \end{pmatrix}$ span the $\boldsymbol{\beta}_\perp$ and $\boldsymbol{\alpha}$ spaces, respectively, as established in Thm. 3.1. Furthermore, the diagonal matrix $\mathbf{D}_2$ again contains all of the non-unit eigenvalues $\{\lambda_{p-r+1}, ..., \lambda_{pk}\}$.

As Eq. 3.13 reveals, the ZIR is the sum of constant and time-varying terms. The time-varying component is along the $\boldsymbol{\alpha}$ subspace, which is consistent with the interpretation that this is the error-correcting (disequilibrium readjustment) space, as discussed in Section 3.2.1. In the limit as $n \to \infty$, this component vanishes, since $|\lambda_i| < 1$ for $i = p - r + 1 \cdots pk$, and the overall ZIR converges to the constant component, $\boldsymbol{\beta}_\perp \mathbf{g}$. Thus, in the absence of a driving input process, the output process converges to a fixed point in the space spanned by the columns of $\boldsymbol{\beta}_\perp$, justifying the interpretation of this subspace as the attractor space.

Recall from Section 3.1 that there exists some $\mathbf{b} \in \mathbb{R}^p$ so that the scalar process $z = \mathbf{b}^T \mathbf{x}$ is AWSS, which implies that the ZIR must decay to zero as time increases. Choosing $\mathbf{b} \in \text{span}\{\boldsymbol{\beta}\}$ produces a ZIR equal to:

$$z_{zir}[n] = \mathbf{b}^T \mathbf{y}_{zir}[n] = \mathbf{b}^T \left(\boldsymbol{\beta}_\perp \mathbf{g} + \boldsymbol{\alpha} \mathbf{D}_2^{n+1} \mathbf{g} \right) = \mathbf{b}^T \boldsymbol{\alpha} \mathbf{D}_2^{n+1} \mathbf{g},$$

which indeed converges to zero as $n \to \infty$ as all of the elements of $\mathbf{D}_2$ have magnitudes less than one.

3.3.2 Zero State Response

In Appendix A, the zero-state response for a general VAR process is characterized by the transfer function matrix of Eq. A.27. In the special case of a cointegrated VAR system

with exactly $p-r$ unit eigenvalues, the transfer function matrix is equal to:

$$\begin{aligned}\mathbf{H}(z) &= \frac{\mathbf{m}_{1,1\ldots p-r}\mathbf{n}_{1\ldots p-r,1}^T}{1-z^{-1}} + \sum_{i=p-r+1}^{pk}\frac{\mathbf{m}_{1,i}\mathbf{n}_{i,1}^T}{1-\lambda_i z^{-1}}, \\ &= \frac{\boldsymbol{\beta}_\perp\left(\boldsymbol{\alpha}_\perp^T\boldsymbol{\Gamma}\boldsymbol{\beta}_\perp\right)^{-1}\boldsymbol{\alpha}_\perp^T}{1-z^{-1}} + \boldsymbol{\alpha}\mathbf{D}_2(z)\left(\boldsymbol{\beta}^T\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}^T. \end{aligned} \tag{3.14}$$

where the diagonal matrix $\mathbf{D}_2(z)$ contains the terms $\frac{1}{1-\lambda_i z^{-1}}$ for $i = p-r+1,..,pk$. Subsequently, the matrix $\mathbf{H}(z)$ is the sum of a marginally unstable component due to the pole at $z=1$, and a BIBO stable component due to all of the remaining poles strictly inside the unit circle. The corresponding time domain transfer function is computed by taking the inverse $\mathcal{Z}$-transform of Eq. 3.14, as follows:

$$\mathbf{h}[n] = \boldsymbol{\beta}_\perp\left(\boldsymbol{\alpha}_\perp^T\boldsymbol{\Gamma}\boldsymbol{\beta}_\perp\right)^{-1}\boldsymbol{\alpha}_\perp^T + \boldsymbol{\alpha}\mathbf{D}_2^n\left(\boldsymbol{\beta}^T\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}^T \text{ for } n \geq 0.$$

The overall zero-state response of a cointegrated VAR is equal to:

$$\begin{aligned}\mathbf{y}_{zsr}[n] &= \mathbf{h}[n] * \mathbf{u}[n] \\ &= \boldsymbol{\beta}_\perp\left(\boldsymbol{\alpha}_\perp^T\boldsymbol{\Gamma}\boldsymbol{\beta}_\perp\right)^{-1}\boldsymbol{\alpha}_\perp^T\sum_{j=0}^{n}\mathbf{u}[j] + \sum_{j=0}^{n}\boldsymbol{\alpha}\mathbf{D}_2^{n-j}\left(\boldsymbol{\beta}^T\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}^T\mathbf{u}[j].\end{aligned}$$

As with the ZIR, it is necessary to confirm that there exists some $\mathbf{b} \in \mathbb{R}^p$ so that the scalar process $z = \mathbf{b}^T\mathbf{x}$ is WSS. Letting $\mathbf{b} \in \text{span}\{\boldsymbol{\beta}\}$, the ZSR is equal to:

$$\begin{aligned}z_{zsr}[n] &= \mathbf{b}^T\mathbf{y}_{zsr}[n] \\ &= \sum_{j=0}^{n}\mathbf{b}^T\boldsymbol{\alpha}\mathbf{D}_2^{n-j}\left(\boldsymbol{\beta}^T\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}^T\mathbf{u}[j],\end{aligned}$$

which is indeed WSS since $|\lambda_i| < 1$ for $i = p-r+1\cdots pk$.

3.3.3 Total System Response

The total output response of a cointegrated VAR process, is computed as the sum of its zero-input and zero-state responses, as follows:

$$\begin{aligned}\mathbf{y}[n] &= \mathbf{y}_{zsr}[n] + \mathbf{y}_{zir}[n] \\ &= \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}_\perp^T \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}_\perp^T \sum_{j=0}^{n} \mathbf{u}[j] + \sum_{j=0}^{n} \boldsymbol{\alpha} \mathbf{D}_2^{n-j} \left(\boldsymbol{\beta}^T \boldsymbol{\alpha}\right)^{-1} \boldsymbol{\beta}^T \mathbf{u}[j] + \boldsymbol{\beta}_\perp \mathbf{g} + \boldsymbol{\alpha} \mathbf{D}_2^{n+1} \mathbf{g}.\end{aligned} \tag{3.15}$$

The first term represents an integrated process, scaled by the matrix factor $\boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}_\perp^T \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}_\perp^T$. The second term represents an AWSS random process, as it is the sum of a set of AWSS random processes formed by convolving the input process $\mathbf{u}[n] = \boldsymbol{\Phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n]$ with an exponentially decaying moving-average sequence. Finally, the last two terms comprising the zero-input response consist of a constant term and a term that decays to zero as n increases.

A simple rearrangement of the terms in Eq. 3.15 yields:

$$\mathbf{y}[n] = \boldsymbol{\beta}_\perp \left(\left(\boldsymbol{\alpha}_\perp^T \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}_\perp^T \sum_{j=1}^{n} \mathbf{u}[j] + \mathbf{g} \right) + \boldsymbol{\alpha} \left(\sum_{j=0}^{n} \mathbf{D}_2^{n-j} \left(\boldsymbol{\beta}^T \boldsymbol{\alpha}\right)^{-1} \boldsymbol{\beta}^T \mathbf{u}[j] + \mathbf{D}_2^{n+1} \mathbf{g} \right) \tag{3.16}$$

This representation for $\mathbf{y}[n]$ clearly shows that a cointegrated VAR process can be decomposed into two components that exist within non-orthogonal subspaces. The first part corresponds to an integrated component plus a constant term within the $\boldsymbol{\beta}_\perp$ subspace, while the second part corresponds to an AWSS random process within the $\boldsymbol{\alpha}$ subspace. From this representation it is clear that the scalar process $z[n] = \mathbf{b}^T \mathbf{y}[n]$ for $\mathbf{b} \in \text{span}\{\boldsymbol{\beta}\}$ is also an AWSS random process, due to the fact that the integrated and constant components are removed as $\mathbf{b}^T \boldsymbol{\beta}_\perp = \mathbf{0}$.

3.3.4 Relationship to Granger Representation Theorem

The total system response of a cointegrated VAR system, defined by Eq. 3.15, is an alternate form of the celebrated Granger Representation Theorem. There are numerous versions of this theorem in the literature, each establishing the relationship between the cointegration property and the various forms of a VAR process. The name was first coined by Engle and

Granger [24], in which the equivalence of the autoregressive, moving-average, and error-correcting forms of a cointegrated VAR process is established. The theorem is subsequently expanded on by Johansen [29], who derived an explicit moving-average representation that decomposes the output into three factors corresponding to a random walk component, stationary component, and an initial value response, as follows:

$$\mathbf{y}[n] \;=\; \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}_\perp^T \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}_\perp^T \sum_{j=1}^{n} \mathbf{u}[j] + \mathbf{C}_1(L)\mathbf{u}[n] + \mathbf{a},$$

where the vector $\mathbf{a}$ depends on the initial conditions and satisfies $\boldsymbol{\beta}^T\mathbf{a} = 0$, and $\mathbf{C}_1(L)$ is a infinite order matrix polynomial of the lag (delay) operator L. Furthermore, the term $\mathbf{C}_1(L)$ satisfies:

$$\frac{1}{\mathbb{I} - z^{-1}\boldsymbol{\Pi}_1 + \ldots + z^{-(k-1)}\boldsymbol{\Pi}_{k-1} + z^{-k}\boldsymbol{\Pi}_k} = \frac{C}{1-z} + \mathbf{C}_1(z),$$

for $z \neq 1$. Johansen implicitly defines $\mathbf{C}_1(L)$ in terms of the characteristic polynomial of the VAR process and the matrix coefficient of the random walk component.

On the other hand, an explicit representation for $\mathbf{C}_1(L)$ is developed in [28]. The coefficients of the polynomial are given by the following recursion:

$$\begin{aligned} \Delta\mathbf{C}_1[n] &= \boldsymbol{\Pi}\mathbf{C}_1[n] + \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta\mathbf{C}_1[n-i], \\ \mathbf{C}_1[0] &= \mathbb{I} - \mathbf{C}, \\ \mathbf{C}_1[-1]\cdots\mathbf{C}_1[-(k-1)] &= -\mathbf{C}. \end{aligned} \tag{3.17}$$

Note the similarity of Eq. 3.17 to the VAR-ECM form given in Eq. 3.5. While this work goes one step further than Johansen's representation by providing an explicit closed-form representation for each of the coefficients, the resulting representation is not a pure moving-average equation due to the recursive definition for $\mathbf{C}_1(L)$.

A state-space approach to analyzing cointegrated VAR processes, similar to the approach taken here, is presented in [9]. The coefficient matrices preceding each factor in Eq. 3.15 are expressed in terms of the Jordan form of the state-space model, rather than as functions

of the base block vectors $\{\mathbf{m}_{1,i}, \mathbf{n}_{i,1}\}$ of the original system transition matrix.

In contrast to existing representations, the total system response dervied here and given in Eq. 3.15 highlights the underlying geometry of a cointegrated system. The nonstationary integrated and AWSS components are explicitly defined in terms of the $\{\boldsymbol{\alpha}, \boldsymbol{\alpha}_\perp, \boldsymbol{\beta}, \boldsymbol{\beta}_\perp\}$ subspaces of the matrix $\boldsymbol{\Pi}$, without the need for any recursive specifications. From this form it is immediately clear why choosing any vector from the span of $\boldsymbol{\beta}$ produces a WSS scalar random process.

3.3.5 Order of Integration

The total system response given in Eq. 3.15 contains only a single integrator, and therefore encodes the belief that each component process of the vector process is at most integrated of order one. In this section, a restriction on the parameters of the VAR model is given that can be used to test that the resulting component processes are indeed $I(1)$, and not integrated of order two, denoted as $I(2)$. Two forms of the restriction are presented, one in terms of the eigenstructure of the state transition matrix $\mathbf{A}$, and the other in terms of the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

First consider the following example, which elucidates the relationship between the unit eigenvalues of $\mathbf{A}$ and the order of integration of the process.

Example 3.3.

Consider a cointegrated VAR(2) process:

$$\mathbf{x}[n] = \boldsymbol{\Pi}_1\mathbf{x}[n-1] + \boldsymbol{\Pi}_2\mathbf{x}[n-2] + \boldsymbol{\epsilon}[n].$$

When expressed in state-space form, suppose that the state transition matrix, $\mathbf{A} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 \\ \mathbb{I} & \mathbf{0} \end{pmatrix}$, has eigenvalues $\lambda = \{1, 1, 0.25, 0.5\}$. While the algebraic multiplicity of the unit eigenvalue is two, further assume the associated eigenspace has reduced rank, implying that the geometric multiplicity is equal to one. As a result, $\mathbf{A}$ is not diagonalizable, but rather may be factored into Jordan canonical form, as discussed in Appendix A.2.3. Accordingly, the state transition matrix is equal to $\mathbf{A} = \mathbf{M}\mathbf{D}\mathbf{M}^{-1}$, where the columns of the matrix $\mathbf{M}$ contain the

generalized eigenvectors, and the matrix $\mathbf{D}$ contains the *Jordan blocks* along the diagonal. For the VAR(2) process under consideration here, these two matrices are given by:

$$\mathbf{D} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0.5 \end{pmatrix}, \qquad \mathbf{M} = \begin{pmatrix} \mathbf{m}_{1,1} & \mathbf{m}_{1,2} & \mathbf{m}_{1,3} & \mathbf{m}_{1,4} \\ \mathbf{m}_{1,1} & \mathbf{m}_{1,2} - \mathbf{m}_{1,1} & 4\mathbf{m}_{1,3} & 2\mathbf{m}_{1,4} \end{pmatrix}.$$

While it is not possible to transform the system to modal coordinates, it is possible to apply the transformation $\mathbf{r} = \mathbf{M}^{-1}\mathbf{q}$, yielding a system similar to the form given in Eq. A.16. Here the modes are not fully decoupled, but are separated as much as possible. The system transfer function can be computed according to Eq. A.26, as follows:

$$\begin{aligned} \mathbf{H}(z) &= \tilde{\mathbf{c}}^T \left(z\mathbb{I} - \tilde{\mathbf{A}} \right)^{-1} \tilde{\mathbf{b}} + \tilde{\mathbf{d}}, \\ &= \tilde{\mathbf{c}}^T \begin{pmatrix} z-1 & 1 & 0 & 0 \\ 0 & z-1 & 0 & 0 \\ 0 & 0 & z-0.25 & 0 \\ 0 & 0 & 0 & z-0.5 \end{pmatrix}^{-1} \tilde{\mathbf{b}} + \tilde{\mathbf{d}}, \\ &= \tilde{\mathbf{c}}^T \begin{pmatrix} \frac{1}{(z-1)} & \frac{1}{(z-1)^2} & 0 & 0 \\ 0 & \frac{1}{(z-1)} & 0 & 0 \\ 0 & 0 & \frac{1}{(z-0.25)} & 0 \\ 0 & 0 & 0 & \frac{1}{(z-0.5)} \end{pmatrix} \tilde{\mathbf{b}} + \tilde{\mathbf{d}}, \end{aligned}$$

where the last line follows by computing the matrix function $f(\mathbf{A}) = (z\mathbb{I} - \mathbf{A})^{-1}$ using the characteristic polynomial of $\mathbf{A}$, as described in Chapter 2 of [21]. The transfer function contains both $\frac{1}{(z-1)}$ and $\frac{1}{(z-1)^2}$ terms, so that the inverse $\mathcal{Z}$-transform contains terms of the form $\sum_{i=0}^{n} \mathbf{u}[i]$ and $\sum_{j=0}^{n} \sum_{i=0}^{j} \mathbf{u}[i]$, resulting in an $I(2)$ output process. ■

As this example demonstrates, there is a direct relationship between the geometric and algebraic multiplicities for the unit eigenvalues and the resulting order of integration. This, in turn, naturally leads to an equivalent restriction in terms of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ matrices. Both restrictions are given in Theorem 3.3 below.

Theorem 3.3.

A cointegrated vector autoregressive process is integrated of order one so long as:

- *The unit eigenvalues associated with the state transition matrix* $\mathbf{A}$ *have equivalent geometric and algebraic multiplicities, or equivalently,*
- *the condition* $\boldsymbol{\alpha}_{\perp}^{T}\boldsymbol{\Gamma}\boldsymbol{\beta}_{\perp} = \mathbf{0}$ *holds.*

The proof of Thm 3.3 is given in Appendix 3.A. The first restriction, expressed in terms of the dimension of the eigen space of $\mathbf{A}$ corresponding to the unit eigenvalues, prevents anything but a single integrator from appearing at the output. The second restriction, expressed in terms of the subspaces of the matrix $\boldsymbol{\Pi}$, is closely related to the $I(1)$ condition given in [29], which states that only the determinant of $\boldsymbol{\alpha}_{\perp}^{T}\boldsymbol{\Gamma}\boldsymbol{\beta}_{\perp}$ must be non-zero in order for the system to be integrated of order one.

3.4 Estimation

This section presents an overview of the procedure needed in order to estimate the parameters of a cointegrated VAR process from a random sample of data. As depicted in Fig. 3-3, the presence of a unit root must first be verified in each of the constituent time series. Once confirmed, the number of cointegrating relations (r) and common trends ($p-r$) must be estimated, and only then can the VAR model parameters can be computed. Each block of Fig. 3-3 is described in detail below. The maximum likelihood analysis used both to estimate r and the VAR model parameters follows the work of Johansen [30]. While not discussed here, many alternative estimation procedures can be found in the literature, including the non-parametric work of [15], the subspace methods of [8, 10, 63], the Bayesian techniques described in [32, 33, 61], the canonical correlations analysis used by [14, 16], and the principal components analysis method proposed by [57].

3.4.1 Unit Root Test

The cointegrated VAR model defined here assumes that the shaping filter for each constituent time series contains a single pole at unity, while all of the remaining poles lie strictly inside the unit circle. This implies that while the signal is integrated of order one, the resulting first-difference series is WSS or AWSS, depending on the choice of initial con-

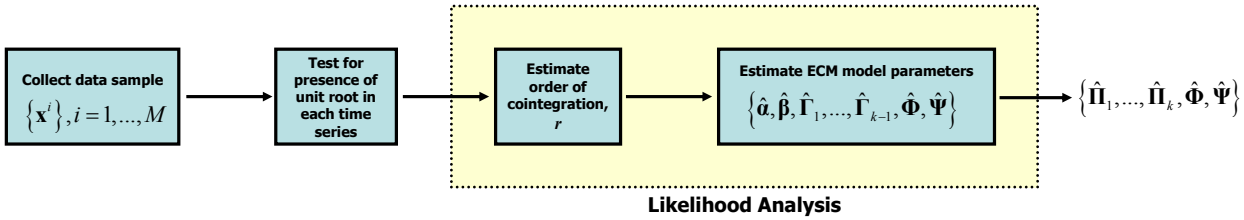


Figure 3-3. Cointegrated VAR Estimation Procedure.

dition. The presence of unit pole (unit root) in each data set must be statistically verified before the parameters of the cointegrated VAR can be estimated. The most popular statistical test for the presence of a unit root is the Dickey-Fuller (DF) test [22], described here.

Let $x_i[n]$ denote the scalar time series comprising the i^{th} component of the vector process $\mathbf{x}[n]$. In the DF test, one first estimates a model of the form:

$$\Delta x_i[n] = a_0 + a_1 n + a_2 x_i[n-1] + \epsilon[n],$$

by ordinary least squares (OLS) regression. Absence or presence of the unit root is determined by a likelihood ratio hypothesis test against a null hypothesis that $a_2 = 0$. Lagged first difference terms may be included to remove any serial autocorrelation present in the data, resulting in what is known as the Augmented Dickey-Fuller (ADF) test. In this case the regression model may be written as follows:

$$\Delta x_i[n] = a_0 + a_1 n + a_2 x_i[n-1] + \sum_{i=1}^{k} b_i \Delta x_i[n-i] + \epsilon[n],$$

in which the parameters are again fit via an OLS procedure.

3.4.2 Likelihood Analysis

Both the order of cointegration and estimates for the cointegrated VAR model parameters are determined from an analysis of the likelihood function of the observed data sample under the assumption that the stochastic input, $\boldsymbol{\epsilon}_n$, is i.i.d. according to a Normal distribution with zero mean and covariance matrix $\boldsymbol{\Psi}$. The ML estimator given in Section A.4 does not enforce the restriction that $\boldsymbol{\Pi}$ has reduced rank and can be factored as $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^T$. A new estimator can be derived starting from the ECM form of a VAR process, given by Eq. 3.9,

as this description explicitly contains the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The set of parameters to be estimated is now given by $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}_1, ..., \boldsymbol{\Gamma}_{k-1}, \boldsymbol{\Phi}, \boldsymbol{\Psi}\}$, which in turn can be used with Eq. 3.7 to determine the parameters $\{\boldsymbol{\Pi}_1, ..., \boldsymbol{\Pi}_k\}$. For notational convenience, Eq. 3.9 is rewritten in linear regression form, as follows:

$$\mathbf{z}_0[n] = \boldsymbol{\alpha}\boldsymbol{\beta}^T \mathbf{z}_1[n] + \boldsymbol{\theta}^T \mathbf{z}_2[n] + \boldsymbol{\epsilon}[n], \tag{3.18}$$

where:

$$\mathbf{z}_0 = \Delta\mathbf{x}[n], \quad \mathbf{z}_1 = \mathbf{x}[n-1], \quad \mathbf{z}_2 = \begin{pmatrix} \Delta\mathbf{x}[n-1] \\ \vdots \\ \Delta\mathbf{x}[n-k+1] \\ \mathbf{d}[n] \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\Gamma}_1 \\ \vdots \\ \boldsymbol{\Gamma}_{k-1} \\ \boldsymbol{\Phi} \end{pmatrix}.$$

Given a random sample $\left\{\mathbf{x}^1, ..., \mathbf{x}^M\right\}$, the likelihood function is equal to:

$$\begin{aligned} L\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\Psi}\right) &= Pr_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\Psi}}\left(\mathbf{x}^1, ..., \mathbf{x}^M\right), \\ &= \prod_{i=1}^{M} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Psi}|^{1/2}} \exp\left[-0.5\left(\mathbf{z}_0^i - \boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{z}_1^i - \boldsymbol{\theta}^T\mathbf{z}_2^i\right)^T \boldsymbol{\Psi}^{-1}\left(\mathbf{z}_0^i - \boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{z}_1^i - \boldsymbol{\theta}^T\mathbf{z}_2^i\right)\right], \end{aligned}$$

with corresponding log-likelihood function:

$$\begin{aligned} \ell\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\Psi}\right) &= -\frac{pM}{2}\log(2\pi) - \frac{M}{2}\log\left(|\boldsymbol{\Psi}|\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{M}\left(\mathbf{z}_0^i - \boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{z}_1^i - \boldsymbol{\theta}^T\mathbf{z}_2^i\right)^T \boldsymbol{\Psi}^{-1}\left(\mathbf{z}_0^i - \boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{z}_1^i - \boldsymbol{\theta}^T\mathbf{z}_2^i\right). \end{aligned} \tag{3.19}$$

The estimators are defined as solutions to the following optimization problem:

$$\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Psi}} = \underset{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\Psi}}{\arg\max}\ \ell\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\Psi}\right),$$

and are computed according to the following two-step procedure, known in the literature as reduced-rank regression and first introduced by Anderson [3]. First, the log-likelihood function is differentiated with respect to $\boldsymbol{\theta}$, which produces an estimator that is a function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, given by:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left(\left(\sum_{i=1}^{M}\mathbf{z}_0^i\mathbf{z}_2^{iT}\right) - \alpha\beta^T\left(\sum_{i=1}^{M}\mathbf{z}_1^i\mathbf{z}_2^{iT}\right)\right)\left(\sum_{i=1}^{M}\mathbf{z}_2^i\mathbf{z}_2^{iT}\right)^{-1}. \tag{3.20}$$

Next, substitute Eq. 3.20 back into Eq. 3.18, resulting in a new regression equation, given by:

$$\mathbf{w}_0[n] = \boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{w}_1[n] + \tilde{\boldsymbol{\epsilon}}[n], \tag{3.21}$$

with:

$$\begin{aligned}
\mathbf{w}_0 &= \mathbf{z}_0 - \left(\sum_{i=1}^{M}\mathbf{z}_0^i\mathbf{z}_2^{iT}\right)\left(\sum_{i=1}^{M}\mathbf{z}_2^i\mathbf{z}_2^{iT}\right)^{-1}\mathbf{z}_2, \\
\mathbf{w}_1 &= \mathbf{z}_1 - \left(\sum_{i=1}^{M}\mathbf{z}_1^i\mathbf{z}_2^{iT}\right)\left(\sum_{i=1}^{M}\mathbf{z}_2^i\mathbf{z}_2^{iT}\right)^{-1}\mathbf{z}_2.
\end{aligned}$$

For fixed $\boldsymbol{\beta}$, estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}$ are obtained from Eq. 3.21, as follows:

$$\hat{\boldsymbol{\alpha}}^T(\boldsymbol{\beta}) = \left(\sum_{i=1}^{M}\boldsymbol{\beta}^T\mathbf{w}_1^i\mathbf{w}_1^{iT}\boldsymbol{\beta}\right)^{-1}\left(\sum_{i=1}^{M}\boldsymbol{\beta}^T\mathbf{w}_1^i\mathbf{w}_0^{iT}\right), \tag{3.22}$$

$$\begin{aligned}
\hat{\boldsymbol{\Psi}}(\boldsymbol{\beta}) &= \frac{1}{M}\sum_{i=1}^{M}\left(\mathbf{w}_0^i - \hat{\boldsymbol{\alpha}}\boldsymbol{\beta}^T\mathbf{w}_1^i\right)\left(\mathbf{w}_0^i - \hat{\boldsymbol{\alpha}}\boldsymbol{\beta}^T\mathbf{w}_1^i\right)^T, \\
&= \frac{1}{M}\sum_{i=1}^{M}\mathbf{w}_0^i\mathbf{w}_0^{iT} - \frac{2}{M}\hat{\boldsymbol{\alpha}}\boldsymbol{\beta}^T\sum_{i=1}^{M}\mathbf{w}_1^i\mathbf{w}_0^{iT} + \frac{1}{M}\hat{\boldsymbol{\alpha}}\boldsymbol{\beta}^T\left(\sum_{i=1}^{M}\mathbf{w}_1^i\mathbf{w}_1^{iT}\right)\boldsymbol{\beta}\hat{\boldsymbol{\alpha}}^T, \\
&= \frac{1}{M}\sum_{i=1}^{M}\mathbf{w}_0^i\mathbf{w}_0^{iT} - \frac{1}{M}\hat{\boldsymbol{\alpha}}\boldsymbol{\beta}^T\left(\sum_{i=1}^{M}\mathbf{w}_1^i\mathbf{w}_1^{iT}\right)\boldsymbol{\beta}\hat{\boldsymbol{\alpha}}^T,
\end{aligned} \tag{3.23}$$

where the last line follows from the fact that $\sum_{i=1}^{M}\boldsymbol{\beta}^T\mathbf{w}_1^i\mathbf{w}_0^{iT} = \sum_{i=1}^{M}\boldsymbol{\beta}^T\mathbf{w}_1^i\mathbf{w}_1^{iT}\boldsymbol{\beta}\hat{\boldsymbol{\alpha}}^T$. All that remains is to estimate $\boldsymbol{\beta}$. To this end, observe that substitution of the ML estimates for $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}$ back into Eq. 3.19 yields a log-likelihood function that is proportional to $\left|\hat{\boldsymbol{\Psi}}\right|$. For notational convenience, define:

$$\mathbf{S}_{ij} = \frac{1}{M}\sum_{k=1}^{M}\mathbf{w}_k^i\mathbf{w}_j^{kT},$$

so that the determinant of $\hat{\mathbf{\Psi}}$ can be expressed as follows:

$$\begin{aligned}\left|\hat{\mathbf{\Psi}}\right| &= \left|\mathbf{S}_{00} - \hat{\boldsymbol{\alpha}}\boldsymbol{\beta}^T\mathbf{S}_{11}\boldsymbol{\beta}\hat{\boldsymbol{\alpha}}^T\right|, \\ &= \left|\mathbf{S}_{00} - \mathbf{S}_{01}\boldsymbol{\beta}\left(\boldsymbol{\beta}^T\mathbf{S}_{11}\boldsymbol{\beta}\right)^{-1}\boldsymbol{\beta}^T\mathbf{S}_{10}\right|, \\ &= \frac{|\mathbf{S}_{00}|\left|\boldsymbol{\beta}^T\left(\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\right)\boldsymbol{\beta}\right|}{\left|\boldsymbol{\beta}^T\mathbf{S}_{11}\boldsymbol{\beta}\right|}. \end{aligned} \tag{3.24}$$

According to Lemma A.8 of [30], Eq. 3.24 is maximized by choosing:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_r\end{pmatrix}, \tag{3.25}$$

which represents the eigenvectors corresponding to the r largest eigenvalues that satisfy:

$$\lambda_i\mathbf{S}_{11}\mathbf{v}_i = \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\mathbf{v}_i.$$

It is also shown in Lemma A.8 of [30] that the corresponding maximal value for $\left|\hat{\mathbf{\Psi}}\right|$ is $|\mathbf{S}_{00}|\prod_{i=1}^{r}\lambda_i$.

The ML estimators for the parameters of a cointegrated VAR are given by Eqs. 3.25, 3.22, 3.23, and 3.20, and all that remains is to estimate the value of r, the rank of the matrix $\boldsymbol{\beta}$ (i.e. the dimension of the cointegrating space). This parameter must either be known a priori or determined from the data. One estimation procedure, as described in Ch. 12 of [30], is based on a sequence of likelihood ratio tests. Let $H(r)$ denote the hypothesis that there are r cointegrating relations, and let $H(p)$ denote the hypothesis that there are p relations, i.e. that the matrix $\mathbf{\Pi}$ has full rank. Next, let Q_r denote the likelihood ratio test statistic of $H(r)$ versus $H(p)$, defined as follows:

$$Q_r = -2\log\frac{|\mathbf{S}_{00}|\prod_{i=1}^{r}\lambda_i}{|\mathbf{S}_{00}|\prod_{i=1}^{p}\lambda_i} = -\frac{1}{M}\sum_{i=r+1}^{p}\lambda_i.$$

The estimator for the order of cointegration at a given confidence level, $\hat{r}$, is given by the smallest value of r such that $Q_r \leq c_{p-r}$, where c_i denote the corresponding critical value, as tabulated in [30].

3.5 Literature Review

3.5.1 Previous Work in Portfolio Theory

Cointegration has been the focus of a small set of research on trading and portfolio selection. Early work focused on the use of statistical arbitrage techniques to trade an AWSS linear combination of the cointegrated price or log-price time series, formed by selecting a portfolio weight vector from within the span of the cointegrating space, $\boldsymbol{\beta}$. One such technique is a mean-reverting scheme, detailed in [31], in which the entire portfolio is bought when the portfolio value deviates from its mean by some predetermined threshold such as one standard deviation, and the position is closed when the signal mean reverts. Such portfolios are often referred to as the *Beta portfolios*, as discussed in detail in Section 4.1.

Another area of research focuses on the construction of tracking portfolios that are cointegrated with a market index or benchmark [2]. Drawing on ideas from *enhanced indexation* [54], replicating portfolios are formed that track a benchmark return plus or minus some additional fixed return, so that the spread on these two portfolios may be traded. For example, one may construct a two portfolios, one whose value is cointegrated with the $S\&P$ 500 index plus 5%, the other which is cointegrated with the $S\&P$ 500 index minus 5%. The investor can then take a long position in the plus portfolio, and a short position in the minus portfolio, earning, on average, the 10% spread.

The work of Burgess draws on both the idea of cointegrated replicating portfolios and statistical arbitrage [18]. First, a portfolio is constructed so that the value of the portfolio is cointegrated with the value of some target asset, such as a particular stock. The resulting system, cointegrated by construction, can then be used as the basis for a pairs "relative value" trading scheme, in which positions in the two assets are taken based on whether the current prices are "cheap" or "expensive" relative to the long-term equilibrium value.

A continuous-time formulation of the portfolio choice problem with cointegrated assets is studied by Lucas [40]. Here it is assumed that the investor makes decisions based on constant relative risk averse (CRRA) preferences, such as those captured by a log utility function, as compared to the mean-variance approach adopted in this thesis. In Lucas'

model, the investor acts to maximize the present value of his or her expected utility across the entire trading period, and both short- and long-term time horizons are considered. The results of Chapter 4 of this thesis agree with the general theme of the results in [40], mainly that for short periods both the error-correction ($\boldsymbol{\alpha}$) and cointegration ($\boldsymbol{\beta}$) forces play a role in the choice of portfolio weights, while for long time horizons the cointegration effects dominate. However, the details differ due to the different underlying assumptions concerning investor preferences.

3.5.2 Applications in Econometrics

In addition to modeling tradable asset prices or returns, cointegrated VAR processes have played a notable role in the econometrics literature. For example, such models are often used in the study of exchange rate dynamics [5, 6, 23]. For studies of cointegration in fixed income markets, see [26, 65, 68]. More recently, cointegration has been used in the study of the consumption wealth ratio. In [41], it is shown that consumption, asset wealth, and labor income are cointegrated, and this idea is further exploited in [34] to build a linear predictor of asset returns.

3.A Proofs of Chapter 3 Theorems

Proof of Theorem 3.1.

A set of basis vectors for the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ subspaces are determined in terms of the eigenvectors of the state transition matrix $\mathbf{A}$. According to Eq. 3.27, the nullspace of $\mathbf{\Pi}$ is spanned by the set of vectors $\{\mathbf{m}_{1,i}\}$ for $i = 1, ..., p-r$, which correspond to the base blocks of the $p-r$ eigenvectors of $\mathbf{A}$ with unit eigenvalue. Similarly, Eq. A.14 with $\lambda_i = 1$ implies that the left nullspace of $\mathbf{\Pi}$ is spanned by the set of vectors $\left\{\mathbf{n}_{i,1}^T\right\}$ from the base blocks of the corresponding left eigenvectors. Furthermore, according to the SVD of $\mathbf{\Pi}$ given in 3.B, the right and left nullspaces of $\mathbf{\Pi}$ are also spanned by the columns of $\boldsymbol{\beta}_\perp$ and $\boldsymbol{\alpha}_\perp$, respectively. Hence, $\text{sp}\{\mathbf{m}_{1,i}\} = \text{sp}\{\boldsymbol{\beta}_\perp\}$, and the $\text{sp}\{\mathbf{n}_{i,1}\} = \text{sp}\{\boldsymbol{\alpha}_\perp\}$.

Basis vectors for the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ subspaces are determined as follows:

$$
\begin{aligned}
\mathbf{\Pi} &= \begin{pmatrix} \mathbb{I}_p \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}^T \mathbf{A} \begin{pmatrix} \mathbb{I}_p \\ \mathbb{I}_p \\ \vdots \\ \mathbb{I}_p \end{pmatrix} - \mathbb{I}_p = \begin{pmatrix} \mathbb{I}_p \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}^T \mathbf{M}\mathbf{D}\mathbf{M}^{-1} \begin{pmatrix} \mathbb{I}_p \\ \mathbb{I}_p \\ \vdots \\ \mathbb{I}_p \end{pmatrix} - \begin{pmatrix} \mathbb{I}_p \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}^T \mathbf{M}\mathbf{M}^{-1} \begin{pmatrix} \mathbb{I}_p \\ \mathbb{I}_p \\ \vdots \\ \mathbb{I}_p \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{m}_{1,1} & \mathbf{m}_{1,2} & \cdots & \mathbf{m}_{1,pk} \end{pmatrix} \begin{pmatrix} \lambda_1 - 1 & 0 & \cdots & 0 \\ 0 & \lambda_2 - 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{pk} - 1 \end{pmatrix} \begin{pmatrix} \mathbf{n}_{1,1}^T + \mathbf{n}_{1,2}^T + \cdots + \mathbf{n}_{1,k}^T \\ \mathbf{n}_{2,1}^T + \mathbf{n}_{2,2}^T + \cdots + \mathbf{n}_{2,k}^T \\ \vdots \\ \mathbf{n}_{pk,1}^T + \mathbf{n}_{pk,2}^T + \cdots + \mathbf{n}_{pk,k}^T \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{m}_{1,p-r+1} & \cdots & \mathbf{m}_{1,pk} \end{pmatrix} \begin{pmatrix} \lambda_{p-r+1} - 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{pk} - 1 \end{pmatrix} \begin{pmatrix} \sum_{j=1}^k \mathbf{n}_{2,j}^T \\ \vdots \\ \sum_{j=1}^k \mathbf{n}_{pk,j}^T \end{pmatrix}.
\end{aligned}
$$

Setting

$$
\boldsymbol{\alpha} = \begin{pmatrix} \mathbf{m}_{1,p-r+1} & \cdots & \mathbf{m}_{1,pk} \end{pmatrix} \begin{pmatrix} \lambda_{p-r+1} - 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{pk} - 1 \end{pmatrix}^{1/2}
$$

and

$$\boldsymbol{\beta}^T = \begin{pmatrix} \lambda_{p-r+1} - 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{pk} - 1 \end{pmatrix}^{1/2} \begin{pmatrix} \sum_{j=1}^{k} \mathbf{n}_{2,j}^T \\ \vdots \\ \sum_{j=1}^{k} \mathbf{n}_{pk,j}^T \end{pmatrix},$$

the result follows.

Proof of Corollary 3.1.

Here, the eigendecomposition of $\mathbf{A}$ is expressed as a function of the four subspaces of $\mathbf{\Pi}$. First, without loss of generality, the matrix $\boldsymbol{\beta}_\perp$ can be represented as:

$$\boldsymbol{\beta}_\perp = \mathbf{m}_{1,1\ldots p-r} \triangleq \begin{pmatrix} \mathbf{m}_{1,1} & \mathbf{m}_{1,2} & \cdots & \mathbf{m}_{1,p-r} \end{pmatrix}.$$

Second, according to Eq. A.15, the following must be true when $\lambda_i = 1$:

$$\begin{aligned} \mathbf{n}_i^T \mathbf{m}_i &= 1, \\ \mathbf{n}_{i,1}^T \left(\sum_{j=2}^{k} \sum_{k=1}^{j-1} \lambda_i^{-k} \mathbf{\Pi}_j \right) \mathbf{m}_{1,i} &= 1, \\ \mathbf{n}_{i,1}^T \left(\sum_{j=2}^{k} (j-1) \mathbf{\Pi}_j \right) \mathbf{m}_{1,i} &= 1, \\ \mathbf{n}_{i,1}^T \mathbf{\Gamma} \mathbf{m}_{1,i} &= 1, \end{aligned} \tag{3.26}$$

where

$$\mathbf{\Gamma} \triangleq \mathbb{I} - \sum_{j=1}^{k-1} \mathbf{\Gamma}_j = \sum_{j=2}^{k} (j-1) \mathbf{\Pi}_j,$$

and $\mathbf{\Gamma}_j$ is defined according to Eq. 3.7. For the specific case where $\mathbf{m}_{1,1\ldots p-r} = \boldsymbol{\beta}_\perp$, the corresponding $\mathbf{n}_{1\ldots p-r,1}$ are constrained as:

$$\mathbf{n}_{1\ldots p-r,1} \triangleq \begin{pmatrix} \mathbf{n}_{1,1}^T \\ \mathbf{n}_{2,1}^T \\ \vdots \\ \mathbf{n}_{p-r,1}^T \end{pmatrix} = \mathbf{C} \boldsymbol{\alpha}_\perp^T = \left(\boldsymbol{\alpha}_\perp^T \mathbf{\Gamma} \boldsymbol{\beta}_\perp \right)^{-1} \boldsymbol{\alpha}_\perp^T,$$

where the normalization matrix $\mathbf{C} \in \mathbb{R}^{p-r \times p-r}$ is chosen so that Eq. 3.26 is satisfied. For $\mathbf{m}_{1,p-r+1,\ldots,p} = \boldsymbol{\alpha}$, the corresponding $\mathbf{n}_{p-r+1\ldots p,1}$ are computed so that:

$$\boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}_\perp^T \mathbf{\Gamma} \boldsymbol{\beta}_\perp \right)^{-1} \boldsymbol{\alpha}_\perp^T + \boldsymbol{\alpha} \mathbf{n}_{p-r+1\ldots p,1}^T = \mathbb{I}$$

implying that $\mathbf{n}_{p-r+1\ldots p,1}^T = \left(\boldsymbol{\beta}^T \boldsymbol{\alpha} \right)^{-1} \boldsymbol{\beta}^T$. Therefore, the eigendecomposition of $\mathbf{A}$ is equal

to:

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\beta}_\perp & \boldsymbol{\alpha} \end{pmatrix} \begin{pmatrix} \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix} \begin{pmatrix} \left(\boldsymbol{\alpha}_\perp^T \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}_\perp^T \\ \left(\boldsymbol{\beta}^T \boldsymbol{\alpha}\right)^{-1} \boldsymbol{\beta}^T \end{pmatrix},$$

where the diagonal matrix $\mathbf{D}_2$ contains all of the non-unit eigenvalues.

Proof of Theorem 3.2.

First, it is shown that Restriction 3.2 implies Restriction 3.1. In section A.2.1, the block structure for the eigenvectors of a vector autoregressive system is derived, as given by Eq. A.6. According to Eq. A.7, the *base vector* satisfies:

$$\left(\lambda_i^{-1}\boldsymbol{\Pi}_1 + \lambda_i^{-2}\boldsymbol{\Pi}_2 + \cdots + \lambda_i^{-k}\boldsymbol{\Pi}_k - \mathbb{I}_p\right)\mathbf{m}_{1,i} \;=\; 0.$$

According to Restriction 3.2, in a cointegrated VAR system, there is at least one eigenvalue at $\lambda_i = 1$, so that the following must be true:

$$\begin{aligned} \left(\boldsymbol{\Pi}_1 + \boldsymbol{\Pi}_2 + \cdots + \boldsymbol{\Pi}_k - \mathbb{I}_p\right)\mathbf{m}_{1,i} &= 0, \\ \boldsymbol{\Pi}\mathbf{m}_{1,i} &= 0, \end{aligned} \tag{3.27}$$

where the last line follows from Eq. 3.6. Thus, it is immediately clear that $\boldsymbol{\Pi}$ has reduced rank, since $\mathbf{m}_{1,i}$ is a non-trivial member of the nullspace of $\boldsymbol{\Pi}$. If the matrix $\mathbf{A}$ has more than one unit eigenvalue, then the rank of $\boldsymbol{\Pi}$ is equal to the rank of the matrix formed by concatenating the base blocks $\mathbf{m}_{1,i}$ for each of the corresponding eigenvectors. Let r denote the rank of the matrix $\boldsymbol{\Pi}$, which implies that the matrix $\mathbf{A}$ has at least $p - r$ unit eigenvalues. Since $\boldsymbol{\Pi}$ has rank $r < p$ it can be factored according to Eq. 3.8, and indeed Restriction 3.2 for a cointegrated system implies Restriction 3.1.

To establish the converse, observe that the eigenvalues of $\mathbf{A}$ satisfy $\det(\mathbf{A} - \lambda\mathbb{I}_p) = \mathbf{0}$, which is equivalently given by the expression in Eq. A.4. Since $\boldsymbol{\Pi}$ has reduced rank and therefore its determinant equals zero, $\lambda_i = 1$ is clearly a solution. Hence Restriction 3.1 implies Restriction 3.2.

Proof of Theorem 3.3.

First, consider the condition to guarantee that each component process is at most $I(1)$, in terms of the eigenspace associated with the unit eigenvalues. When the geometric multiplicity is less than the algebraic multiplicity, there will be at least one Jordan block of the form:

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}. \tag{3.28}$$

As a result, when the transfer function if computed according to Eq. A.26, the inverse term will contain a block of the form:

$$\begin{pmatrix} \frac{1}{(z-1)} & \frac{1}{(z-1)^2} & \cdots & \frac{1}{(z-1)^k} \\ 0 & \frac{1}{(z-1)} & \ddots & \vdots \\ 0 & 0 & \ddots & \frac{1}{(z-1)^2} \\ 0 & 0 & 0 & \frac{1}{(z-1)} \end{pmatrix} \text{[21]}$$

and higher-order integrators will appear at the output, which can only be prevented by requiring that the geometric and algebraic multiplicities agree.

Second, consider the condition to guarantee that each component process is at most $I(1)$, in terms of the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. When the geometric and algebraic multiplicities do not agree, the two generalized eigenvectors , as defined in equations A.20 and A.6, are related as follows:

$$\begin{aligned} (\mathbf{A} - \lambda\mathbb{I})\,\mathbf{m}_2 &= \mathbf{m}_1 \\ \begin{pmatrix} \boldsymbol{\Pi}_1 - \lambda_i\mathbb{I} & \boldsymbol{\Pi}_2 & \cdots & \boldsymbol{\Pi}_{k-1} & \boldsymbol{\Pi}_k \\ \mathbb{I}_p & -\lambda_i\mathbb{I}_p & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\lambda_i\mathbb{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbb{I}_p & -\lambda_i\mathbb{I}_p \end{pmatrix} \begin{pmatrix} \mathbf{m}_{1,2} \\ \mathbf{m}_{1,2} - \mathbf{m}_{1,1} \\ \vdots \\ \mathbf{m}_{1,2} - (k-2)\mathbf{m}_{1,1} \\ \mathbf{m}_{1,2} - (k-1)\mathbf{m}_{1,1} \end{pmatrix} &= \begin{pmatrix} \mathbf{m}_{1,1} \\ \mathbf{m}_{1,1} \\ \vdots \\ \mathbf{m}_{1,1} \\ \mathbf{m}_{1,1} \end{pmatrix} \end{aligned}$$

Multiplying out the top row yields:

$$\begin{aligned}
\mathbf{m}_{1,1} &= (\mathbf{\Pi}_1 - \lambda_i \mathbb{I})\,\mathbf{m}_{2,1} + \mathbf{\Pi}_2\,(\mathbf{m}_{2,1} - \mathbf{m}_{1,1}) \\
&\quad + \cdots + \ \mathbf{\Pi}_{k-1}\,(\mathbf{m}_{1,2} - (k-2)\mathbf{m}_{1,1}) + \mathbf{\Pi}_k\,(\mathbf{m}_{1,2} - (k-1)\mathbf{m}_{1,1}) \\
\mathbf{0} &= (\mathbf{\Pi}_1 + \mathbf{\Pi}_2 + \cdots + \mathbf{\Pi}_{k-1} + \mathbf{\Pi}_k - \lambda_i \mathbb{I})\,\mathbf{m}_{1,2} \\
&\quad - (\mathbf{\Pi}_2 + 2\mathbf{\Pi}_3 + \cdots + (k-2)\mathbf{\Pi}_{k-1} + (k-1)\mathbf{\Pi}_k - \mathbb{I})\,\mathbf{m}_{1,1} \\
\mathbf{0} &= \mathbf{\Pi}\mathbf{m}_{1,2} - \mathbf{\Gamma}\mathbf{m}_{1,1} \\
\mathbf{\Gamma}\mathbf{m}_{1,1} &= \mathbf{\Pi}\mathbf{m}_{1,2} \\
\mathbf{\Gamma}\mathbf{m}_{1,1} &= \boldsymbol{\alpha}\boldsymbol{\beta}^T\mathbf{m}_{1,2} \\
\boldsymbol{\alpha}_\perp^T\mathbf{\Gamma}\mathbf{m}_{1,1} &= \mathbf{0} \\
\boldsymbol{\alpha}_\perp^T\mathbf{\Gamma}\boldsymbol{\beta}_\perp &= \mathbf{0}.
\end{aligned} \tag{3.29}$$

3.B Singular Value Decompositon of $\mathbf{\Pi}$

Recall from Eq. 3.8 that the reduced rank matrix $\mathbf{\Pi}$ can be factored according to $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^T$ where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{p\times r}$. The matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be defined according to the singular value decomposition (SVD) of $\mathbf{\Pi}$, as derived here.

The SVD of $\mathbf{\Pi}$ is given by:

$$\mathbf{\Pi} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

where $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{p\times p}$ are unitary matrices and $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ is a diagonal matrix of singular values. The columns of the matrices $\mathbf{U}$ and $\mathbf{V}^T$ contain basis vectors for the *four fundamental subspaces* of $\mathbf{\Pi}$. The first r columns of $\mathbf{U}$ span the *column space* or *range* of $\mathbf{\Pi}$, while the last $p-r$ columns span the *left nullspace*. The first r columns of $\mathbf{V}^T$ span the *row space* of $\mathbf{\Pi}$ or range of $\mathbf{\Pi}^T$, while the last $p-r$ columns span the *right nullspace*. Under the assumption that $\mathbf{\Pi}$ has reduced rank $r < p$, onlt the first r singular values, denoted by $\{\sigma_1, \cdots, \sigma_r\}$, are non-zero. The matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are defined by:

$$\begin{aligned}
\mathbf{\Pi} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{pmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_p\end{pmatrix} \begin{pmatrix}\sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \mathbf{0} & \end{pmatrix} \begin{pmatrix}\mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \\ \mathbf{v}_{r+1}^T \\ \vdots \\ \mathbf{v}_p^T\end{pmatrix} \\
&= \underbrace{\begin{pmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_r\end{pmatrix} \begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{pmatrix}^{1/2}}_{\boldsymbol{\alpha}} \underbrace{\begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{pmatrix}^{1/2} \begin{pmatrix}\mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T\end{pmatrix}}_{\boldsymbol{\beta}^T} \qquad (3.30)
\end{aligned}$$

The matrix $\boldsymbol{\alpha}$ is equal to the non-zero columns of $\mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}$ and the matrix $\boldsymbol{\beta}^T$ is equal the non-zero rows of $\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{V}^T$. Equation 3.30 also indicates that the columns of $\boldsymbol{\alpha}$ form basis for the column space of $\mathbf{\Pi}$, while the columns of $\boldsymbol{\beta}$ form a basis for the row space of $\mathbf{\Pi}$. Hence, the columns of $\boldsymbol{\beta}_\perp$ form a basis for the nullspace of $\mathbf{\Pi}$ and the columns of $\boldsymbol{\alpha}_\perp$ form a basis for the left nullspace of $\mathbf{\Pi}$.

Chapter 4

Static Portfolio Choice

This chapter considers static multistage portfolio construction over an N-stage investment horizon, where inter-stage portfolio rebalancing is prohibited. It is assumed that the log-prices of the assets evolve according to a cointegrated vector autoregressive process, as described in the previous chapter. The investor makes a single decision at the beginning of the first stage and the resulting portfolio is held without modification for all N periods. This scenario often occurs in the real world, as professional portfolio managers and individual investors frequently hold a position for extended time periods (e.g., weeks or months) despite the availability of daily (or intra-day) price data and market news.

Two investment strategies for the static asset allocation problem are presented and compared. In the first strategy, the portfolio vector is chosen from within the span of the cointegrating space (i.e., span$\{\boldsymbol{\beta}\}$). Such portfolios encode the belief that the current log-prices represent a state of relative asset mispricing (i.e., temporary disequilibrium), and that the prices will revert back to their steady-state levels within the attractor (i.e., equilibrium) space. While these portfolios are popular among industry practitioners, it is shown here that they are not mean-variance optimal, in the sense of Section 2.3.1. Thus, a second strategy is considered, in which the additional requirement is imposed that the investor act in accordance with the Markowitz mean-variance framework. Given this preference function, closed-form solutions for the mean-variance optimal (MVO) static portfolios are derived as functions of the cointegrated VAR model parameters and investment horizon length. The resulting asset allocation vectors are interpreted geometrically within the context of the cointegrated VAR model. Of particular interest is the relationship between the MVO portfolio weight vector and the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ subspaces, both for fixed N and in the limit as N increases. It is shown that only in the limit of an infinite trading horizon does the MVO

portfolio equal the Beta portfolio. For finite N, there is an additional, positive expected return to be gained from choosing the portfolio not only in the direction of finite variance, but also in the direction of expected change.

The organization of this chapter is as follows. In Section 4.1 the static Beta portfolio is presented. Next, in Section 4.2, the static Markowitz mean-variance asset allocation problem is presented, and the MVO portfolio weight vector is derived in this setting, both with and without a budget constraint. In Section 4.3, the properties of these portfolios are further explored. In particular, in Section 4.3.1, the statistics of the per-stage portfolio returns are characterized; in Section 4.3.2, the ability of the MVO solution to capture short-term predictability is analyzed; in Section 4.3.3, the asymptotic properties of the MVO solution are derived; and lastly, in Section 4.3.4, the mean-variance tradeoff is explored through the use of a leverage constraint.

4.1 The Beta Portfolio

Consider the static portfolio choice problem, where a single portfolio, $\mathbf{w}$, is constructed and held without modification over an investment time horizon of N stages. Furthermore, assume that the asset log-prices, $\mathbf{x}_k \in \mathbb{R}^p$, are well-modeled by a first-order[1] cointegrated vector autoregressive process, detailed in Chapter 3, as follows:

$$\mathbf{x}[n+1] = \mathbf{\Pi}_1 \mathbf{x}[n] + \mathbf{\Phi d}[n] + \boldsymbol{\epsilon}[n], \tag{4.1}$$

with stochastic input $\boldsymbol{\epsilon}[n] \sim N(\mathbf{0}, \mathbf{\Psi})$, constant deterministic input $\mathbf{\Phi d}[n] = \boldsymbol{\phi}$, and initial condition, $\mathbf{x}_0$. The vector $\boldsymbol{\phi}$, also referred to as the *drift* term, is included in order to capture the overall linear growth trend present in the historical log-prices of most assets. The corresponding log-return of the assets after N stages is defined as:

$$\mathbf{r}[N] = \mathbf{x}[N] - \mathbf{x}[0].$$

[1] While throughout this chapter primary attention is given to the first-order VAR system, extensions to higher-order systems, achieved by augmenting the state space, are also briefly discussed.

The variance of both $\mathbf{x}[N]$ and $\mathbf{r}[N]$ is given by:

$$\text{var}\left[\mathbf{r}[N]\right] = \text{var}\left[\mathbf{x}[N]\right] = \sum_{k=1}^{N} \mathbf{\Pi}_1^{N-k} \mathbf{\Psi} \mathbf{\Pi}_1^{(N-k)T}.$$

Due to the cointegration assumption, the state transition matrix $\mathbf{A} = \mathbf{\Pi}_1$ contains at least one eigenvalue at unity. Therefore, the process variance grows linearly over time, and the variance of any estimator of future log-prices or returns is asymptotically unbounded in the limit as the length of investment horizon increases.

While $\mathbf{x}_N$ is difficult to predict, it is known that in the long-run the zero-input response (ZIR), derived in Section 3.3.1, converges to a fixed point within the attractor space, $\text{span}\{\boldsymbol{\beta}_\perp\}$. Consequently, it is often suggested [2, 31] that the portfolio weight vector be selected in the direction of the orthogonal projection from $\mathbf{x}[0]$ to this space (i.e., $\mathbf{w} \in \text{span}\{\boldsymbol{\beta}\}$), appropriately earning it the name *Beta portfolio*. The realized return, $r[n] = \boldsymbol{\beta}^T \mathbf{r}[n]$, evolves according to an AWSS random process, and subsequently the Beta portfolio is commonly referred to as the portfolio with asymptotically convergent variance.

The following example illustrates how the Beta portfolio is constructed for a system of three assets with up to two integrators.

Example 4.1.

Consider a system of three risky assets where the log-prices are assumed to evolve according to a cointegrated VAR process without drift, as follows:

$$\mathbf{x}[n+1] = \mathbf{\Pi}_1 \mathbf{x}[n] + \boldsymbol{\epsilon}[n].$$

First, assume the eigendecomposition for the matrix $\mathbf{\Pi}_1$ is given by:

$$\mathbf{\Pi}_1 = \mathbf{M}\mathbf{D}\mathbf{M}^{-1} = \begin{pmatrix} 0.1533 & -0.4674 & -0.3198 \\ 0.7832 & -0.8826 & -0.8160 \\ 0.6026 & -0.0498 & 0.4815 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0.7 \end{pmatrix} \begin{pmatrix} 0.1533 & -0.4674 & -0.3198 \\ 0.7832 & -0.8826 & -0.8160 \\ 0.6026 & -0.0498 & 0.4815 \end{pmatrix}^{-1}.$$

This system contains a single unit eigenvalue (i.e., one common trend), implying that the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which define the error-correcting and cointegrating subspaces, respec-

tively, span two-dimensional subspaces, as depicted in Figure 4-1(a). Furthermore, the attractor space, defined by the matrix $\boldsymbol{\beta}_\perp$, spans a one-dimensional subspace. The direction of the Beta portfolio, $\mathbf{w}$, is chosen by computing the orthogonal projection of the current state of the system, $\mathbf{x}_0$, back to the attractor, such that $\mathbf{w} \in \text{span}\{\boldsymbol{\beta}\}$. While any vector from within the cointegrating space forms a portfolio whose value has finite variance, only the vector corresponding to the orthogonal projection of $\mathbf{x}_0$ onto $\text{span}\{\boldsymbol{\beta}_\perp\}$ encodes the belief that the current log-prices represent a state of temporary disequilibrium, or relative asset mispricing, and that over time the process will revert back to equilibrium, settling at a point within the attractor space.

Second, consider a system where the eigendecomposition for the matrix $\boldsymbol{\Pi}_1$ contains two unit eigenvalues, as follows:

$$\boldsymbol{\Pi}_1 = \mathbf{MDM}^{-1} = \begin{pmatrix} -0.4674 & -0.3198 & 0.1533 \\ -0.8826 & -0.8160 & 0.7832 \\ -0.0498 & 0.4815 & 0.6026 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.7 \end{pmatrix} \begin{pmatrix} -0.4674 & -0.3198 & 0.1533 \\ -0.8826 & -0.8160 & 0.7832 \\ -0.0498 & 0.4815 & 0.6026 \end{pmatrix}^{-1}.$$

The underlying geometry for this system is shown in Fig. 4-1(b). Here, the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ define one-dimensional subspaces of $\mathbb{R}^3$, while the matrix $\boldsymbol{\beta}_\perp$ defines a two-dimensional attractor space. While there are two choices for the portfolio direction (i.e., $\mathbf{w} = \pm\boldsymbol{\beta}$), the one corresponding to the projection of $\mathbf{x}[0]$ onto $\text{span}\{\boldsymbol{\beta}_\perp\}$ is chosen. ■

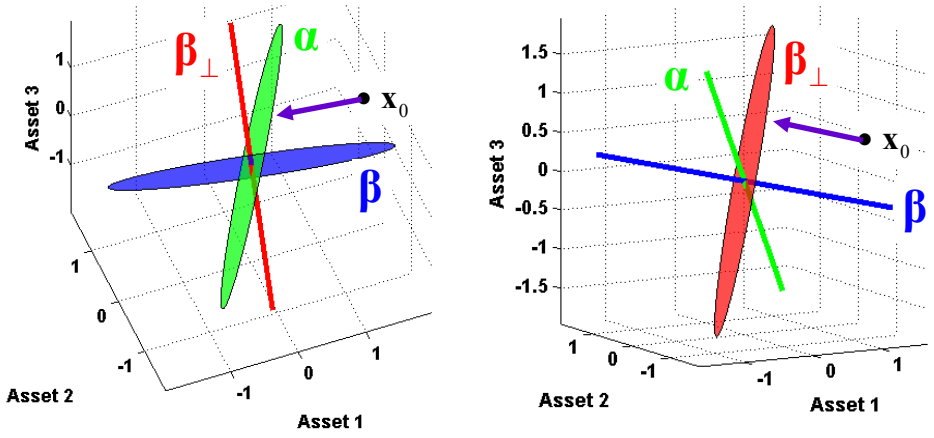
As this example illustrates, the justification for the Beta portfolio is readily understood using a geometric interpretation of a cointegrated VAR process. The change in the asset log-prices, $\Delta\mathbf{x} = \mathbf{x}_N - \mathbf{x}_0$, can be expanded in a $\{\boldsymbol{\beta}, \boldsymbol{\beta}_\perp\}$ basis, as follows:

$$\Delta\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2,$$

where $c_1, c_2 \in \mathbb{R}, \mathbf{b}_1 \in \text{span}\{\boldsymbol{\beta}\}, \mathbf{b}_2 \in \text{span}\{\boldsymbol{\beta}_\perp\}$. The corresponding total log-return is computed as:

$$r_N = \mathbf{w}^T\Delta\mathbf{x} = c_1\boldsymbol{\beta}^T\boldsymbol{\beta},$$

and therefore the Beta portfolio investor only earns profit due to the component of $\Delta\mathbf{x}$ in the $\boldsymbol{\beta}$ direction, and is indifferent to movement along the $\boldsymbol{\beta}_\perp$ direction. While the portfolio



(a) One common trend. (b) Two common trends.

Figure 4-1. Portfolio weight vectors in the three asset system considered in Example 4.1. The portfolio is chosen by computing the orthogonal projection of the current state of the system, $\mathbf{x}_0$, back to the attractor, $\text{sp}\{\boldsymbol{\beta}_\perp\}$, which corresponds to a vector from within the span of the cointegrating space, $\text{sp}\{\boldsymbol{\beta}\}$.

direction is initially chosen independent of any investor preference function, there is an implicit assumption that the investor prefers to *hedge against* uncertainty in the shared common trends of the underlying assets. However, for finite N, such portfolios are not mean-variance optimal in the sense of Section 2.3.1, as discussed next.

4.2 Mean-Variance Optimal Portfolio Construction

Consider again the static portfolio choice problem of Section 4.1, where now the additional assumption is imposed that the investor acts in accordance with the Markowitz mean-variance framework of Section 2.3.1. Specifically, the objective is to maximize the expected value of the cumulative portfolio log-return over an investment length of N stages, given an equality constraint on the corresponding portfolio return variance. Formally, the static mean-variance portfolio choice problem, P_0, is given by:

$$\left.\begin{aligned} \{\mathbf{w}_N^*\} = \underset{\mathbf{w}_N}{\arg\max} \;\; & \mathbf{w}_N^T \boldsymbol{\mu}_{\mathbf{r}_N} \\ \text{s.t.} \;\; & \mathbf{w}_N^T \boldsymbol{\Sigma}_{\mathbf{r}_N} \mathbf{w}_N = \sigma_0^2 \end{aligned}\right\} \; P_0,$$

where the N stage log-return vector, $\mathbf{r}_N \triangleq \mathbf{x}_N - \mathbf{x}_0$, is a Gaussian random vector with mean $E_{t_0}[\mathbf{r}_N] = \boldsymbol{\mu}_{\mathbf{r}_N}$ and covariance matrix $\text{var}_{t_0}[\mathbf{r}_N] = \boldsymbol{\Sigma}_{\mathbf{r}_N}$. The expectation and variance

operators are taken with respect to the information available at the start of the investment horizon, denoted as t_0. Both leveraged and non-leveraged portfolios are considered, in Sections 4.2.1 and 4.2.2 respectively, with the latter case achieved by adding a budget constraint of the form $\mathbf{w}_N^T\mathbf{1} = 1$ to problem P_0. Without the budget constraint, the net leverage (i.e., size) of the portfolio is limited only by the allowable risk parameter, σ_0.

4.2.1 Case 1: No Budget Constraint

Let $\lambda_L \in \mathbb{R}$ be a Lagrange multiplier, so that problem P_0 can be rewritten as follows:

$$\left.\{\mathbf{w}_N^*, \lambda_L^*\} = \underset{\mathbf{w}_N, \lambda_L}{\arg\max}\ \mathbf{w}_N^T\boldsymbol{\mu}_{\mathbf{r}_N} - \lambda_L\left(\mathbf{w}_N^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{w}_N - \sigma_0^2\right) \quad \right\}\ P_0'.$$

The objective function is quadratic in $\mathbf{w}_N$, and admits the well-known closed-form solution:

$$\mathbf{w}_N^* = \frac{1}{2\lambda_L^*}\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N} = \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}_{\mathbf{r}_N}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}}}\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}. \tag{4.2}$$

It is shown that for finite N, there is an additional, positive expected return to be gained from choosing the portfolio weight vector not only in the $\boldsymbol{\beta}$ direction, but also with a component in the direction of expected change (i.e., the $\boldsymbol{\mu}_{\mathbf{r}_N}$ direction).

Solving for $\boldsymbol{\mu}_{\mathbf{r}_N}$ The mean of the cumulative return over N stages is computed according to the cointegrated VAR model of Eq. 4.1 and is given by:

$$\boldsymbol{\mu}_{\mathbf{r}_N} = E\left[\mathbf{r}_N\right] = E\left[\mathbf{x}_N - \mathbf{x}_0\right] = \left(\boldsymbol{\Pi}_1^N - \mathbb{I}\right)\mathbf{x}_0 + \left(\sum_{i=0}^{N-1}\boldsymbol{\Pi}_1^i\right)\boldsymbol{\phi}. \tag{4.3}$$

When this cointegrated VAR process is expressed in state-space form, the state transition matrix is equal to $\mathbf{A} = \boldsymbol{\Pi}_1$, with corresponding eigendecomposition $\mathbf{A} = \mathbf{M}\mathbf{D}\mathbf{M}^{-1}$. Furthermore, assume that $\mathbf{A}$ has exactly one unit eigenvalue (*i.e.*, $\lambda_1 = 1$), while the remaining $p-1$ eigenvalues lie strictly inside the unit circle (i.e., $|\lambda_i| < 1,\ i = 2, ..., p$). Letting $\mathbf{m}_i$ denote the columns of the matrix $\mathbf{M}$ and $\mathbf{n}_i^T$ denote the rows of the matrix $\mathbf{M}^{-1}$, the expected

cumulative N-stage return is equal to:

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{r}_N} &= \mathbf{M}\left(\mathbf{D}^N - \mathbb{I}\right)\mathbf{M}^{-1}\mathbf{x}_0 + \mathbf{M}\left(\sum_{i=0}^{N-1}\mathbf{D}^i\right)\mathbf{M}^{-1}\boldsymbol{\phi}, \\
&= N\mathbf{m}_1\mathbf{n}_1^T\boldsymbol{\phi} + \sum_{i=2}^{p}\left(\left(\lambda_i^N - 1\right)\mathbf{m}_i\mathbf{n}_i^T\mathbf{x}_0 + \frac{1-\lambda_i^N}{1-\lambda_i}\mathbf{m}_i\mathbf{n}_i^T\boldsymbol{\phi}\right), \qquad (4.4) \\
&= \sum_{i=1}^{p} c_i\mathbf{m}_i, \qquad (4.5)
\end{aligned}$$

where $c_1 = N\mathbf{n}_1^T\boldsymbol{\phi}$ and $c_i = \mathbf{n}_i^T\left(\left(\lambda_i^N - 1\right)\mathbf{x}_0 + \frac{1-\lambda_i^N}{1-\lambda_i}\boldsymbol{\phi}\right)$ for $i = 2, ..., p$. As Eq. 4.5 shows, the expected change in the asset log-prices can be expressed as a linear combination of the eigenvectors of the matrix $\mathbf{A} = \boldsymbol{\Pi}_1$. When drift is present (i.e., $\boldsymbol{\phi} \neq \mathbf{0}$), the weight on the component in the direction $\mathbf{m}_1$ grows over time, while the weights on the remaining eigenvectors converge to finite quantities since $|\lambda_i| < 1$ for $i = 2, ..., p$. Recall from Theorem 3.1 that the vector $\mathbf{m}_1$ forms a basis for the $\boldsymbol{\beta}_\perp$ subspace, while the vectors $\{\mathbf{m}_2, ..., \mathbf{m}_p\}$ form a basis for the $\boldsymbol{\alpha}$ subspace. As Eq. 4.5 indicates, only in the absence of drift does the expected N stage return fall within the $\boldsymbol{\alpha}$ subspace.

In the case of higher-order VAR systems, $\boldsymbol{\mu}_{\mathbf{r}_N}$ can be readily computed by augmenting the state-space according to Eq. A.2. First, define:

$$\begin{aligned}
\mathbf{x}_{N:N-k+1} &\triangleq \begin{pmatrix}\mathbf{x}_N & \mathbf{x}_{N-1} & \cdots & \mathbf{x}_{N-k+1}\end{pmatrix}^T \\
\boldsymbol{\phi}_{N:N-k+1} &\triangleq \begin{pmatrix}\boldsymbol{\phi} & \mathbf{0} & \cdots & \mathbf{0}\end{pmatrix}^T \\
\mathbf{T} &\triangleq \begin{pmatrix}\mathbb{I}_p & \mathbf{0}_{p\times(k-1)p}\end{pmatrix}
\end{aligned}$$

which implies that Eq. 4.3 can be rewritten as:

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{r}_N} &= \mathbf{T}\left(\mathbf{A}^N - \mathbb{I}\right)\mathbf{x}_{0:-(k-1)} + \mathbf{T}\left(\sum_{i=0}^{N-1}\mathbf{A}^i\right)\boldsymbol{\phi}_{0:-(k-1)} \\
&= \mathbf{TM}\left(\mathbf{D}^N - \mathbb{I}\right)\mathbf{M}^{-1}\mathbf{x}_{0:-(k-1)} + \mathbf{TM}\left(\sum_{i=0}^{N-1}\mathbf{D}^i\right)\mathbf{M}^{-1}\boldsymbol{\phi}_{0:-(k-1)}, \\
&= N\mathbf{m}_{1,i}\mathbf{n}_1^T\boldsymbol{\phi} + \sum_{i=2}^{p}\left(\lambda_i^N - 1\right)\mathbf{m}_{1,i}\mathbf{n}_i^T\mathbf{x}_0 + \frac{1-\lambda_i^N}{1-\lambda_i}\mathbf{m}_{1,i}\mathbf{n}_i^T\boldsymbol{\phi}, \\
&= \sum_{i=1}^{p} c_i\mathbf{m}_{1,i},
\end{aligned}$$

Thus, in higher-order VAR systems, the direction of expected change is equal to a linear combination of the *base blocks* of the eigenvectors of $\mathbf{A}$, denoted by $\mathbf{m}_{1,i}$, when represented in the block form of Eq. A.6.

Solving for $\boldsymbol{\Sigma}_{\mathbf{r}_N}$ The covariance matrix of the total return can be computed directly from Eq. 4.1 as follows:

$$\boldsymbol{\Sigma}_{\mathbf{r}_N} = \mathrm{var}\left[\mathbf{r}_N\right] = \mathrm{var}\left[\mathbf{x}_N - \mathbf{x}_0\right] = \sum_{i=0}^{N-1}\mathbf{A}^i\boldsymbol{\Psi}\left(\mathbf{A}^i\right)^T, \tag{4.6}$$

or can be computed recursively, according to the matrix difference equation:

$$\boldsymbol{\Sigma}_{\mathbf{r}_N} = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{r}_{N-1}}\mathbf{A}^T + \boldsymbol{\Psi}.$$

For higher order VAR systems, the covariance matrix is again computed by augmenting the state-space, as follows:

$$\boldsymbol{\Sigma}_{\mathbf{r}_N} = \mathrm{var}\left[\mathbf{r}_N\right] = \mathrm{var}\left[\mathbf{x}_N - \mathbf{x}_0\right] = \sum_{i=0}^{N-1}\mathbf{T}\mathbf{A}^i\widetilde{\boldsymbol{\Psi}}\left(\mathbf{A}^i\right)^T\mathbf{T}^T, \tag{4.7}$$

where the augmented covariance matrix $\widetilde{\boldsymbol{\Psi}} \in \mathbb{R}^{pk\times pk}$ is defined as:

$$\widetilde{\boldsymbol{\Psi}} = \mathrm{diag}\begin{pmatrix}\boldsymbol{\Psi} & \mathbf{0} & \cdots & \mathbf{0}\end{pmatrix}. \tag{4.8}$$

As shown here, the first and second order statistics of the asset cumulative N stage log-returns are readily computed given only a model for the evolution of the asset log-prices

over time. The asymptotic properties of $\boldsymbol{\mu}_{\mathbf{r}_N}$ and $\boldsymbol{\Sigma}_{\mathbf{r}_N}$ are discussed in Section 4.3.3. The next section extends the analysis presented here for the case where a budget constraint is included in the mean-variance framework.

4.2.2 Case 2: With Budget Constraint

In order to ensure that the investor allocates all of the initial wealth among the available assets, a budget constraint of the form $\mathbf{w}^T\mathbf{1} = 1$ is introduced. Recall from Section 2.2.3, that such a constraint can be enforced using the following affine relation:

$$\begin{aligned} \mathbf{w} &= \mathbf{c} + \mathbf{D}\mathbf{v} \\ \mathbf{c} &= \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}^T, \quad \mathbf{D} = \begin{pmatrix} \mathbb{I}_{p-1} \\ -\mathbf{1}^T \end{pmatrix} \end{aligned}$$

and subsequently solving for the optimal reduced dimension portfolio weight vector, $\mathbf{v}^*$. Under this simple transformation, Problem P_0 can be rewritten as Problem P_0'', as follows:

$$\left.\begin{aligned} \mathbf{v}_N^* = \arg\max_{\mathbf{v}_N} \ & \left(\mathbf{c} + \mathbf{D}\mathbf{v}_N\right)^T \boldsymbol{\mu}_{\mathbf{r}_N} \\ \text{s.t.} \ & \left(\mathbf{c} + \mathbf{D}\mathbf{v}_N\right)^T \boldsymbol{\Sigma}_{\mathbf{r}_N} \left(\mathbf{c} + \mathbf{D}\mathbf{v}_N\right) = \sigma_0^2, \end{aligned}\right\} \ P_0'',$$

which, in turn, can be expressed using Lagrange multiplier λ_L, as follows:

$$\mathbf{v}_N^*, \lambda_L^* = \arg\max_{\mathbf{v}_N} \ \left(\mathbf{c} + \mathbf{D}\mathbf{v}_N\right)^T \boldsymbol{\mu}_{\mathbf{r}_N} - \lambda_L \left(\left(\mathbf{c} + \mathbf{D}\mathbf{v}_N\right)^T \boldsymbol{\Sigma}_{\mathbf{r}_N} \left(\mathbf{c} + \mathbf{D}\mathbf{v}_N\right) - \sigma_0^2 \right).$$

Taking the derivative with respect to $\mathbf{v}_N$ and setting the result equal to zero yields:

$$\mathbf{v}_N^* = \frac{1}{2\lambda_L^*} \left(\mathbf{D}^T \boldsymbol{\Sigma}_{\mathbf{r}_N} \mathbf{D}\right)^{-1} \mathbf{D}^T \left(\boldsymbol{\mu}_{\mathbf{r}_N} - 2\lambda_L^* \boldsymbol{\Sigma}_{\mathbf{r}_N} \mathbf{c}\right),$$

which implies that the optimal portfolio weight vector, $\mathbf{w}_N^*$, is given by:

$$\mathbf{w}_N^* = \mathbf{c} + \frac{1}{2\lambda_L^*} \mathbf{D} \left(\mathbf{D}^T \boldsymbol{\Sigma}_{\mathbf{r}_N} \mathbf{D}\right)^{-1} \mathbf{D}^T \left(\boldsymbol{\mu}_{\mathbf{r}_N} - 2\lambda_L^* \boldsymbol{\Sigma}_{\mathbf{r}_N} \mathbf{c}\right).$$

In order to determine λ_L^*, first let $\mathbf{w}_N = \frac{1}{2\lambda_L}\mathbf{f} + \mathbf{g}$, where:

$$\begin{aligned}\mathbf{f} &= \mathbf{D}\left(\mathbf{D}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{D}\right)^{-1}\mathbf{D}^T\boldsymbol{\mu}_{\mathbf{r}_N},\\ \mathbf{g} &= \mathbf{c} - \mathbf{D}\left(\mathbf{D}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{D}\right)^{-1}\mathbf{D}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{c}.\end{aligned}$$

The relationship between λ_L and σ_0^2 is given by the following quadratic equation:

$$\begin{aligned}\sigma_0^2 &= (\mathbf{c}+\mathbf{D}\mathbf{v}_N)^T\boldsymbol{\Sigma}_{\mathbf{r}_N}(\mathbf{c}+\mathbf{D}\mathbf{v}_N),\\ &= \left(\frac{1}{2\lambda_L}\mathbf{f}+\mathbf{g}\right)^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\left(\frac{1}{2\lambda_L}\mathbf{f}+\mathbf{g}\right),\\ &= \lambda_L^{-2}0.25\mathbf{f}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{f} + \lambda_L^{-1}\mathbf{f}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{g} + \mathbf{g}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{g}. \qquad (4.9)\end{aligned}$$

While there are two values of λ_L for every value of σ_0, one of the roots is extraneous, earning a lower expected value for a given level of risk than the other, and must be discarded. An alternative derivation of the budget-constrained solution, which uses an additional Lagrange multiplier rather than the linear transformation of Section 2.2.3, is given in Appendix 4.B.

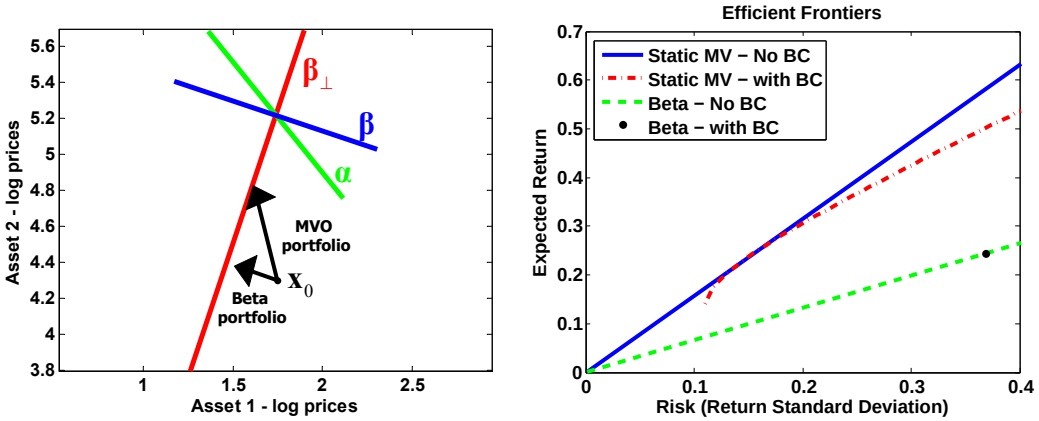
The following example illustrates the differences between the Beta and MVO solutions, both with and without the presence of a budget constraint.

Example 4.2.

Consider a system of two assets, in which the log-prices are assumed to evolve over an investment horizon of two stages according to the first-order cointegrated VAR system of Eq. 4.1, with:

$$\boldsymbol{\Pi}_1 = \begin{pmatrix}0.7878 & 0.0707\\ 0.2634 & 0.9122\end{pmatrix} \quad \text{and} \quad \boldsymbol{\Psi} = \begin{pmatrix}0.2^2 & 0\\ 0 & 0.07^2\end{pmatrix}.$$

The initial value of the log-prices is arbitrarily chosen as $\mathbf{x}_0 = \begin{pmatrix}1.75 & 4.3\end{pmatrix}^T$ and the drift is assumed to be $\boldsymbol{\phi} = \mathbf{0}$. The underlying geometry of this system is depicted in Fig. 4-2(a), and the efficient frontiers for a two-stage trading strategy are shown in Fig. 4-2(b), both for the mean-variance optimal and Beta strategies. The two curves corresponding to the mean-variance optimal portfolios, with and without a budget constraint, achieve a higher expected return for a given level of risk (standard deviation) as compared to the efficient



(a) Underlying cointegrated VAR system geometry in Example 4.2.

(b) Efficient Frontiers for Beta vs. MVO portfolios in two stage, two asset example.

Figure 4-2. Efficient Frontiers for Beta vs. MVO portfolios in two stage, two asset example.

frontiers for the Beta portfolio. As there is only one scaling of the $\boldsymbol{\beta}$ vector that also satisfies $\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{r}_N} \boldsymbol{\beta} = \sigma_0^2$, the corresponding efficient frontier only contains a single point, as shown. As Fig. 4-2(b) illustrates, the efficient frontier corresponding to the case where the budget constraint is omitted is tangent to the curve with the budget constraint included. This is due to the fact that omission of the budget constraint implies that the investor can borrow from a risk-free source for free (i.e., $r_f = 0$), thereby achieving any convex combination of a portfolio of the risk-free asset and a portfolio of all risky assets. In this case, the optimal portfolio of risky assets the investor seeks is the one that maximizes the slope of the efficient frontier line, defined by the line tangent to the red curve that goes through the origin. ■

As Ex. 4.2 illustrates, the two MVO solutions significantly outperform the Beta portfolio, as they capitalize not only on expected movements in the direction of $\boldsymbol{\beta}$, but also take advantage of the short-term predictability of a cointegrated VAR process along the $\boldsymbol{\alpha}$ direction.

4.3 Portfolio Properties

The following section explores the properties of the static Beta and MVO portfolio weight vectors, both with and without a budget constraint through a series of illustrative examples. First, in Section 4.3.1, the per-stage portfolio return distributions are derived, and the inter-

stage return correlation structure is explored. Second, in Section 4.3.2, the nature of the predictable component in a cointegrated VAR is quantified. Third, in Section 4.3.3, a complete asymptotic analysis of the MVO solution is presented, and the relationship between the direction of the optimal portfolio weight vector and the subspaces of a cointegrated VAR is revealed. Finally, in Section 4.3.4, the mean-variance tradeoff is explored.

4.3.1 Per-stage Return Statistics

As stated in Section 4.2, the random vector corresponding to the N stage individual asset returns is Gaussian with mean $\boldsymbol{\mu}_{\mathbf{r}_N}$, defined in Eq. 4.3, and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{r}_N}$, defined in Eq. 4.6. In static asset allocation, the portfolio weight vector, $\mathbf{w}_N$, is determined by the information available at time t_0, and therefore the cumulative return on the portfolio, r_N, is a scalar Gaussian random variable with mean $\mathbf{w}^T\boldsymbol{\mu}_{\mathbf{r}_N}$ and variance $\mathbf{w}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{w}$. In a two-stage problem, the per-stage portfolio returns are denoted by $r_1 = \mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_0)$ and $r_2 = \mathbf{w}^T(\mathbf{x}_2 - \mathbf{x}_1)$. Each of these returns is Gaussian, with distributions parametrized as follows:

$$\begin{aligned} r_1 &\sim N\left(\mathbf{w}^T\boldsymbol{\Pi}\mathbf{x}_0, \mathbf{w}^T\boldsymbol{\Psi}\mathbf{w}\right) \\ r_2 &\sim N\left(\mathbf{w}^T\boldsymbol{\Pi}\boldsymbol{\Pi}_1\mathbf{x}_0, \mathbf{w}^T\left(\boldsymbol{\Psi} + \boldsymbol{\Pi}\boldsymbol{\Psi}\boldsymbol{\Pi}^T\right)\mathbf{w}\right) \end{aligned}$$

Furthermore, the covariance between r_1 and r_2 is given by:

$$\begin{aligned} \operatorname{cov}[r_1, r_2] &= E[r_1 r_2] - E[r_1]E[r_2] \\ &= E\left[\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_0)(\mathbf{x}_2 - \mathbf{x}_1)^T\mathbf{w}\right] - E\left[\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_0)\right]E\left[\mathbf{w}^T(\mathbf{x}_2 - \mathbf{x}_1)\right] \\ &= \mathbf{w}^T E\left[(\boldsymbol{\Pi}\mathbf{x}_0 + \boldsymbol{\epsilon}_1)(\boldsymbol{\Pi}(\boldsymbol{\Pi}_1\mathbf{x}_0 + \boldsymbol{\epsilon}_1) + \boldsymbol{\epsilon}_2)^T\right]\mathbf{w} - \mathbf{w}^T\boldsymbol{\Pi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Pi}_1^T\boldsymbol{\Pi}^T\mathbf{w} \\ &= \mathbf{w}^T\left(\boldsymbol{\Pi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Pi}_1^T\boldsymbol{\Pi}^T + \boldsymbol{\Psi}\boldsymbol{\Pi}^T\right)\mathbf{w} - \mathbf{w}^T\boldsymbol{\Pi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Pi}_1^T\boldsymbol{\Pi}^T\mathbf{w} \\ &= \mathbf{w}^T\boldsymbol{\Psi}\boldsymbol{\Pi}^T\mathbf{w}. \end{aligned} \tag{4.10}$$

Thus, in static portfolio choice, the inter-stage return covariance is a simple quadratic form of the portfolio weight vector. The following example compares the per-stage portfolio return statistics between the Beta and MVO portfolios.

Example 4.3.

Consider again the system of two assets from Ex. 4.2. The Beta portfolio that also satisfies the budget constraint of $\mathbf{w}^T\mathbf{1} = 1$, corresponds to a total portfolio return standard deviation of $\sigma_0 = 0.37$. Table 4.1 displays the per-stage return statistics for all three strategies when executed at this risk level, and reveals two interesting characteristics of static portfolios. First, for the two strategies with no budget constraint, there is a direct relationship between the total expected return and the degree of negative correlation between the inter-stage portfolio returns. Since the total return, r_N, is defined as the sum of the per-stage log-returns, the total variance is computed as:

$$\text{var}\left[r_N\right] = \text{var}\left[r_1\right] + \text{var}\left[r_2\right] + 2\text{cov}\left[r_1, r_2\right].$$

Thus, higher negative correlation implies that the per-stage variances can increase, while the total variance remains constant.

The second property of static portfolios revealed in Table 4.1 is that the increased per-stage portfolio return variance is realized through the use of leverage. As the portfolio weights for the MVO solution with the budget constraint indicate (Table 4.1, line 3), the investor enters into a long position in asset 2 using 310% of the initial wealth, which is, in turn, financed by a short position in asset 1 using 210% of the initial wealth. The presence of an explicit risk-free asset is not required since the net position is unlevered. However, while this trading scheme is able to outperform the Beta strategy, it is the MVO solution without the budget constraint that achieves the highest expected return of all three strategies, due its use of *both* high negative inter-stage return correlation and net leverage. ■

As Ex. 4.3 illustrates, both the Beta and MVO portfolios exhibit negative inter-stage

Trading Strategy	**Budget Constraint**	**Portfolio Weights**			**Stage 1,** r_1		**Stage 2,** r_2		**Correlation**	**Total,** r_T	
		Asset 1	**Asset 2**	**Total Leverage**	**Mean**	**Std** (σ_1)	**Mean**	**Std** (σ_2)	$\rho = \frac{cov(r_1,r_2)}{\sigma_1\sigma_2}$	**Mean**	**Std** (σ_T)
Beta	YES	-1.50	0.50	1.00	0.14	0.30	0.10	0.33	-0.29	0.24	0.37
MVO	NO	-0.87	3.45	2.58	0.35	0.30	0.24	0.37	-0.41	0.59	0.37
MVO	YES	-2.10	3.10	1.00	0.30	0.33	0.21	0.37	-0.46	0.50	0.37

Table 4.1. Second-order statistics for static trading strategies in a two-stage example with total risk budget $\sigma_0 = 0.37$.

portfolio return correlations. This naturally raises the question as to whether negative correlation alone explains the increase in expected return for a given level of standard deviation realized by the MVO solution over the Beta solution. Furthermore, one might wonder whether the MVO solution corresponds to the portfolio weight vector that yields the largest negative correlation between the first and second stage returns over all possible vector directions. These questions are explored next.

Example 4.4.

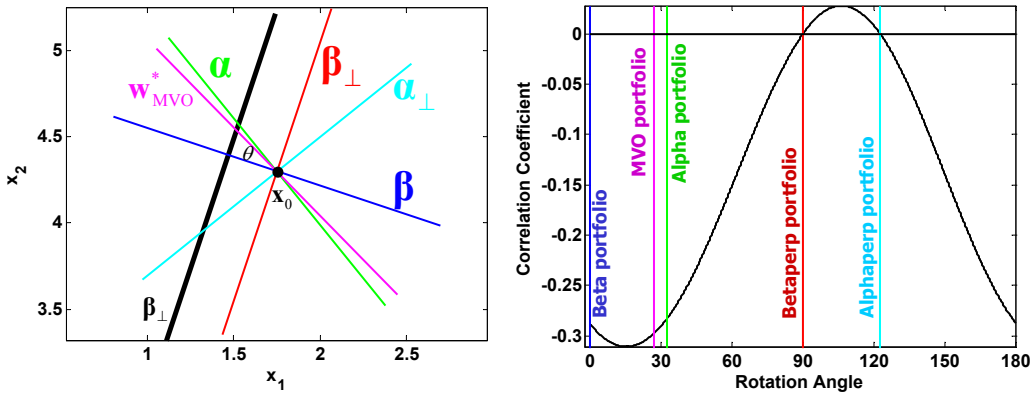
Consider a first-order cointegrated VAR process with the following parameters:

$$\mathbf{\Pi}_1 = \begin{pmatrix} 0.7878 & 0.0707 \\ 0.2634 & 0.9122 \end{pmatrix}, \qquad \mathbf{\Psi} = \mathbb{I}_2, \qquad \boldsymbol{\phi} = \mathbf{0}.$$

Figure 4-3(a) depicts the direction of the vectors $\{\boldsymbol{\alpha}, \boldsymbol{\alpha}_\perp, \boldsymbol{\beta}, \boldsymbol{\beta}_\perp\}$ and the two-stage MVO portfolio weight vector $\mathbf{w}_2^*$ centered at an arbitrary initial condition of $\mathbf{x}_0 = \begin{pmatrix} 1.75 & 4.3 \end{pmatrix}^T$. In order to explore the relationship between the direction of the portfolio policy and the degree of negative correlation achieved between the inter-stage portfolio returns, a set of portfolios are determined by rotating the portfolio weight vector over a range of angles $\theta = [0, \pi]$ from the Beta direction. Figure 4-3(b) displays the correlation coefficient between the returns for the first and second stages as a function of the rotation angle. Also highlighted is the correlation level achieved by the MVO solution and by four vectors corresponding to the $\{\boldsymbol{\alpha}, \boldsymbol{\alpha}_\perp, \boldsymbol{\beta}, \boldsymbol{\beta}_\perp\}$ subspaces. While the MVO solution is able to realize a larger degree of negative correlation than the Beta solution, it is not the direction with maximum negative correlation. This in turn suggests that negative correlation is only part of the story, and must be considered alongside individual asset and total portfolio leverage. ■

In addition to demonstrating that the MVO solution does not maximize the negative correlation between the inter-stage returns, the preceding example also revealed that the correlation coefficient between the inter-stage portfolio returns is not negative for all portfolio vector directions. In order to understand this result, first represent an arbitrary portfolio weight vector $\mathbf{w}$ using the non-orthogonal basis $\{\boldsymbol{\alpha}_\perp, \boldsymbol{\beta}\}$, as follows:

$$\mathbf{w} = c_1 \boldsymbol{\beta} + c_2 \boldsymbol{\alpha}_\perp, \tag{4.11}$$



(a) Geometry of system in Ex. 4.4. The vectors $\{\boldsymbol{\alpha}, \boldsymbol{\alpha}_\perp, \boldsymbol{\beta}, \boldsymbol{\beta}_\perp\}$ and the MVO optimal portfolio weight vector are depicted centered at an arbitrary initial condition of $\mathbf{x}_0 = (1.75 \quad 4.3)^T$.

(b) Correlation coefficient between first and second stage returns as a function of the portfolio vector direction. While the MVO solution realizes a larger degree of negative correlation than the Beta solution, it does not correspond to the portfolio direction which maximizes the negative correlation.

Figure 4-3. Correlation coefficient between first and second stage returns as a function of the portfolio vector direction.

where the constants $c_1, c_2 \in \mathbb{R}$ are computed according to the oblique projections:

$$c_1 = \left(\boldsymbol{\alpha}^T\boldsymbol{\beta}\right)^{-1}\boldsymbol{\alpha}^T\mathbf{w}, \quad c_2 = \left(\boldsymbol{\beta}_\perp^T\boldsymbol{\alpha}_\perp\right)^{-1}\boldsymbol{\beta}_\perp^T\mathbf{w}.$$

Next, substitute Eq. 4.11 into Eq. 4.10, the covariance between the first and second stage returns, which yields:

$$\operatorname{cov}\left[r_1, r_2\right] = c_1^2\boldsymbol{\beta}^T\boldsymbol{\Psi}\boldsymbol{\beta}\boldsymbol{\alpha}^T\boldsymbol{\beta} + c_2^2\boldsymbol{\alpha}_\perp^T\boldsymbol{\Psi}\boldsymbol{\beta}\boldsymbol{\alpha}^T\boldsymbol{\alpha}_\perp + c_1c_2\boldsymbol{\beta}^T\boldsymbol{\Psi}\boldsymbol{\beta}\boldsymbol{\alpha}^T\boldsymbol{\alpha}_\perp + c_1c_2\boldsymbol{\alpha}_\perp^T\boldsymbol{\Psi}\boldsymbol{\beta}\boldsymbol{\alpha}^T\boldsymbol{\beta}.$$

Due to the fact that $\boldsymbol{\alpha}^T\boldsymbol{\alpha}_\perp = 0$, the second and third terms are equivalently zero. The first term, which corresponds to a portfolio purely in the $\boldsymbol{\beta}$ direction, is always negative, as established in Thm. 4.1 below.

Theorem 4.1.

Consider a two-stage asset allocation problem in which the log-price process $\mathbf{x} \in \mathbb{R}^p$ *is assumed to evolve according to a first-order cointegrated VAR system. The matrix* $\boldsymbol{\Pi}_1$ *contains exactly* $p-1$ *unit eigenvalues, so that the matrix* $\boldsymbol{\Pi}$ *can be factored as the outer product of two* $p-$*dimensional vectors, as* $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^T$. *When the static portfolio weight*

vector is chosen according to $\mathbf{w} \propto \boldsymbol{\beta}$*, the covariance (equivalently the correlation coefficient) between the portfolio returns for the first and second stages is negative.*

The proof of Thm. 4.1 is given in Appendix 4.A. Thus, the covariance between the first and second stage returns is positive when the following condition is met:

$$c_1 \boldsymbol{\beta}^T \boldsymbol{\Psi} \boldsymbol{\beta} < -c_2 \boldsymbol{\alpha}_\perp^T \boldsymbol{\Psi} \boldsymbol{\beta} \quad \rightarrow \quad \mathbf{w}^T \boldsymbol{\Psi} \boldsymbol{\beta} < 0,$$

which follows using the fact that $\boldsymbol{\alpha}^T \boldsymbol{\beta} < 0$, as established in the proof of Thm. 4.1.

As a final point of interest, consider the two portfolio weight vectors corresponding to zero inter-stage return covariance. These can be directly determined through examination of Eq. 4.10, which can be factored as follows:

$$\operatorname{cov}[r_1, r_2] = \mathbf{w}^T (\boldsymbol{\Psi} \boldsymbol{\beta}) \boldsymbol{\alpha}^T \mathbf{w}.$$

Hence, choosing $\mathbf{w} \propto (\boldsymbol{\Psi} \boldsymbol{\beta})_\perp$ or $\mathbf{w} \propto \boldsymbol{\alpha}_\perp$ produces a correlation coefficient of zero. In Ex. 4.4, the covariance matrix of the input was chosen as $\boldsymbol{\Psi} = \mathbb{I}_2$, and therefore the two zero crossings of the correlation coefficient function occur at $\mathbf{w} = \{\boldsymbol{\beta}_\perp, \boldsymbol{\alpha}_\perp\}$. The zero crossing at $\mathbf{w} = \boldsymbol{\alpha}_\perp$ can be best understood by examining the total system response for a cointegrated VAR process defined in Eq. 3.16. The portfolio value (i.e., log price) at any given point is given by the scalar process $v[n] = \mathbf{w}^T \mathbf{x}[n]$. Therefore, when $\mathbf{w} = \boldsymbol{\alpha}_\perp$, the second term of $\mathbf{x}[n]$ corresponding to the AWSS random process drops out, leaving the portfolio value to be defined by a nonstationary integrated process. This in turn implies that the log-return process is an i.i.d. random process, with per-stage returns that are uncorrelated.

4.3.2 Short-term Predictability

The increased expected return earned by the MVO portfolio over the Beta portfolio for a given level of risk is due to the ability of the MVO solution to capture not only expected movements in the direction back towards the attractor ($\boldsymbol{\beta}_\perp$), but also short-term movements along the error-correcting space, $\boldsymbol{\alpha}$. While the eigenvector associated with the unit root of a cointegrated VAR system defines the long-term equilibrium relationship among the assets, the remaining eigenvalues and eigenvectors provide a source of predictability exploited by

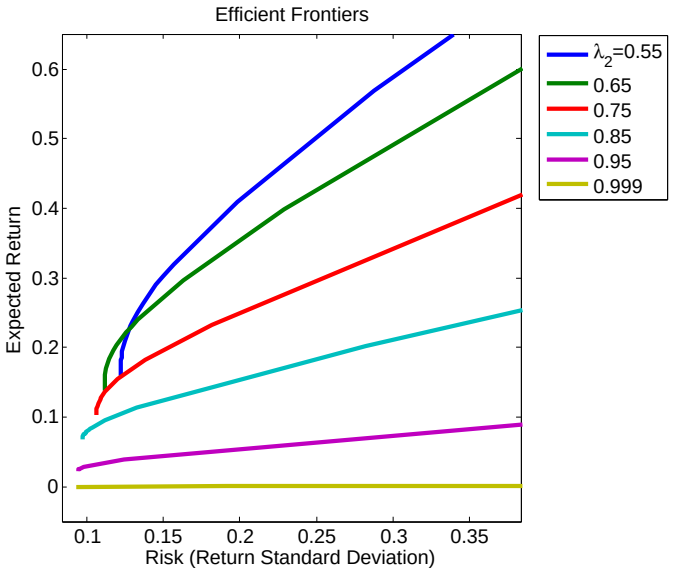


Figure 4-4. Capturing short-term predictability in a cointegrated VAR system. The largest expected returns are realized for the smallest values of λ_2. In the limit as $\lambda_2 \to 1$, the time series approach two independent random walks, and all predictability disappears.

the MVO trading strategy, as explored next.

Example 4.5.

For the cointegrated VAR system considered in Ex. 4.2, the eigendecomposition of the state transition matrix is given by:

$$\mathbf{A} = \mathbf{\Pi}_1 = \begin{pmatrix} 0.7878 & 0.0707 \\ 0.2634 & 0.9122 \end{pmatrix} = \begin{pmatrix} -0.6274 & -0.3162 \\ 0.7787 & -0.9487 \end{pmatrix} \begin{pmatrix} 0.7 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -0.6274 & -0.3162 \\ 0.7787 & -0.9487 \end{pmatrix}^{-1}.$$

Figure 4-4 contains a family of efficient frontiers (all with the budget constraint included) produced as the non-unit eigenvalue, λ_2, is varied over the set $\{0.55, 0.65, 0.75, 0.85, 0.95, 0.999\}$. In the limit as $\lambda_2 \to 1$, the matrix $\mathbf{\Pi}_1$ approaches the identity matrix, and the two underlying time series become uncorrelated random walks. In that limit, all the predictability disappears, evidenced by the fact that the efficient frontier flattens out to near zero expected return for all levels of risk. ■

4.3.3 Asymptotic Analysis

Having computed closed-form expressions for $\boldsymbol{\mu}_{\mathbf{r}_N}$ and $\mathbf{\Sigma}_{\mathbf{r}_N}$ in Section 4.2, both for systems with and without log-price drift and a budget constraint, the asymptotic properties of the

resulting portfolio weight vectors are now investigated. Of particular interest is whether the portfolio vectors converge to a steady-state solution as N increases, and how this value, if it exists, is related to the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ subspaces of the cointegrated VAR system. It is shown that the MVO portfolio weight vector satisfies $\mathbf{w}_N^* \in \text{sp}\{\boldsymbol{\beta}\}$ only in the limit of an infinite trading horizon when drift is omitted, both with and without a budget constraint. To begin, the behavior of the mean of the log-return, $\boldsymbol{\mu}_{\mathbf{r}_N}$, is characterized in Theorem 4.2.

Theorem 4.2.

Consider a cointegrated VAR(1) system, where $\mathbf{A} = \boldsymbol{\Pi}_1$ *has exactly one unit eigenvalue (i.e.,* $\lambda_1 = 1$*), while the remaining* $p-1$ *eigenvalues all lie strictly inside the unit circle (i.e.,* $|\lambda_i| < 1$*). As the length of the trading horizon increases, the mean of the asset returns behaves according to:*

$$\boldsymbol{\mu}_{\mathbf{r}} = \lim_{N\to\infty} \boldsymbol{\mu}_{\mathbf{r}_N} = \begin{cases} -\sum_{i=2}^{n} \mathbf{m}_i \mathbf{n}_i^T \mathbf{x}_0 & \textit{if } \boldsymbol{\phi} = \mathbf{0} \textit{ (no drift)}, \\ \infty & \textit{otherwise (drift).} \end{cases}$$

The proof of Thm. 4.2 is detailed in Appendix 4.A. As the theorem states, the vector of expected assets returns over an infinite trading horizon converges to a steady-state solution only when a constant drift term is not included in the cointegrated VAR log-price model. One may be tempted to hypothesize that in the presence of drift, a steady-state solution for the portfolio weight vector does not exist, since the portfolio weight vector explicitly depends on the expected return, as given by Eq. 4.2. However, this is *not* the case. The asymptotic properties of $\boldsymbol{\Sigma}_{\mathbf{r}_N}$ must be examined first, and the relationship between $\boldsymbol{\mu}_{\mathbf{r}}$ and $\boldsymbol{\Sigma}_{\mathbf{r}}$ must be jointly considered.

In order for $\boldsymbol{\Sigma}_{\mathbf{r}_N}$ to admit a steady-state solution, $\boldsymbol{\Sigma}_{\mathbf{r}}$, it must satisfy the discrete-time Lyapunov equation, given by:

$$\boldsymbol{\Sigma}_{\mathbf{r}} - \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{r}}\mathbf{A}^T - \boldsymbol{\Psi} = 0.$$

However, due to fact that $\mathbf{A}$ has an eigenvalue at unity, the matrix difference equation given by Eq. 4.7 is asymptotically unstable, and $\boldsymbol{\Sigma}_{\mathbf{r}_N}$ has one eigenvalue that diverges as N increases. Fortunately, the optimal portfolio weights, given in Eq. 4.2, do not directly depend

on $\mathbf{\Sigma}_{\mathbf{r}_N}$, but rather on $\mathbf{\Sigma}_{\mathbf{r}_N}^{-1}$, which is well-behaved. These observations are summarized in Theorem 4.3, which describes the behavior of the eigenvectors and eigenvalues of the return covariance and inverse covariance matrices as a function of the trading horizon, N.

Theorem 4.3.

Consider a cointegrated VAR(1) system, where $\mathbf{A} = \mathbf{\Pi}_1$ *has exactly one unit eigenvalue (i.e.,* $\lambda_1 = 1$*), while the remaining* $p-1$ *eigenvalues lie strictly inside the unit circle (i.e.,* $|\lambda_i| < 1$*). When* $N = 1$*, the* p *eigenvectors of* $\mathbf{\Sigma}_{\mathbf{r}_N}$ *and* $\mathbf{\Sigma}_{\mathbf{r}_N}^{-1}$ *are aligned with the eigenvectors of* $\mathbf{\Psi}$*. These eigenvectors converge to vectors within the column spans of* $\{\boldsymbol{\beta}, \boldsymbol{\beta}_\perp\}$ *as* N *approaches infinity, where* $\boldsymbol{\beta} \in \mathbb{R}^{p \times p-1}$ *and* $\boldsymbol{\beta}_\perp \in \mathbb{R}^{p \times 1}$*. The* $p-1$ *eigenvalues associated with the eigenvectors in span*$\{\boldsymbol{\beta}\}$ *converge to strictly positive, real-valued scalars, while the single eigenvalue associated with the eigenvector proportional to* $\boldsymbol{\beta}_\perp$ *diverges in* $\mathbf{\Sigma}_{\mathbf{r}_N}$ *and converges to zero in* $\mathbf{\Sigma}_{\mathbf{r}_N}^{-1}$*.*

The proof of Thm. 4.3 is given in Appendix 4.A. The main ideas from this theorem are illustrated in Ex. 4.6, below.

Example 4.6.

Consider a system comprised of two assets ($p = 2$) whose log-prices evolve according to Eq. 4.1, with $\mathbf{\Psi} = \mathbb{I}$. The matrix $\mathbf{\Pi}_1$ contains one unit eigenvalue ($\lambda_1 = 1$) and one eigenvalue strictly inside the unit circle ($|\lambda_2| < 1$). Recall from Chapter 3 that the eigenvector associated with λ_1 forms a basis for the attractor space (span$\{\boldsymbol{\beta}_\perp\}$), while the eigenvector associated with λ_2 forms of basis for the error-correcting space (span$\{\boldsymbol{\alpha}\}$). As depicted in Figure 4-5, the principal axes of $\mathbf{\Sigma}_{\mathbf{r}_N}$ are initially aligned with the unit vectors in the plane, and converge to $\{\boldsymbol{\beta}, \boldsymbol{\beta}_\perp\}$ as N increases. The eigenvalue associated with $\boldsymbol{\beta}$ converges to $\gamma = \frac{1}{1-\lambda^2}$, while the eigenvalue associated with $\boldsymbol{\beta}_\perp$ diverges. The eigenvectors of the inverse covariance matrix, $\mathbf{\Sigma}_{\mathbf{r}_N}^{-1}$, are the same eigenvectors as the eigenvectors of $\mathbf{\Sigma}_{\mathbf{r}_N}$, but the eigenvalues of $\mathbf{\Sigma}_{\mathbf{r}_N}^{-1}$ approach $1 - \lambda^2$ and zero. ■

It is now possible to combine the results of Theorems 4.2 and 4.3 in order to derive steady-state portfolio weight vectors for each of the four cases corresponding to the presence or absence of drift and a budget constraint, respectively, as given in Theorem 4.4 below.

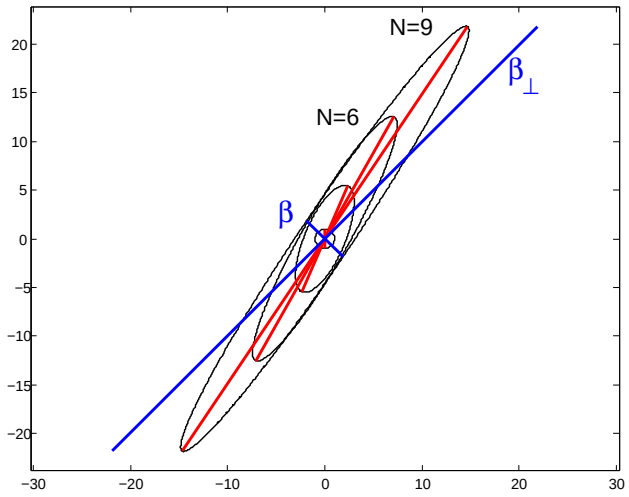


Figure 4-5. Evolution of Covariance Matrix Principal Axes. The lengths of the axes are proportional to the corresponding eigenvalues. In the system considered in Ex. 4.6, the axes are initially aligned with the unit vectors in the plane, and converge to $\{\boldsymbol{\beta}, \boldsymbol{\beta}_\perp\}$ as N increases.

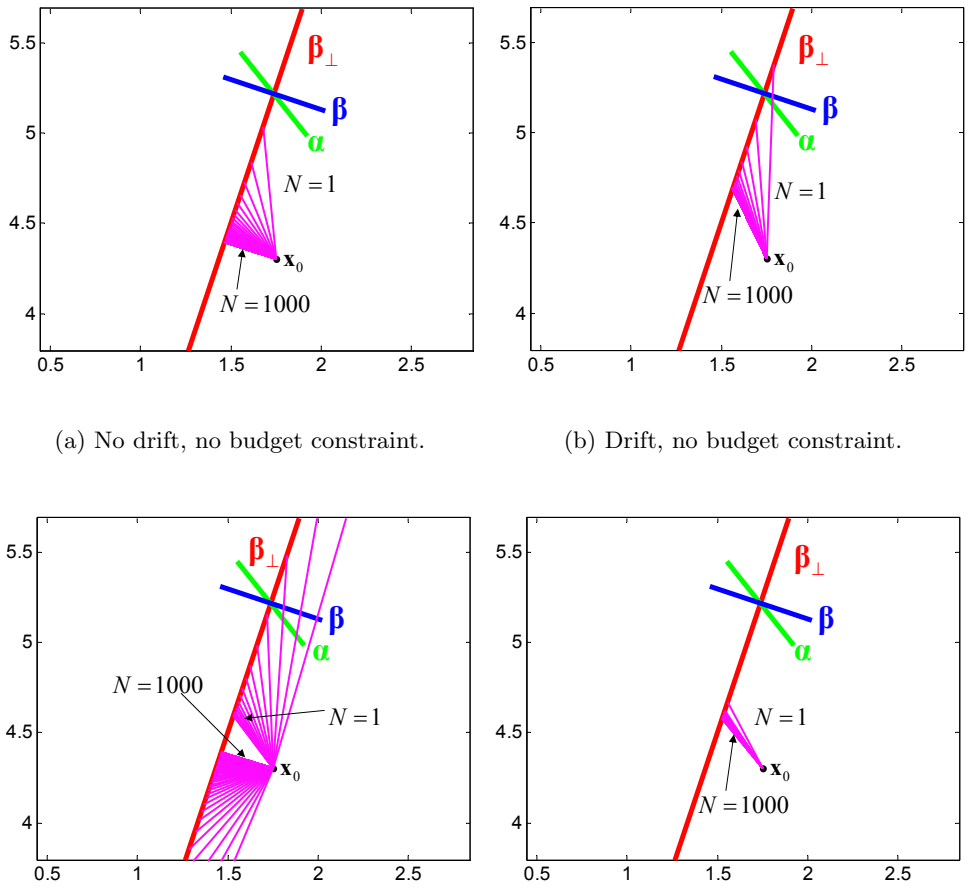
Theorem 4.4.

Consider a cointegrated VAR(1) system, where $\mathbf{A} = \boldsymbol{\Pi}_1$ *has exactly one unit eigenvalue (i.e.,* $\lambda_1 = 1$*), while the remaining* $p-1$ *eigenvalues lie strictly inside the unit circle (i.e.,* $|\lambda_i| < 1$*). The mean vector,* $\boldsymbol{\mu}_{\mathbf{r}_N}$*, and covariance matrix,* $\boldsymbol{\Sigma}_{\mathbf{r}_N}$*, of the expected return over a trading horizon of length* N *in a static setting are given by Eqs. 4.5 and 4.6, respectively. The optimal static mean-variance portfolio weight vector always converges to a steady-state solution. In the case of no drift (i.e.,* $\boldsymbol{\phi} = \mathbf{0}$*), the steady-state solution is in the span of* $\boldsymbol{\beta}$*; when drift is present, the steady-state solution contains components in both the* $\boldsymbol{\beta}$ *and* $\boldsymbol{\beta}_\perp$ *subspaces. These results are the same whether or not a budget constraint is enforced.*

The proof of Thm. 4.4 is given in Appendix 4.A. The results of this theorem are best illustrated through a simple example, presented next.

Example 4.7.

Consider a system of two assets, in which the log-prices are assumed to evolve according to a first-order cointegrated VAR system, as defined in Eq. 4.1, with $\boldsymbol{\Psi} = \begin{pmatrix} 0.2^2 & 0 \\ 0 & 0.07^2 \end{pmatrix}$.



(a) No drift, no budget constraint.

(b) Drift, no budget constraint.

(c) No drift, budget constraint.

(d) Drift, budget constraint.

Figure 4-6. Direction of static portfolio weight vectors as a function of investment length. The steady-state portfolio policy converges to a vector in the span of $\boldsymbol{\beta}$ only in the case where drift is not present.

Throughout this example $\boldsymbol{\Pi}_1 = \begin{pmatrix} 0.7878 & 0.0707 \\ 0.2634 & 0.9122 \end{pmatrix}$, which implies that:

$$\boldsymbol{\Pi} = \boldsymbol{\Pi}_1 - \mathbb{I} = \boldsymbol{\alpha}\boldsymbol{\beta}^T = \begin{pmatrix} -0.2122 & 0.0707 \\ 0.2634 & -0.0878 \end{pmatrix} = \begin{pmatrix} -0.3746 \\ 0.4649 \end{pmatrix} \begin{pmatrix} 0.5665 & -0.1888 \end{pmatrix}.$$

The initial value of the log-prices is assumed to be $\mathbf{x}_0 = \begin{pmatrix} 1.75 & 4.3 \end{pmatrix}^T$ and the drift, when included, is given by $\boldsymbol{\phi} = \begin{pmatrix} .5 & .25 \end{pmatrix}^T$.

For each of the four cases considered in Theorem 4.4, the direction of the N-stage mean-

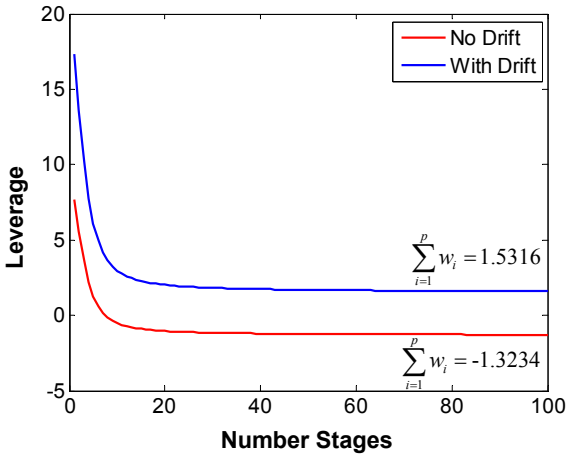


Figure 4-7. Degree of total leverage, $\mathbf{w}^T\mathbf{1}$, when a budget constraint is omitted, as a function of the trade horizon, N, for the system in Example 4.7.

variance optimal portfolio weight vector is computed as N is varied from 1 to 1000, and the results are depicted in Figure 4-6. As the plots confirm, the steady-state solution converges to a direction in the span of $\boldsymbol{\beta}$ only in the case where drift is not present. The trajectory of the optimal portfolio weight vector direction in the third case (no drift, with budget constraint), is particularly interesting. Here, the weight vector initially moves towards $\boldsymbol{\beta}$ in the counter-clockwise direction, but after five steps, $\mathbf{w}_N^*$ reverses its direction of motion, and approaches $\boldsymbol{\beta}$ from the clockwise direction. This abrupt change of direction occurs as the optimal portfolio weight vector must always point in the direction towards, not away from, the attractor space.

When the budget constraint is omitted, the degree of total leverage, $\mathbf{1}^T\mathbf{w}$, can be computed as function of the trade horizon, as shown in Figure 4-7. While the amount of leverage converges regardless of whether or not drift is included in the model, the presence of drift can significantly alter the asymptotic solution. In Ex. 4.7, the overall position is net long 153 percent of initial wealth when drift is included, while the net position is short 132 percent of initial wealth when drift is omitted. ■

4.3.4 Mean-Variance Tradeoff

In addition to the Beta portfolio, another solution to consider inspired by the geometry of a cointegrated VAR system is known as the Alpha portfolio, which occurs when the portfolio

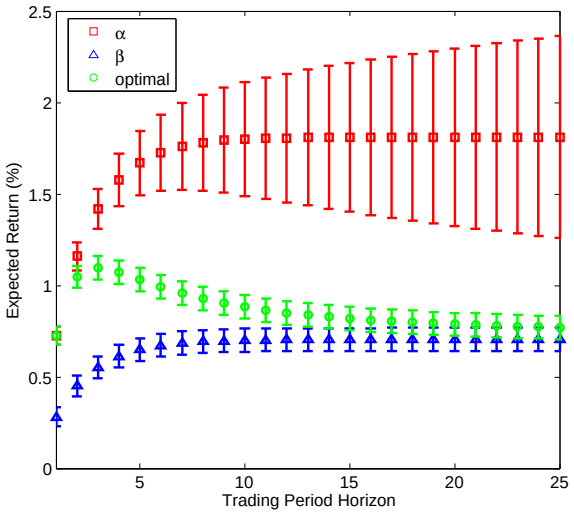


Figure 4-8. Expected return as a function of trading horizon, with portfolio return standard deviation error bars. For the system considered in Ex. 4.8, the mean-variance optimal portfolio is initially aligned with $\boldsymbol{\alpha}$, but as N increases and the variance grows, it is pulled toward $\boldsymbol{\beta}$ in order to satisfy the variance constraint, until they are perfectly aligned in the limit of an infinite trading horizon.

is bought from within the error-correcting space (i.e., span$\{\boldsymbol{\alpha}\}$). Such portfolios may be optimal when the variance constraint of Problem P_0 is replaced by a leverage constraint (i.e., a constraint on the length of the portfolio vector), such as $\mathbf{w}^T\mathbf{w} = 1$. As Thm. 4.5 shows, in this case the optimal action is to choose a portfolio proportional to the direction of the expected log-return (i.e., $\boldsymbol{\mu}_{\mathbf{r}_N}$), which in one particular case is equal to the $\boldsymbol{\alpha}$ direction.

Theorem 4.5.

Consider the following static portfolio choice problem, in which the portfolio is selected to be mean-leverage optimal, in accordance with problem P_0''', as follows:

$$\left.\begin{array}{l} \{\mathbf{w}_N^*\} = \underset{\mathbf{w}_N}{\arg\max}\ \ \mathbf{w}_N^T \boldsymbol{\mu}_{\mathbf{r}_N} \\ \qquad\qquad s.t.\ \ \mathbf{w}_N^T \mathbf{w}_N = 1 \end{array}\right\} \ P_0''',$$

Given a constraint on the degree of portfolio leverage, the optimal portfolio weight vector for a trading horizon of length N is proportional to the direction of expected change, as follows:

$$\mathbf{w}_N^* \propto \boldsymbol{\mu}_{\mathbf{r}_N} = N\mathbf{m}_1\mathbf{n}_1^T\boldsymbol{\phi} + \sum_{i=2}^{p}\left(\lambda_i^N - 1\right)\mathbf{m}_i\mathbf{n}_i^T\mathbf{x}_0 + \frac{1-\lambda_i^N}{1-\lambda_i}\mathbf{m}_i\mathbf{n}_i^T\boldsymbol{\phi}.$$

In the limit of an infinite trading horizon and when $\boldsymbol{\phi} = \mathbf{0}$*, the optimal weight vector is proportional to:*

$$\mathbf{w}_{\infty}^{*} \propto -\sum_{i=2}^{n} \left(\mathbf{n}_i^T \mathbf{x}_0\right) \mathbf{m}_i.$$

The proof of Thm. 4.5 is given in Appendix 4.A. It is important to note that in a first order VAR system with $p = 2$, $\boldsymbol{\phi} = \mathbf{0}$, and the constraint $\mathbf{w}^T\mathbf{w} = 1$, the direction of the optimal portfolio is proportional to $\mathbf{m}_2 \propto \boldsymbol{\alpha}$. The following example utilizes the idea of a leverage constraint in order to further explore the mean-variance trade-off experienced in systems where the log-prices of the assets follow a cointegrated VAR model.

Example 4.8.

A set of three portfolios is compared for a system comprised of two assets whose log-prices are known to evolve according to a cointegrated VAR process. Furthermore, it is assumed that the covariance matrix of the input process, $\boldsymbol{\Psi}$, is proportional to the identity matrix. The first portfolio is constructed to be proportional to the $\boldsymbol{\alpha}$ vector, the second portfolio is chosen so that the weights are proportional to the $\boldsymbol{\beta}$ vector, and the third so that the weights are proportional to the mean-variance optimal (MVO) portfolio, $\mathbf{w}^*$, computed with no budget constraint. Once the three portfolio directions are determined, each vector is normalized so that $||\mathbf{w}||_2 = 1$. As depicted in Fig. 4-8, for a given N, the highest expected return is achieved with the Alpha portfolio, due to the fact that when $p = 2$, the direction of expected change, $\boldsymbol{\mu}_N$, is proportional to $\boldsymbol{\alpha}$ for all N. However, as shown by the corresponding error bars, the variance of this portfolio grows linearly over time due to the unit eigenvalue in the state transition matrix. On the other hand, the Beta portfolio has smaller expected return, but the variance converges to a finite quantity. In the middle is the MVO portfolio. Due to the assumption that $\boldsymbol{\Psi}$ is proportional to the identity matrix, when $N = 1$ the MVO portfolio is aligned with the Alpha portfolio. As N increases and the portfolio return variance grows, the MVO portfolio is pulled toward the $\boldsymbol{\beta}$ direction in order to satisfy the variance constraint. Only in the limit of an infinite trading horizon does the MVO portfolio align with the Beta portfolio. ■

As Example 4.8 demonstrates, the time horizon of the investor matters. For short horizons, there is additional profit earned by buying the portfolio with a component not only in the

direction of finite variance, i.e. the Beta direction, but also in the direction of expected change. However, for long horizons, the variance constraint effectively limits the investor to buy more and more in the Beta direction.

4.A Proofs of Chapter 4 Theorems

Proof of Theorem 4.1.

Consider the static two-stage, two-asset portfolio choice problem with portfolio weight vector $\mathbf{w}$. Since the correlation coefficient is related to the covariance through a positive scale factor (i.e., the product of the standard deviations corresponding to the portfolio returns for the first and second stages), it is sufficient to show that the covariance between the portfolio returns for the first and second stages is always negative when $\mathbf{w}$ is chosen in the Beta direction. Let r_1 denote the return for the first stage, and r_2 denote the return for the second stage, so that the general form of the covariance is computed as:

$$\begin{aligned}
\operatorname{cov}[r_1, r_2] &= E[r_1 r_2] - E[r_1] E[r_2] \\
&= E\left[\mathbf{w}^T (\mathbf{x}_1 - \mathbf{x}_0)(\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{w}\right] - E\left[\mathbf{w}^T (\mathbf{x}_1 - \mathbf{x}_0)\right] E\left[\mathbf{w}^T (\mathbf{x}_2 - \mathbf{x}_1)\right] \\
&= \mathbf{w}^T E\left[(\boldsymbol{\Pi}\mathbf{x}_0 + \boldsymbol{\epsilon}_1)(\boldsymbol{\Pi}(\boldsymbol{\Pi}_1\mathbf{x}_0 + \boldsymbol{\epsilon}_1) + \boldsymbol{\epsilon}_2)^T\right] \mathbf{w} - \mathbf{w}^T \boldsymbol{\Pi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Pi}_1^T\boldsymbol{\Pi}^T\mathbf{w} \\
&= \mathbf{w}^T\left(\boldsymbol{\Pi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Pi}_1^T\boldsymbol{\Pi}^T + \boldsymbol{\Psi}\boldsymbol{\Pi}^T\right)\mathbf{w} - \mathbf{w}^T\boldsymbol{\Pi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Pi}_1^T\boldsymbol{\Pi}^T\mathbf{w} \\
&= \mathbf{w}^T\boldsymbol{\Psi}\boldsymbol{\Pi}^T\mathbf{w}.
\end{aligned}$$

Next consider the case where $\mathbf{w} = c\boldsymbol{\beta}$, where $c = \pm 1$, so that the covariance is equal to $\operatorname{cov}[r_1, r_2] = c^2\boldsymbol{\beta}^T\boldsymbol{\Psi}\boldsymbol{\beta}\boldsymbol{\alpha}^T\boldsymbol{\beta}$. Since $c^2 > 0$ and $\boldsymbol{\beta}^T\boldsymbol{\Psi}\boldsymbol{\beta} > 0$ due to the fact that $\boldsymbol{\Psi}$ is a positive definite covariance matrix, the desired result follows only when $\boldsymbol{\alpha}^T\boldsymbol{\beta} < 0$. As shown next, the latter condition follows as a direct result of the assumption concerning the non-unit eigenvalue of the matrix $\boldsymbol{\Pi}_1$, λ. Recall from Thm. 3.1 that when the matrix $\boldsymbol{\Pi}$ has rank $r = 1$, the first $p-1$ eigenvectors of $\boldsymbol{\Pi}$, $\mathbf{M}_{\boldsymbol{\beta}_\perp} = \{\mathbf{m}_1, ..., \mathbf{m}_{p-1}\}$, form a basis for the $\boldsymbol{\beta}_\perp$ subspace with corresponding eigenvalues of zero, while the remaining eigenvector, $\mathbf{m}_p$, forms a basis for the $\boldsymbol{\alpha}$ subspace, with eigenvalue $-2 < \lambda - 1 < 0$. Now let $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^T$, and let $\mathbf{m}_p = k\boldsymbol{\alpha}$ for some $k \in \mathbb{R}$. The eigendecomposition of $\boldsymbol{\Pi}$ is given by:

$$\begin{aligned}
\boldsymbol{\Pi} &= \begin{pmatrix} \mathbf{M}_{\boldsymbol{\beta}_\perp} & k\boldsymbol{\alpha} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \\ & \lambda - 1 \end{pmatrix} \begin{pmatrix} \mathbf{M}_{\boldsymbol{\beta}_\perp} & k\boldsymbol{\alpha} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \mathbf{M}_{\boldsymbol{\beta}_\perp} & k\boldsymbol{\alpha} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \\ & \lambda - 1 \end{pmatrix} \begin{pmatrix} \left(\boldsymbol{\alpha}_\perp^T \mathbf{M}_{\boldsymbol{\beta}_\perp}\right)^{-1} \boldsymbol{\alpha}_\perp^T \\ k\left(\boldsymbol{\beta}^T\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}^T \end{pmatrix} \\
&= k^2(\lambda - 1)\left(\boldsymbol{\beta}^T\boldsymbol{\alpha}\right)\boldsymbol{\alpha}\boldsymbol{\beta}^T.
\end{aligned}$$

Equating the two factorizations for $\mathbf{\Pi}$ reveals that $k^2(\lambda-1)\left(\boldsymbol{\beta}^T\boldsymbol{\alpha}\right)=1$. Since $k^2>0$ and $-2<\lambda-1<0$, it follows that $\boldsymbol{\beta}^T\boldsymbol{\alpha}=\boldsymbol{\alpha}^T\boldsymbol{\beta}<0$.

Proof of Theorem 4.2.

For fixed N, the mean of the total return is given by Eq. 4.4. Thus the limit as N increases is given by:

$$\lim_{N\to\infty}\ \boldsymbol{\mu}_{\mathbf{r}_N}=\lim_{N\to\infty}\ N\mathbf{m}_1\mathbf{n}_1^T\boldsymbol{\phi}+\sum_{i=2}^{p}\left(\lambda_i^N-1\right)\mathbf{m}_i\mathbf{n}_i^T\mathbf{x}_0+\frac{1-\lambda_i^N}{1-\lambda_i}\mathbf{m}_i\mathbf{n}_i^T\boldsymbol{\phi},$$

Since $\mathbf{A}$ has exactly one unit eigenvalue, i.e. $\lambda_1=1$, while the remaining $p-1$ eigenvalues all lie strictly inside the unit circle, i.e. $|\lambda_i|<1,\ i=2,...,p$, the term $\lambda_i^N\to 0$ as N increases. Therefore, when drift is not present, i.e. $\boldsymbol{\phi}=\mathbf{0}$, the first and third terms drop out, and the mean of the return vector converges to a weighted combination of the eigenvectors of $\mathbf{A}$ corresponding to the non-unit eigenvalues, $-\sum_{i=2}^{n}\mathbf{m}_i\mathbf{n}_i^T\mathbf{x}_0$. In the presence of drift, the first term dominates the summation, thus the overall expected return $\boldsymbol{\mu}_{\mathbf{r}_N}$ diverges. □

Proof of Theorem 4.3.

According to Eq. 4.6, when $N=1$, $\mathbf{\Sigma}_{\mathbf{r}_1}=\mathbf{\Psi}$. Next consider the case where $N>1$ and let $\mathbf{\Psi}=\mathbf{S}\mathbf{S}^T$ and $\mathbf{A}=\mathbf{M}\mathbf{D}\mathbf{M}^{-1}$, so that the N period covariance matrix for $\mathbf{r}_N$, $\mathbf{\Sigma}_{\mathbf{r}_N}$, can be expressed as follows:

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{r}_N} &= \sum_{i=0}^{N-1}\mathbf{M}\mathbf{D}^i\mathbf{M}^{-1}\mathbf{S}\mathbf{S}^T\left(\mathbf{M}^{-1}\right)^T\mathbf{D}^i\mathbf{M}^T\\
&= \mathbf{M}\left[\sum_{i=0}^{N-1}\left(\mathbf{D}^i\mathbf{M}^{-1}\mathbf{S}\right)\left(\mathbf{D}^i\mathbf{M}^{-1}\mathbf{S}\right)^T\right]\mathbf{M}^T\\
&= \mathbf{M}\sum_{i=0}^{N-1}\begin{pmatrix} c_{1,1}^i\mathbf{q}_1^T\mathbf{q}_1 & \cdots & c_{1,n}^i\mathbf{q}_1^T\mathbf{q}_n\\ c_{2,1}^i\mathbf{q}_2^T\mathbf{q}_1 & \cdots & c_{2,n}^i\mathbf{q}_2^T\mathbf{q}_n\\ \vdots & & \vdots\\ c_{n,1}^i\mathbf{q}_n^T\mathbf{q}_1 & \cdots & c_{n,n}^i\mathbf{q}_n^T\mathbf{q}_n\end{pmatrix}\mathbf{M}^T,
\end{aligned}$$

where $c_{i,j}=\lambda_i\lambda_j$ and $\mathbf{q}_i$ is the i^{th} column of the matrix $\mathbf{Q}=\mathbf{M}^{-1}\mathbf{S}$. Using the fact that

$\lambda_1 = 1$, define $\mathbf{K}$ as follows:

$$\mathbf{K} = \begin{pmatrix} N\mathbf{q}_1^T\mathbf{q}_1 & \cdots & \frac{1-c_{1,n}^N}{1-c_{1,n}}\mathbf{q}_1^T\mathbf{q}_n \\ \frac{1-c_{2,1}^N}{1-c_{2,1}}\mathbf{q}_2^T\mathbf{q}_1 & \cdots & \frac{1-c_{2,n}^N}{1-c_{2,n}}\mathbf{q}_2^T\mathbf{q}_n \\ \vdots & & \vdots \\ \frac{1-c_{n,1}^N}{1-c_{n,1}}\mathbf{q}_n^T\mathbf{q}_1 & \cdots & \frac{1-c_{n,n}^N}{1-c_{n,n}}\mathbf{q}_n^T\mathbf{q}_n \end{pmatrix},$$

so that the covariance matrix can be expressed as follows:

$$\boldsymbol{\Sigma}_{\mathbf{r}_N} = \mathbf{M}\mathbf{K}\mathbf{M}^T = \sum_{i=1}^{n}\sum_{j=1}^{n} K_{i,j}\mathbf{m}_i\mathbf{m}_j^T.$$

As N approaches infinity, the first term, $N\mathbf{q}_1^T\mathbf{q}_1\mathbf{m}_1\mathbf{m}_1^T$, dominates the summation, causing the covariance matrix to diverge in the direction of $\mathbf{m}_1 = \boldsymbol{\beta}_\perp$. Hence $\boldsymbol{\Sigma}_\mathbf{r}$ has one eigenvector in the direction of $\boldsymbol{\beta}_\perp$ with corresponding eigenvalue of infinity. Using the fact that the eigenvectors of any real, symmetric matrix are orthogonal, the remaining eigenvectors must form a basis for $\boldsymbol{\beta}$, each with bounded eigenvalue γ_i. The inverse covariance matrix $\boldsymbol{\Sigma}_\mathbf{r}^{-1}$ has the same eigenvectors as $\boldsymbol{\Sigma}_\mathbf{r}$, with eigenvalues of zero and $\frac{1}{\gamma_i}$, respectively.

Proof of Theorem 4.4.

First, consider the case where $\boldsymbol{\phi} = \mathbf{0}$ and no budget constraint is included. The optimal steady-state portfolio weight vector is computed using Eq. 4.2 as follows:

$$\mathbf{w}_\infty^* = \lim_{N\to\infty} \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}_{\mathbf{r}_N}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}}}\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}$$

As described by Theorems 4.2 and 4.3, the term $\boldsymbol{\mu}_{\mathbf{r}_N}$ converges as N increases, while the inverse covariance matrix converges to $\boldsymbol{\Sigma}_\mathbf{r}^{-1} = \begin{pmatrix}\boldsymbol{\beta} & \boldsymbol{\beta}_\perp\end{pmatrix}$. Since the eigenvalues associated with the $\boldsymbol{\beta}_\perp$ eigenvectors asymptotically approach zero, the matrix product $\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}$ converges to a vector in the span of $\boldsymbol{\beta}$. In addition, the scale factor $\frac{\sigma_0}{\sqrt{\boldsymbol{\mu}_{\mathbf{r}_N}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}}}$ also converges to a fixed point, as neither $\boldsymbol{\mu}_{\mathbf{r}_N}$ nor $\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}$ diverges as N increases.

Second, consider the case where drift is present, i.e. $\boldsymbol{\phi} \neq \mathbf{0}$, but the budget constraint is not enforced. Here both $\boldsymbol{\mu}_{\mathbf{r}_N}$ and $\boldsymbol{\Sigma}_{\mathbf{r}_N}$ contain terms that depend on N, and therefore the two limits cannot be examined independently. To this end, recall the definition of $\boldsymbol{\mu}_{\mathbf{r}_N}$

from Eq. 4.4 and let $\boldsymbol{\mu}_{\mathbf{r}_N} = N\boldsymbol{\beta}_\perp + b$, where:

$$b = \sum_{i=2}^{p} \left(\lambda_i^N - 1\right) \mathbf{m}_i \mathbf{n}_i^T \mathbf{x}_0 + \frac{1-\lambda_i^N}{1-\lambda_i} \mathbf{m}_i \mathbf{n}_i^T \boldsymbol{\phi},$$

and $\mathbf{n}_1^T \boldsymbol{\phi} \mathbf{m}_1 \in \text{sp}\{\boldsymbol{\beta}_\perp\}$. Now the product $\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1} \boldsymbol{\mu}_{\mathbf{r}_N}$ can be divided into two pieces, $N\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\beta}_\perp$ and $\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1} b$. The presence of the additional factor of N in the first term has significant impact. While the matrix $N\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}$ still has eigenvectors that approach $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$ as N increases, the eigenvalue associated with $\boldsymbol{\beta}_\perp$ no longer diverges, as was the case in Theorem 4.3. Thus, $N\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\beta}_\perp$ is proportional to $\boldsymbol{\beta}_\perp$, and $\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1} b$ is proportional to $\boldsymbol{\beta}$. Thus the overall asymptotic direction of the optimal portfolio weight vector is a linear combination of these two vectors, and is not purely in the direction of $\boldsymbol{\beta}$, as was the case when drift was not present. Also note that in the case where drift is present, the denominator of the scale factor contains an extra factor of $\sqrt{N}$, so the degree of leverage utilized by this strategy decreases as N increases.

Third, consider the case where $\boldsymbol{\phi} = \mathbf{0}$, but a budget constraint of the form $\mathbf{w}^T\mathbf{1} = 1$ is now included. The optimal steady-state portfolio weight vector is computed using Eq. 4.12 as follows:

$$\mathbf{w}_\infty^* = \lim_{N\to\infty} \frac{1}{2\lambda_L^*} \boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1} \left(\boldsymbol{\mu}_{\mathbf{r}_N} - \gamma_L^* \mathbf{1}\right).$$

Since the quantities $\left\{\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}, \boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}, \mathbf{1}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1}, \boldsymbol{\mu}_{\mathbf{r}_N}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1}\right\}$ all converge to steady-state solutions, λ_L^* and γ_L^* also admit steady-state solutions. Again, the overall portfolio weight vector converges to a vector in the span of $\boldsymbol{\beta}$ as the null-space of the steady-state inverse covariance matrix is equal to $\boldsymbol{\beta}_\perp$.

Finally, consider the case where both drift and a budget constraint are included. As with the second case, the quantity $\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\boldsymbol{\mu}_{\mathbf{r}_N}$ no longer converges to a vector in the span of $\boldsymbol{\beta}$, but instead converges to a vector with components in both the $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$ subspaces. Hence, the optimal steady-state portfolio weight vector is a linear combination of three vectors, with the first two in the span of the $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$ subspaces, and the third one in the direction of $\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1}$.

Proof of Theorem 4.5.

According to the Cauchy-Schwartz inequality, the inner product $\mathbf{w}^T\boldsymbol{\mu}_{\mathbf{r}_N}$ subject to $\mathbf{w}^T\mathbf{w} = 1$ is maximized when the vector $\mathbf{w}$ is chosen to be proportional to $\boldsymbol{\mu}_{\mathbf{r}_N}$. Since $|\lambda_i| < 1$ for $i = 2 \ldots n$ and $\boldsymbol{\Phi} = \mathbf{0}^T$, the limit follows.

4.B Alternative derivation of MVO solution with budget constraint

In the Markowitz mean-variance framework, the budget constraint may be enforced by introducing a second Lagrange multiplier, $\gamma_L \in \mathbb{R}$, to Problem P_0, as follows:

$$\{\mathbf{w}_N^*, \lambda_L^*, \gamma_L^*\} = \underset{\mathbf{w}_N, \lambda_L, \gamma_L}{\arg\max} \; \mathbf{w}_N^T\boldsymbol{\mu}_{\mathbf{r}_N} - \lambda_L\left(\mathbf{w}_N^T\boldsymbol{\Sigma}_{\mathbf{r}_N}\mathbf{w}_N - \sigma_0^2\right) - \gamma_L\left(\mathbf{w}_N^T\mathbf{1} - 1\right) \quad \Big\} \; P_0'',$$

with solution:

$$\mathbf{w}_N^* = \frac{1}{2\lambda_L^*}\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\left(\boldsymbol{\mu}_{\mathbf{r}_N} - \gamma_L^*\mathbf{1}\right) = \frac{1}{2\lambda_L^*}\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\left(\boldsymbol{\mu}_{\mathbf{r}_N} - \left(\frac{\boldsymbol{\mu}_{\mathbf{r}_N}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1} - 2\lambda_L^*}{\mathbf{1}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1}}\right)\mathbf{1}\right). \tag{4.12}$$

In order to determine the value of λ_L^*, again let $\mathbf{w}_N = \frac{1}{2\lambda_L}\mathbf{f} + \mathbf{g}$, where the vectors $\mathbf{f}$ and $\mathbf{g}$ are now defined as follows:

$$\begin{aligned}
\mathbf{f} &= \boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\left(\boldsymbol{\mu}_{\mathbf{r}_N} - \left(\frac{\boldsymbol{\mu}_{\mathbf{r}_N}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1}}\right)\mathbf{1}\right), \\
\mathbf{g} &= -\left(\frac{1}{\mathbf{1}^T\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1}}\right)\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}\mathbf{1}.
\end{aligned}$$

The relationship between λ_L and σ_0^2 is again given by a quadratic function. Due to the presence of term $\boldsymbol{\Sigma}_{\mathbf{r}_N}$, the coefficients in the original expression, given by Eq. 4.9, diverge as N increases. On the other hand, here the coefficients depend on $\boldsymbol{\Sigma}_{\mathbf{r}_N}^{-1}$, which converges as N increases, as shown in Section 4.3.3.

Chapter 5

Dynamic Portfolio Choice

The static MVO portfolios considered in Chapter 4 do not take advantage of new information that becomes available over the lifetime of the trade, such as updated asset prices or realized gains and losses. The investor makes a single decision at the beginning of the investment period, and holds the resultant portfolio for the entire time horizon. On the other hand, dynamic portfolio strategies, such as those considered in this chapter and the next, enable the investor to rebalance the portfolio at intermediate stages, taking into account new information as it becomes available.

Dynamic multistage portfolio construction has been discussed in the literature since the 1960s, as summarized in Section 2.4. When the portfolio choice problem is formulated within the framework of *expected utility maximization*, the optimal sequence of portfolio policies can be readily determined using a dynamic programming approach. On the other hand, it was not known until forty years later how to construct dynamically optimal portfolios when the investor's terminal wealth is mean-variance efficient. In particular, Li and Ng [36] show that the mean-variance preference function can be mapped into the expected utility framework by transforming the desired problem into an auxiliary one that uses a quadratic objective function, that can subsequently be solved using existing methods. They prove that if the investor acts to maximize the expected value of a quadratic function of terminal wealth, the resultant portfolios sequence is also mean-variance efficient.

The method of [36], referred to as the *Separable Embedding* technique, suffers from one fundamental limitation. While their framework allows stochastic modeling of the per-stage asset returns, the parameters of the return distributions must be *known in advance*. This restriction poses a problem when modeling the asset log-prices using a cointegrated vector

autoregressive model, because in this case the per-stage return distributions have state-dependent mean vectors. Thus, the return statistics are not known at time t_0, as required by [36]. Here, a new algorithm is presented that extends the separable embedding technique to the case of stochastic asset return distributions by incorporating the use of Monte Carlo and importance sampling methods to solve the resulting dynamic programming problem[1]. While the new method is presented within the context of cointegrated log-price dynamics, the methodology is general and can be applied to any setting in which the asset return distributions are changing over time in a stochastic manner.

The organization of this chapter is as follows. In Section 5.1.1, the dynamic mean-variance portfolio choice problem is presented, and the Principle of Separable Embedding is applied in order to derive an auxiliary quadratic utility problem. In Section 5.1.2, the solution to auxiliary problem is derived using the dynamic programming algorithm. A Monte Carlo based algorithm for solving a set of recursive equations required for the computation of the optimal portfolio policies is detailed in Section 5.1.3. The mapping from the auxiliary problem back to the original mean-variance problem is given in Section 5.1.4. Section 5.1.5 represents the auxiliary quadratic utility problem using the linear quadratic regulator framework. The inclusion of a budget constraint, risk-free asset, and no short-sale constraint are discussed in Sections 5.2.1, 5.2.2, and 5.2.3, respectively. Section 5.3.1 analyzes the per-stage portfolio return statistics, and finally, the asymptotic properties of the dynamic MVO solution are explored in Section 5.3.2.

5.1 Mean-Variance Optimal Portfolio Construction

In this section, the optimal sequence of mean-variance efficient portfolio policies is computed for the case where inter-stage rebalancing is allowed. Unlike the static solution of Chapter 4, in which a single portfolio weight vector is held without modification across all N stages of the investment horizon, here the portfolio weights at the beginning of each stage are allowed vary in a dynamic fashion according to some function of the observed state, which includes both updated asset log-prices and realized portfolio gains (losses). A detailed

[1] One attempt to generalize the work of [36] for the case of asset returns with stochastic parameters in *continuous-time* is given in Chapter 6 of [66]. However, the model used to describe the evolution of the asset prices over time does not account for *inter-temporal* correlations among the assets, as captured by a system of cointegrated assets.

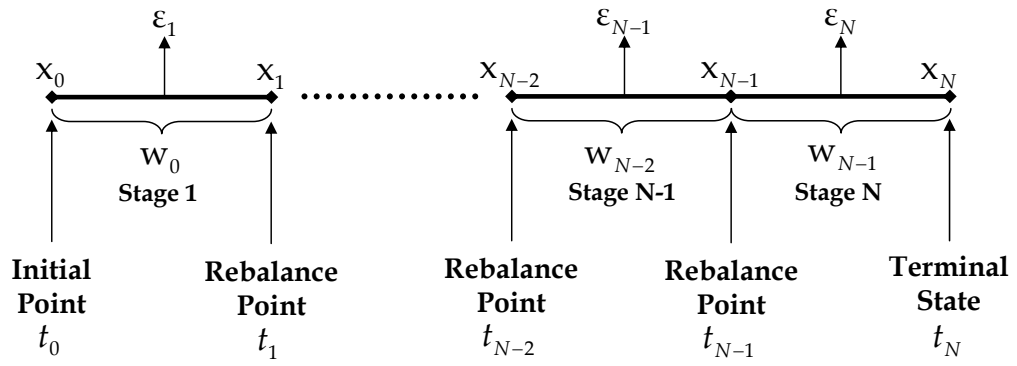


Figure 5-1. Multistage dynamic portfolio choice problem.

problem formulation is presented next.

5.1.1 Problem Formulation

Let $\mathbf{x}_k$ be a p-dimensional random vector representing the log-prices of a group of assets that are assumed to evolve according to a first-order cointegrated vector autoregressive process, as follows:

$$\mathbf{x}[n+1] = \mathbf{\Pi}_1 \mathbf{x}[n] + \mathbf{\Phi d}[n] + \boldsymbol{\epsilon}[n], \tag{5.1}$$

with stochastic input $\boldsymbol{\epsilon}[n] \sim N(\mathbf{0}, \mathbf{\Psi})$ and constant deterministic drift $\mathbf{\Phi d}[n] = \boldsymbol{\phi}$. The drift is included in order to capture the overall linear growth trend present in the historical log-prices of most assets. A detailed description of such models is given in Chapter 3. Even though attention is focused on the first-order VAR system, extensions to higher-order systems are easily realized by augmenting the state space.

The multistage Markowitz mean-variance portfolio choice framework presented in Section 2.4.1 is utilized here. The investor acts so that the *terminal return* is mean-variance efficient. Under such preferences, the objective is to maximize the expected value of the portfolio terminal return, subject to an equality constraint on its variance. Under the assumption that the asset log-prices evolve according to a cointegrated VAR process, the terminal return is computed as the sum of the per-stage log-returns (i.e., change in the log-prices). Formally,

the optimization problem, P_0, can be stated as follows:

$$\left.\begin{array}{l} \left\{\mathbf{w}_0^*, \mathbf{w}_1^*, \ldots, \mathbf{w}_{N-1}^*\right\} = \underset{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{N-1}}{\arg\max} \; E_{t_0}\left[\sum_{k=0}^{N-1} \mathbf{w}_k^T \mathbf{r}_{k+1}\right] \\ \qquad \text{s.t. } \operatorname{var}_{t_0}\left[\sum_{k=0}^{N-1} \mathbf{w}_k^T \mathbf{r}_{k+1}\right] = \sigma_0^2, \end{array}\right\} \; P_0$$

where the per-period vector of individual asset returns, $\mathbf{r}_k$, is defined by:

$$\mathbf{r}_k = \Delta \mathbf{x}_k = \mathbf{x}_k - \mathbf{x}_{k-1}. \tag{5.2}$$

All expectation and variance operators are computed with respect to the information available at the start of the investment horizon, time t_0. The inner product represented by $\mathbf{w}_k^T \mathbf{r}_{k+1}$ denotes the return of the portfolio at stage k. The portfolio weight vector $\mathbf{w}_k$ represents the relative percentage of wealth to allocate to each asset with positive and negative weights indicating long and short positions, respectively. The portfolio is allowed to be leveraged (i.e., the market value of the portfolio may exceed the available wealth), and inclusion of a budget constraint of the form $\mathbf{1}^T \mathbf{w}_k = 1$ is considered in Section 5.2.1. Without a budget constraint, the degree of leverage is limited only by the allowable risk parameter, σ_0^2. Recall from Section 2.2.3, that omission of a budget constraint is inherently equivalent to including a risk-free asset in the model with a rate of $r_f = 0$.

As discussed in Section 2.4.1, it is not possible to solve Problem P_0 exactly using dynamic programming techniques, as the objective function is not additive in the number of stages due to the (potentially) non-zero covariance terms. However, as Li and Ng [36] show, it is possible to first solve a related, auxiliary problem in order to learn the optimal sequence of MVO portfolios. The resulting technique, known as the *(Principle of Separable Embedding)*, is stated in Theorem 5.1 below.

Theorem 5.1.

Consider the problem of maximizing the expected value of a quadratic function of the cumulative N stage portfolio return, as follows:

$$\left\{\mathbf{w}_0^*, \mathbf{w}_1^*, \ldots, \mathbf{w}_{N-1}^*\right\} = \underset{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{N-1}}{\arg\max} \; E_{t_0}\left[\gamma_N \sum_{k=0}^{N-1} \mathbf{w}_k^T \mathbf{r}_{k+1} - \lambda_N \left(\sum_{k=0}^{N-1} \mathbf{w}_k^T \mathbf{r}_{k+1}\right)^2\right] \; \Bigg\} \; P_1,$$

where the scale factors γ_N and λ_N define the shape of the utility function, and are deterministic quantities set by the investor in accordance with his risk preferences. The optimal sequence of portfolio policies, $\left\{\mathbf{w}_0^*, \ldots, \mathbf{w}_{N-1}^*\right\}$ for Problem P_0 is also optimal for Problem P_1, for some appropriate choice of $\{\gamma_N, \lambda_N\}$.

Hence, the original mean-variance portfolio choice problem is *embedded* into an auxiliary problem that is *separable* in the sense required by dynamic programming. The proof of Thm. 5.1 is given in Appendix 5.A.

In applying Thm. 5.1 in order to compute the portfolio weight vector functions, $\mathbf{w}_k^*(\mathbf{x}_k)$, Li and Ng assume that the parameters of the return distributions are *known exactly* at time t_0. However, under the assumption that the asset log-prices evolve according to a cointegrated vector autoregressive model, the per-stage return distributions have state-dependent mean vectors, and are therefore not known at the start of the investment horizon. To address this limitation, a novel extension to their algorithm is developed here, that integrates the use of Monte Carlo and importance sampling methods within the separable embedding framework in order to solve the dynamic mean-variance problem for cointegrated log-price dynamics.

5.1.2 Optimal Portfolio Policies

In this section, the solution to Problem P_1 is derived using the dynamic programming algorithm. Here, care is taken to explicitly highlight how the assumption that the asset log-prices evolve according to a cointegrated VAR process prevents the solution in [36] from being applied directly, thus necessitating the use of the Monte Carlo methods derived in Section 5.1.3.

To begin the derivation, let the value function at the last stage, denoted as $J_N^*(r_N)$, be defined by:

$$J_N^*(r_N) = U(r_N) = \gamma_N r_N - \lambda_N r_N^2.$$

According to the Bellman principle of optimality, at time t_{N-1}, the investor acts to maximize the sum of his current reward and the expected reward-to-go:

$$\mathbf{w}_{N-1}^* = \underset{\mathbf{w}_{N-1}}{\arg\max}\, E_{t_{N-1}}\left[0 + J_N(r_N)\right].$$

The conditional expectation is computed with respect to the information that is available at time t_{N-1}, implying that the only source of randomness is $\mathbf{x}_N$, or equivalently, $\boldsymbol{\epsilon}_N$. Here, the total return can be decomposed into two terms, as $r_N = r_{N-1} + \mathbf{w}_{N-1}^T (\mathbf{x}_N - \mathbf{x}_{N-1})$, where the first term is a known, deterministic quantity with respect to the information available at time t_{N-1}, while the second term is random. Thus, the value function at time t_{N-1} is given by:

$$\begin{aligned}
J_{N-1}^* (r_{N-1}) &= \max_{\mathbf{w}_{N-1}} E_{t_{N-1}} [J_N (r_N)] = \max_{\mathbf{w}_{N-1}} E_{t_{N-1}} \left[\gamma_N r_N - \lambda_N r_N^2\right] \\
&= \max_{\mathbf{w}_{N-1}} E_{t_{N-1}} \Big[\gamma_N \left(r_{N-1} + \mathbf{w}_{N-1}^T (\mathbf{x}_N - \mathbf{x}_{N-1})\right) \\
&\quad -\lambda_N \left(r_{N-1} + \mathbf{w}_{N-1}^T (\mathbf{x}_N - \mathbf{x}_{N-1})\right)^2\Big] \\
&= \max_{\mathbf{w}_{N-1}} \gamma_N r_{N-1} + \gamma_N \mathbf{w}_{N-1}^T E [\mathbf{x}_N - \mathbf{x}_{N-1}] - \lambda_N r_{N-1}^2 \\
&\quad -2\lambda_N r_{N-1} \mathbf{w}_{N-1}^T E [\mathbf{x}_N - \mathbf{x}_{N-1}] \\
&\quad -\lambda_N \mathbf{w}_{N-1}^T E \left[(\mathbf{x}_N - \mathbf{x}_{N-1}) (\mathbf{x}_N - \mathbf{x}_{N-1})^T\right] \mathbf{w}_{N-1} \\
&= \max_{\mathbf{w}_{N-1}} \gamma_N r_{N-1} + \gamma_N \mathbf{w}_{N-1}^T \mathbf{m}_{N-1} - \lambda_N r_{N-1}^2 - 2\lambda_N r_{N-1} \mathbf{w}_{N-1}^T \mathbf{m}_{N-1} \\
&\quad -\lambda_N \mathbf{w}_{N-1}^T \mathbf{S}_{N-1} \mathbf{w}_{N-1}, \qquad (5.3)
\end{aligned}$$

where the vector $\mathbf{m}_N$ and the matrix $\mathbf{S}_N$ are defined as follows:

$$\begin{aligned}
\mathbf{m}_{N-1} &= E_{t_{N-1}} [\mathbf{x}_N - \mathbf{x}_{N-1}] \\
\mathbf{S}_{N-1} &= E_{t_{N-1}} \left[(\mathbf{x}_N - \mathbf{x}_{N-1}) (\mathbf{x}_N - \mathbf{x}_{N-1})^T\right].
\end{aligned}$$

When it is assumed that the log-prices evolve according to the cointegrated VAR model of Eq. 5.1, the above two moments admit closed-form solutions, given by:

$$\mathbf{m}_{N-1} = \boldsymbol{\Pi} \mathbf{x}_{N-1} + \boldsymbol{\phi} \qquad (5.4)$$

$$\mathbf{S}_{N-1} = \boldsymbol{\Psi} + \mathbf{m}_{N-1} \mathbf{m}_{N-1}^T. \qquad (5.5)$$

Note that since $\mathbf{m}_{N-1}$ and $\mathbf{S}_{N-1}$ are functions of the state, $\mathbf{x}_{N-1}$, they are also random variables. At times these quantities will be denoted as $\mathbf{m}_{N-1} (\mathbf{x}_{N-1})$ and $\mathbf{S}_{N-1} (\mathbf{x}_{N-1})$ in order to emphasize their inherent state dependence.

The optimal portfolio policy for the last stage, $\mathbf{w}_{N-1}^*$, can be found by setting the derivative

of Eq. 5.3 equal to zero, and solving for the portfolio weight vector, which yields:

$$\mathbf{w}^*_{N-1} = \frac{1}{2\lambda_N}\left(\gamma_N - 2\lambda_N r_{N-1}\right)\mathbf{S}^{-1}_{N-1}\left(\mathbf{x}_{N-1}\right)\mathbf{m}_{N-1}\left(\mathbf{x}_{N-1}\right). \tag{5.6}$$

Therefore, the optimal portfolio policy at time t_{N-1} is not only a function of the observed asset log-prices, $\mathbf{x}_{N-1}$, but also of the cumulative return to date, r_{N-1}.

The expression for $\mathbf{w}^*_{N-1}$ can now be plugged back into Eq. 5.3, and after some algebra, it is possible to show that the value function can again be expressed as a quadratic function of the cumulative, realized return up to time t_{N-1}, as follows:

$$J^*_{N-1}\left(r_{N-1}\right) = \gamma_{N-1} r_{N-1} - \lambda_{N-1} r^2_{N-1} + c_{N-1},$$

where:

$$\gamma_{N-1} = \gamma_N\left(1 - \mathbf{m}^T_{N-1}\left(\mathbf{x}_{N-1}\right)\mathbf{S}^{-1}_{N-1}\left(\mathbf{x}_{N-1}\right)\mathbf{m}_{N-1}\left(\mathbf{x}_{N-1}\right)\right), \tag{5.7}$$

$$\lambda_{N-1} = \lambda_N\left(1 - \mathbf{m}^T_{N-1}\left(\mathbf{x}_{N-1}\right)\mathbf{S}^{-1}_{N-1}\left(\mathbf{x}_{N-1}\right)\mathbf{m}_{N-1}\left(\mathbf{x}_{N-1}\right)\right), \tag{5.8}$$

$$c_{N-1} = 0.25\frac{\gamma_N}{\lambda_N}\mathbf{m}^T_{N-1}\left(\mathbf{x}_{N-1}\right)\mathbf{S}^{-1}_{N-1}\left(\mathbf{x}_{N-1}\right)\mathbf{m}_{N-1}\left(\mathbf{x}_{N-1}\right). \tag{5.9}$$

It is important to observe that the coefficients in the quadratic objective function, γ_{N-1} and λ_{N-1}, defined by Eqs. 5.7 and 5.8, respectively, are themselves random variables, as the terms $\mathbf{m}_{N-1}$ and $\mathbf{S}_{N-1}$ depend on the random variable $\mathbf{x}_{N-1}$. It is in this crucial manner that the case of cointegrated log-prices considered in this thesis differs from the assumptions of [36].

Applying the Bellman recursion once more, the value function at time t_{N-2}, is given by:

$$\begin{aligned}
J^*_{N-2}(r_{N-2}) &= \max_{\mathbf{w}_{N-2}} E_{t_{N-2}}\left[J_{N-1}(r_{N-1})\right] \\
&= \max_{\mathbf{w}_{N-2}} E_{t_{N-2}}\left[\gamma_{N-1} r_{N-1} - \lambda_{N-1} r^2_{N-1} + c_{N-1}\right] \\
&= \max_{\mathbf{w}_{N-2}} E_{t_{N-2}}\Big[\gamma_{N-1}\left(r_{N-2} + \mathbf{w}^T_{N-2}\left(\mathbf{x}_{N-1} - \mathbf{x}_{N-2}\right)\right) \\
&\qquad -\lambda_{N-1}\left(r_{N-2} + \mathbf{w}^T_{N-2}\left(\mathbf{x}_{N-1} - \mathbf{x}_{N-2}\right)\right)^2 + c_{N-1}\Big] \\
&= \max_{\mathbf{w}_{N-2}} \gamma_N r_{N-2}\nu_{N-2} + \gamma_N \mathbf{w}^T_{N-2}\mathbf{m}_{N-2} - \lambda_N r^2_{N-2}\nu_{N-2} \\
&\qquad -2\lambda_N r_{N-2}\mathbf{w}^T_{N-2}\mathbf{m}_{N-2} - \lambda_N \mathbf{w}^T_{N-2}\mathbf{S}_{N-2}\mathbf{w}_{N-2} + E[c_{N-1}], \quad (5.10)
\end{aligned}$$

where ν_{N-2}, $\mathbf{m}_{N-2}$ and $\mathbf{S}_{N-2}$ are defined as follows:

$$\begin{aligned}
z_{N-2} &= 1 - \mathbf{m}^T_{N-1}(\mathbf{x}_{N-1})\,\mathbf{S}^{-1}_{N-1}(\mathbf{x}_{N-1})\,\mathbf{m}_{N-1}(\mathbf{x}_{N-1}) && (5.11)\\
\nu_{N-2} &= E_{t_{N-2}}\left[z_{N-2}\right] && (5.12)\\
\mathbf{m}_{N-2} &= E_{t_{N-2}}\left[z_{N-2}\left(\mathbf{x}_{N-1} - \mathbf{x}_{N-2}\right)\right] && (5.13)\\
\mathbf{S}_{N-2} &= E_{t_{N-2}}\left[z_{N-2}\left(\mathbf{x}_{N-1} - \mathbf{x}_{N-2}\right)\left(\mathbf{x}_{N-1} - \mathbf{x}_{N-2}\right)^T\right]. && (5.14)
\end{aligned}$$

With respect to the information available at time t_{N-2}, only $\mathbf{x}_{N-2}$ is known, and therefore the expectations are computed over the random variable $\mathbf{x}_{N-1}$. Only the constant terms λ_N and γ_N can be pulled out from the expectation, whereas the terms z_{N-2} and $(\mathbf{x}_{N-1} - \mathbf{x}_{N-2})$ depend on $\mathbf{x}_{N-1}$ and are therefore random. This is in direct contrast to the derivation found in the Appendix of [35], where the entire term corresponding to λ_{N-1} is pulled out of the expectation.

Comparing Equations 5.3 and 5.10 reveals that by setting $\nu_{N-1} = 1$, the two expressions have identical structure. Thus, Eqs. 5.12 through 5.14 can be rewritten as follows:

$$\begin{aligned}
\nu_{N-2} &= E_{t_{N-2}}\left[\left(\nu_{N-1} - \mathbf{m}^T_{N-1}\mathbf{S}^{-1}_{N-1}\mathbf{m}_{N-1}\right)\right], \\
\mathbf{m}_{N-2} &= E_{t_{N-2}}\left[\left(\nu_{N-1} - \mathbf{m}^T_{N-1}\mathbf{S}^{-1}_{N-1}\mathbf{m}_{N-1}\right)\left(\mathbf{x}_{N-1} - \mathbf{x}_{N-2}\right)\right], \\
\mathbf{S}_{N-2} &= E_{t_{N-2}}\left[\left(\nu_{N-1} - \mathbf{m}^T_{N-1}\mathbf{S}^{-1}_{N-1}\mathbf{m}_{N-1}\right)\left(\mathbf{x}_{N-1} - \mathbf{x}_{N-2}\right)\left(\mathbf{x}_{N-1} - \mathbf{x}_{N-2}\right)^T\right].
\end{aligned}$$

The optimal portfolio policy for the second to last stage, $\mathbf{w}_{N-2}$, can be found by differentiating Eq. 5.10 and setting the resulting expression equal to zero, yielding:

$$\mathbf{w}^*_{N-2} = \frac{1}{2\lambda_N}\left(\gamma_N - 2\lambda_N r_{N-2}\right)\mathbf{S}^{-1}_{N-2}\left(\mathbf{x}_{N-2}\right)\mathbf{m}_{N-2}\left(\mathbf{x}_{N-2}\right). \tag{5.15}$$

The function for the portfolio weight vector two stages from the end has identical structure to that portfolio weight vector for the last stage, given in Equation 5.6. This suggests a generalization of the results derived in this section. The optimal portfolio weight vector corresponding to time t_k, when the investor acts so as to maximize the expected utility of terminal wealth according to a quadratic utility function, is given by:

$$\mathbf{w}^*_k\left(\mathbf{x}_k, r_k\right) = \frac{1}{2\lambda_N}\left(\gamma_N - 2\lambda_N r_k\right)\mathbf{S}^{-1}_k\left(\mathbf{x}_k\right)\mathbf{m}_k\left(\mathbf{x}_k\right), \tag{5.16}$$

where the moment functions $\nu_k\left(\mathbf{x}_k\right)$, $\mathbf{m}_k\left(\mathbf{x}_k\right)$ and $\mathbf{S}_k\left(\mathbf{x}_k\right)$ evolve recursively according to:

$$\nu_k\left(\mathbf{x}_k\right) = E_{t_k}\left[\nu_{k+1} - \mathbf{m}^T_{k+1}\mathbf{S}^{-1}_{k+1}\mathbf{m}_{k+1}\right], \tag{5.17}$$

$$\mathbf{m}_k\left(\mathbf{x}_k\right) = E_{t_k}\left[\left(\nu_{k+1} - \mathbf{m}^T_{k+1}\mathbf{S}^{-1}_{k+1}\mathbf{m}_{k+1}\right)\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right], \tag{5.18}$$

$$\mathbf{S}_k\left(\mathbf{x}_k\right) = E_{t_k}\left[\left(\nu_{k+1} - \mathbf{m}^T_{k+1}\mathbf{S}^{-1}_{k+1}\mathbf{m}_{k+1}\right)\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)^T\right]. \tag{5.19}$$

Thus, an exact solution to Problem P_1 has been found in terms of the functions $\{\mathbf{m}_k\left(\mathbf{x}_k\right), \mathbf{S}_k\left(\mathbf{x}_k\right)\}$, the deterministic constants $\{\gamma_N, \lambda_N\}$, and the cumulative portfolio return, r_k. In order to map Problem P_1 back into the original mean-variance portfolio choice framework given by Problem P_0, one must determine the relationship between the choice of $\{\gamma_N, \lambda_N\}$ and the total variance constraint, σ^2_0, as discussed in Section 5.1.4. However, it is first necessary to derive a technique for solving Equations 5.17 through 5.19, which do not admit closed-form solutions. An algorithm based on Monte Carlo and importance sampling methods is presented next, in Section 5.1.3.

5.1.3 A Monte Carlo Based Algorithm

While the functions $\mathbf{m}_{N-1}$ and $\mathbf{S}_{N-1}$ can easily be computed in closed-form, as given by Eqs. 5.4 and 5.5, the remaining terms ν_k, $\mathbf{m}_k$ and $\mathbf{S}_k$ do not admit analytic solutions. For example, at time t_{N-2}, all of the required expectations involve the term $\left(1 - \mathbf{m}^T_N\mathbf{S}^{-1}_N\mathbf{m}_N\right)$,

which is equal to (with drift $\boldsymbol{\phi} = \mathbf{0}$):

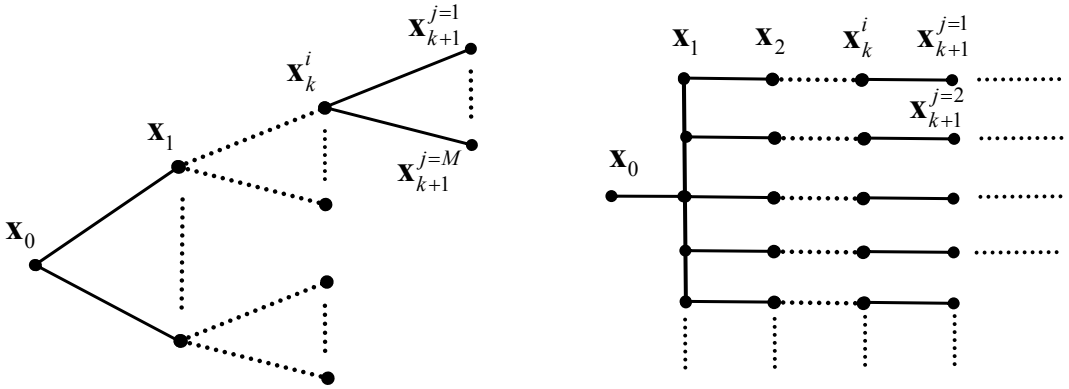
$$
\begin{aligned}
&1 - \mathbf{m}_{N-1}^T \mathbf{S}_{N-1}^{-1} \mathbf{m}_{N-1} \\
&\quad = 1 - \mathbf{x}_{N-1}^T \boldsymbol{\Pi}^T \left(\boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1} \boldsymbol{\Pi} \mathbf{x}_{N-1} \left(1 + \mathbf{x}_{N-1}^T \boldsymbol{\Pi}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Pi} \mathbf{x}_{N-1} \right)^{-1} \mathbf{x}_1^T \boldsymbol{\Pi}^T \boldsymbol{\Psi}^{-1} \right) \boldsymbol{\Pi} x_{N-1} \\
&\quad = 1 - \mathbf{x}_{N-1}^T \boldsymbol{\Pi}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Pi} \mathbf{x}_{N-1} + \frac{\left(\mathbf{x}_{N-1}^T \boldsymbol{\Pi}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Pi} \mathbf{x}_{N-1} \right)^2}{1 + \mathbf{x}_{N-1}^T \boldsymbol{\Pi}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Pi} \mathbf{x}_{N-1}} \\
&\quad = \frac{1}{1 + \mathbf{x}_{N-1}^T \boldsymbol{\Pi}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Pi} \mathbf{x}_{N-1}},
\end{aligned}
$$

where the first equality follows by applying the matrix inversion lemma to $\mathbf{S}_{N-1}$. The presence of a quadratic function of $\mathbf{x}_{N-1}$ in the denominator makes ν_{N-2}, $\mathbf{m}_{N-2}$, $\mathbf{S}_{N-2}$, and all subsequent moments, difficult to compute analytically[2].

In lieu of closed-form solutions, the sequence of optimal portfolio policies must, therefore, be computed numerically. Specifically, the functions $\nu_k(\mathbf{x}_k)$, $\mathbf{m}_k(\mathbf{x}_k)$ and $\mathbf{S}_k(\mathbf{x}_k)$ can be approximated through a two stage Monte-Carlo scheme consisting of a training phase, in which these functions are evaluated over a set of sample paths of the log-price process, and a testing phase, in which the resulting functional approximations are applied to real data. This approach bears resemblance to the Approximate Dynamic Programming (ADP) framework [48], in which the value function is directly approximated over a set of sample paths of the underlying state *and* control spaces.

In order to implement a Monte Carlo based approximation scheme, sample paths of the asset log-price process are required. One approach is to simulate the log-price process according to an M-ary tree, as shown in Figure 5-2(a), where each node of the tree contains the $p-$dimensional vector $\mathbf{x}_k$. On the one hand, this scheme suffers from the curse of dimensionality, since the size of the tree grows exponentially with the number of stages. However, a tree structure allows for easy computation of each expectation as a sample mean. For each node in the tree, a set of M samples from the conditional distribution $f\left(\mathbf{x}_{k+1} | \mathbf{x}_k^i\right)$ are available at the next level, indexed by j, and therefore Eqs. 5.17 through 5.19 may be

[2] In the case where $\boldsymbol{\Pi} = \mathbf{0}$, which corresponds to a system of p independent walks, the expressions for $\mathbf{m}_k$ and $\mathbf{S}_k$ do admit closed-form solutions and the Monte-Carlo algorithm described here is not needed.



(a) Sample paths of log-price process are generated according to a tree. The size of the tree grows exponentially with the number of stages.

(b) In order to avoid the curse of dimensionality, a fixed number of sample paths of the log-price process are generated.

Figure 5-2. Sample path simulation schemes.

approximated using a backwards recursion algorithm, as follows:

$$\begin{aligned}
\nu_k\left(\mathbf{x}_k^i\right) &= E_{t_k}\left[\nu_{k+1} - \mathbf{m}_{k+1}^T \mathbf{S}_{k+1}^{-1} \mathbf{m}_{k+1}\right] \\
&\simeq \frac{1}{M}\sum_{j=1}^{M}\left(\nu_{k+1}^j - \left(\mathbf{m}_{k+1}^j\right)^T \left(\mathbf{S}_{k+1}^j\right)^{-1} \mathbf{m}_{k+1}^j\right), \\
\mathbf{m}_k\left(\mathbf{x}_k^i\right) &= E_{t_k}\left[\left(\nu_{k+1} - \mathbf{m}_{k+1}^T \mathbf{S}_{k+1}^{-1} \mathbf{m}_{k+1}\right)\left(\mathbf{x}_{k+1} - \mathbf{x}_k^i\right)\right] \\
&\simeq \frac{1}{M}\sum_{j=1}^{M}\left(\nu_{k+1}^j - \left(\mathbf{m}_{k+1}^j\right)^T \left(\mathbf{S}_{k+1}^j\right)^{-1} \mathbf{m}_{k+1}^j\right)\left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^i\right), \\
\mathbf{S}_k\left(\mathbf{x}_k^i\right) &= E_{t_k}\left[\left(\nu_{k+1} - \mathbf{m}_{k+1}^T \mathbf{S}_{k+1}^{-1} \mathbf{m}_{k+1}\right)\left(\mathbf{x}_{k+1} - \mathbf{x}_k^i\right)\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)^T\right] \\
&\simeq \frac{1}{M}\sum_{j=1}^{M}\left(\nu_{k+1}^j - \left(\mathbf{m}_{k+1}^j\right)^T \left(\mathbf{S}_{k+1}^j\right)^{-1} \mathbf{m}_{k+1}^j\right)\left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^i\right)\left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^i\right)^T.
\end{aligned}$$

While the tree structure enables direct computation of the required moments at each level using the sample mean estimator, a total of M^N sample paths must be generated. Hence, this method quickly becomes computationally intractable[3].

A second approach, depicted in Figure 5-2(b), addresses the curse of dimensionality by generating a fixed number of sample paths of the log-price process. Here, samples of

[3]Note that a tree structure is also required for computation of the desired moments using a numerical integration approach, which also suffers from the curse of dimensionality

$f(\mathbf{x}_{k+1}|\,\mathbf{x}_0)$, not $f\left(\mathbf{x}_{k+1}|\,\mathbf{x}_k^i\right)$, are available, and therefore importance sampling methods (see Appendix B) must be utilized in order to approximate the functions $\{\nu_k, \mathbf{m}_k, \mathbf{S}_k\}$ using a sample mean estimator. Specifically, each component in the sample mean summation must be scaled by a weighting factor, w_{ij}, defined by:

$$w_{ij} = \frac{f\left(\mathbf{x}_{k+1}^j \middle|\, \mathbf{x}_k^i\right)}{f\left(\mathbf{x}_{k+1}^j \middle|\, \mathbf{x}_0\right)}.$$

Equations 5.17 through 5.19 may now be approximated using the modified backwards recursion algorithm:

$$\begin{aligned}
\nu_k\left(\mathbf{x}_k^i\right) &\simeq \frac{1}{M}\sum_{j=1}^{M} w_{ij}\left(\nu_{k+1}^j - \left(\mathbf{m}_{k+1}^j\right)^T\left(\mathbf{S}_{k+1}^j\right)^{-1}\mathbf{m}_{k+1}^j\right),\\
\mathbf{m}_k\left(\mathbf{x}_k^i\right) &\simeq \frac{1}{M}\sum_{j=1}^{M} w_{ij}\left(\nu_{k+1}^j - \left(\mathbf{m}_{k+1}^j\right)^T\left(\mathbf{S}_{k+1}^j\right)^{-1}\mathbf{m}_{k+1}^j\right)\left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^j\right),\\
\mathbf{S}_k\left(\mathbf{x}_k^i\right) &\simeq \frac{1}{M}\sum_{j=1}^{M} w_{ij}\left(\nu_{k+1}^j - \left(\mathbf{m}_{k+1}^j\right)^T\left(\mathbf{S}_{k+1}^j\right)^{-1}\mathbf{m}_{k+1}^j\right)\left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^j\right)\left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^j\right)^T,
\end{aligned}$$

where M now represents the total number of sample paths generated in the training set and j indices the samples from the conditional distribution $f(\mathbf{x}_{k+1}|\,\mathbf{x}_0)$. The full training phase procedure is summarized in Algorithm 1.

Once the training phase is complete, the grid of simulated log-prices and corresponding values of ν_k, $\mathbf{m}_k$ and $\mathbf{S}_k$ can be used to determine the optimal portfolio weights for a new sequence of test log-prices. However, with the exception of the initial state, an exact log-price match will not be available in the training set. Thus, one must estimate the values of $\mathbf{m}_k$, and $\mathbf{S}_k$ by interpolating over the available samples. One possible scheme, based on the idea of a kernel-density estimate [53], is to weight the training values according to their "closeness" to the current test value. The scalar version of this interpolation scheme is illustrated in Fig. 5-3. A multivariate Gaussian kernel is selected, with a mean equal to the current test state, denoted by $\mathbf{x}_k^{\text{test}}$, and covariance matrix equal to the one period covariance of the underlying cointegrated VAR process. The corresponding weights are

Algorithm 1 Training Phase: Approximating $\nu_k(\mathbf{x}_k)$, $\mathbf{m}_k(\mathbf{x}_k)$ and $\mathbf{S}_k(\mathbf{x}_k)$.

STEP 1: Generate Sample Paths
for $k = 1$ to N **do**
  for $i = 1$ to M **do**
    $\mathbf{x}_k^i \leftarrow \mathbf{\Pi}_1 \mathbf{x}_{k-1}^i + \boldsymbol{\phi} + \boldsymbol{\epsilon}_k^i$
  end for
end for
STEP 2: Initialization
for $i = 1$ to M **do**
  $\nu_{N-1}^i\left(\mathbf{x}_{N-1}^i\right) \leftarrow 1$
  $\mathbf{m}_{N-1}^i\left(\mathbf{x}_{N-1}^i\right) \leftarrow \mathbf{\Pi}\mathbf{x}_{N-1}^i + \boldsymbol{\phi}$
  $\mathbf{S}_{N-1}^i\left(\mathbf{x}_{N-1}^i\right) \leftarrow \mathbf{\Psi} + \mathbf{m}_{N-1}^i\left(\mathbf{x}_{N-1}^i\right)\mathbf{m}_{N-1}^{iT}\left(\mathbf{x}_{N-1}^i\right)$
end for
STEP 3: Run backwards recursion
for $k = N-2$ to 0 **do**
  for $i = 1$ to M **do**
    for $k = 1$ to M **do**
      $w_{ij} \leftarrow \frac{f\left(\mathbf{x}_{k+1}^j|\mathbf{x}_k^i\right)}{f\left(\mathbf{x}_{k+1}^j|\mathbf{x}_0\right)}$
      $z_j \leftarrow \nu_{k+1}^k\left(\mathbf{x}_{k+1}^k\right) - \left(\mathbf{m}_{k+1}^j\left(\mathbf{x}_{k+1}^j\right)\right)^T \left(\mathbf{S}_{k+1}^j\left(\mathbf{x}_{k+1}^j\right)\right)^{-1} \mathbf{m}_{k+1}^j\left(\mathbf{x}_{k+1}^j\right)$
    end for
    $\nu_k\left(\mathbf{x}_k^i\right) \leftarrow \frac{1}{M}\sum_{j=1}^M w_{ij} z_j$
    $\mathbf{m}_k\left(\mathbf{x}_k^i\right) \leftarrow \frac{1}{M}\sum_{j=1}^M w_{ik} z_j \left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^j\right)$
    $\mathbf{S}_k\left(\mathbf{x}_k^i\right) \leftarrow \frac{1}{M}\sum_{j=1}^M w_{ik} z_j \left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^j\right)\left(\mathbf{x}_{k+1}^j - \mathbf{x}_k^j\right)^T$
  end for
end for

computed as follows:

$$v_{\text{test},j} = f\left(\mathbf{x}_k^j|\mathbf{x}_k^{\text{test}}\right) \sim N\left(\mathbf{x}_k^j; \mathbf{x}_k^{\text{test}}, \mathbf{\Psi}\right).$$

In addition, the weights must be normalized so that $\sum_{j=1}^M v_j = 1$. Now, the two functions needed for the optimal portfolio weight vector are approximated as:

$$\begin{aligned}
\mathbf{m}_k\left(\mathbf{x}_k^{\text{test}}\right) &\simeq \sum_{j=1}^M v_{\text{test},j}\mathbf{m}_k^j, \\
\mathbf{S}_k\left(\mathbf{x}_k^{\text{test}}\right) &\simeq \sum_{j=1}^M v_{\text{test},j}\mathbf{S}_k^j,
\end{aligned}$$

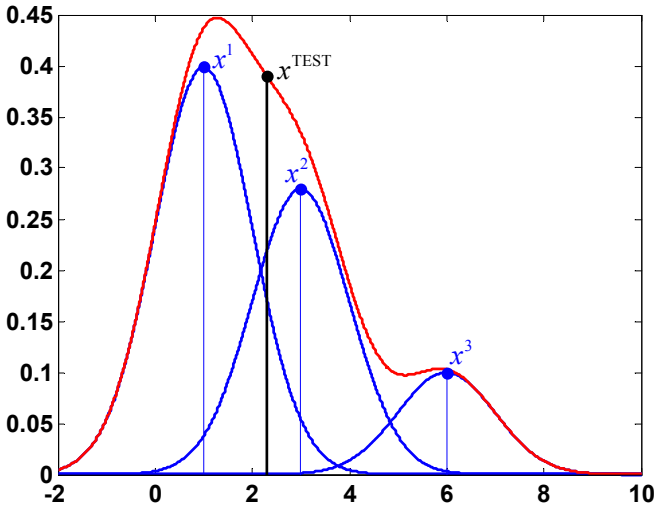


Figure 5-3. Illustration of kernel density estimate.

The portfolio weight vector at time t_k is computed as follows:

$$\mathbf{w}_k^* \left(\mathbf{x}_k^{\text{test}}, r_k\right) \quad = \quad \frac{1}{2\lambda_N} \left(\gamma_N - 2\lambda_N r_k\right) \mathbf{S}_k^{-1} \left(\mathbf{x}_k^{\text{test}}\right) \mathbf{m}_k \left(\mathbf{x}_k^{\text{test}}\right).$$

The full testing phase procedure is summarized in Algorithm 2.

Thus, it is possible to utilize Monte Carlo and importance sampling methods to derive an algorithm for estimating a set of state-dependent recursive moments required for the computation of the dynamically optimal sequence of mean-variance portfolio policies, given by Eq. 5.16. The resulting procedure consists of a training phase, in which the functions are approximated over a grid of sample paths of the underlying log-price process, and a testing phase, in which the functions are evaluated on new data by interpolating between the training samples using an estimator based on a Gaussian kernel. The last step in deriving the full dynamic MVO solution is to determine how to initially select the constants γ_N and λ_N of Problem P_1 so that the variance constraint from Problem P_0 is satisfied, as addressed in Section 5.1.4 next.

5.1.4 Computation of Total Risk

Mapping the quadratic utility auxiliary Problem P_1 back to the original mean-variance objective defined in Problem P_0 is achieved by determining the relationship between γ_N and λ_N with σ_0, as depicted in Figure 5-4. First, observe that the constant γ_N can be factored

Algorithm 2 Testing Phase: Determining the optimal sequence of asset allocation policies.

STEP 1: Fix $\{\gamma_N, \lambda_N\}$
STEP 2: Compute optimal portfolio weight vectors
for $k = 0$ to $N-1$ **do**
  for $i = 1$ to M **do**
    for $j = 1$ to M **do**
      $v_{ij} \leftarrow f\left(\mathbf{x}_k^j | \mathbf{x}_k^i\right)$
    end for
    $v_{ij} \leftarrow \frac{v_{ij}}{\sum_j v_{ij}}$
    $\mathbf{m}_k\left(\mathbf{x}_k^i\right) \leftarrow \sum_{j=1}^{M} v_{ij} \mathbf{m}_k^j\left(\mathbf{x}_k^j\right)$
    $\mathbf{S}_k\left(\mathbf{x}_k^i\right) \leftarrow \sum_{j=1}^{M} v_{ij} \mathbf{S}_k^j\left(\mathbf{x}_k^j\right)$
    $\mathbf{w}_{ik}^*\left(\mathbf{x}_k^i, r_k\right) \leftarrow \frac{1}{2\lambda_N}\left(\gamma_N - 2\lambda_N r_k\right) \mathbf{S}_k^{-1}\left(\mathbf{x}_k^i\right) \mathbf{m}_k\left(\mathbf{x}_k^i\right)$
  end for
end for

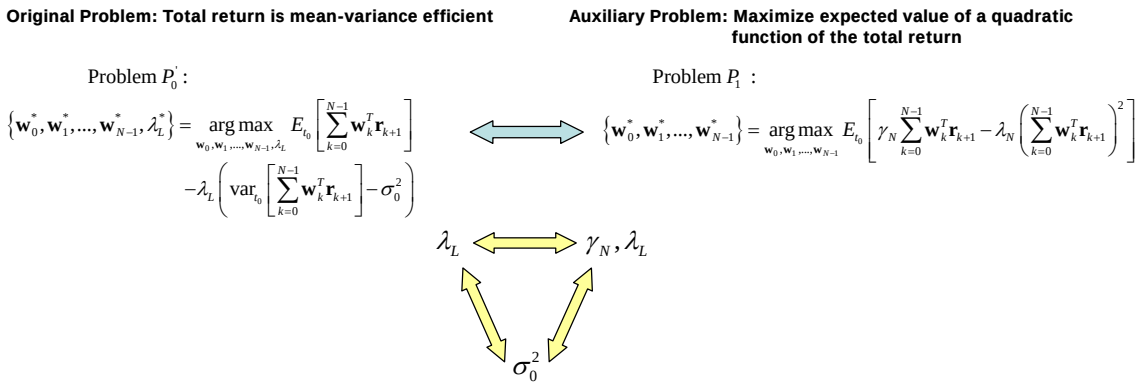


Figure 5-4. The mapping from problem P_1 back to the original mean-variance objective given in Problem P_0' can be established by directly determining the relationship between γ_N and λ_N with σ_0^2, bypassing the intermediate step of mapping from γ_N and λ_N to λ_L.

out of the objective function in Problem P_1, without affecting the overall solution. The resulting quadratic term is scaled by $\frac{\lambda_N}{\gamma_N}$, and therefore it is only this ratio, not the absolute values of these constants, that matters. As a result, γ_N may be set to unity, leaving only one free parameter.

Recall from Eq. 5.16 that the portfolio weight vector for the first stage is given by:

$$\mathbf{w}_0 \quad = \quad \frac{\gamma_N}{2\lambda_N} \mathbf{S}_0^{-1} \mathbf{m}_0 = \frac{1}{\lambda_N} \mathbf{v}_0. \tag{5.20}$$

The above expression emphasizes the relationship between the portfolio weight vector and λ_N. For the second stage, the relationship between $\mathbf{w}_1$ and λ_N is given by:

$$\begin{aligned}\mathbf{w}_1 &= \frac{1}{2\lambda_N}\left(\gamma_N - 2\lambda_N r_1\right)\mathbf{S}_1^{-1}\mathbf{m}_1 = \frac{1}{2\lambda_N}\left(\gamma_N - 2\lambda_N \mathbf{w}_0^T\left(\mathbf{x}_1 - \mathbf{x}_0\right)\right)\mathbf{S}_1^{-1}\mathbf{m}_1, \\ &= \frac{1}{2\lambda_N}\left(\gamma_N - 2\mathbf{v}_0^T\left(\mathbf{x}_1 - \mathbf{x}_0\right)\right)\mathbf{S}_1^{-1}\mathbf{m}_1 = \frac{1}{\lambda_N}\mathbf{v}_1,\end{aligned}$$

which has identical structure to Eq. 5.20. A similar procedure can be carried out for all of the remaining stages, generating a recursive representation for the portfolio weights with the term $\frac{1}{\lambda_N}$ factored out. This factorization facilitates the easy computation of the variance of the total cumulative portfolio return, σ_0^2, as follows:

$$\sigma_0^2 = \mathrm{var}_{t_0}\left[\sum_{k=1}^{N} r_k\right] = \mathrm{var}_{t_0}\left[\sum_{k=0}^{N-1} \mathbf{w}_k^T\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right] = \mathrm{var}_{t_0}\left[\frac{1}{\lambda_1}\sum_{k=0}^{N-1} \mathbf{v}_k^T\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right].$$

The relationship between λ_N and σ_0 is given by:

$$\lambda_N = \frac{\mathrm{std}_{t_0}\left[\sum_{k=0}^{N-1} \mathbf{v}_k^T\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right]}{\sigma_0} \quad . \tag{5.21}$$

Due to the lack of closed-form expressions for the terms $\mathbf{m}_k$ and $\mathbf{S}_k$, the above expression for λ_N cannot be computed in closed-form, and Monte Carlo methods must again be employed. However, the numerator only needs to be estimated once for $\sigma_0 = 1$, as additional values scale according to Eq. 5.21.

5.1.5 Formulation as Linear-Quadratic Regulator

One important class of sequential decision making problems are known as the Linear-Quadratic Regulator (LQR) problem, which is characterized by linear state evolution and an additive per-stage cost that is a quadratic function of the state and control variables. These problems have been extensively studied in the literature [13], and can be solved analytically using the well-known Riccati equations. While the original dynamic mean-variance objective given in Problem P_0 cannot be directly mapped into this framework as explained in Section 2.4, the quadratic utility version defined by Problem P_1 can. The resulting system is a special case of an LQR, in which the state evolution matrices are themselves stochastic. While formulation within the LQR framework provides a convenient and well-known

representation for the problem of constructing MVO portfolios in systems where the log-prices evolve according to cointegrated VAR dynamics, it does not eliminate accompanying computational difficulties, as shown here.

In order to map Problem P_1 into the LQR framework, let $\mathbf{y}_k \in \mathbb{R}^{n\times 1}$ denote the system state, defined as:

$$\mathbf{y}_k \quad = \quad \begin{pmatrix} r_k \\ \mathbf{w}_k \end{pmatrix},$$

where r_k and $\mathbf{w}_k$ denote the cumulative return and portfolio weight vector at time t_k, respectively. The state vector evolves according to the following linear system:

$$\mathbf{y}_{k+1} = \mathbf{A}_k \mathbf{y}_k + \mathbf{B}\mathbf{u}_k + \boldsymbol{\epsilon}_k,$$

where the vector $\mathbf{u}_k \in \mathbb{R}^{m\times 1}$ represents the set of actions to be taken at stage k, here corresponding to the asset rebalancing task, and the vector $\boldsymbol{\epsilon}_k \in \mathbb{R}^{n\times 1}$ corresponds to the stochastic input process driving the log-price process, defined by Eq. 5.1. The matrices $\mathbf{A}_k \in \mathbb{R}^{n\times n}$ and $\mathbf{B}_k \in \mathbb{R}^{n\times m}$ are given by:

$$\mathbf{A}_k = \begin{pmatrix} 1 & \Delta\mathbf{x}_{k+1}^T \\ \mathbf{0} & \mathbb{I} \end{pmatrix}, \qquad \mathbf{B}_k = \begin{pmatrix} \Delta\mathbf{x}_{k+1}^T \\ \mathbb{I} \end{pmatrix}.$$

The asset log-prices, $\mathbf{x}_k$, are not included as part of the state vector, but rather appear within the state transition matrices. This "trick" works due to the additional assumption that the control action does not influence the asset log-prices (i.e., the trade does not *move the market*). In this way, it is possible to express an otherwise non-linear system using the LQR framework, at the expense of creating time-dependent, random system matrices. The optimal set of actions is chosen to maximize the following quadratic cost function:

$$\underset{\epsilon_k, k=0,\ldots,N-1}{E} \left\{ \mathbf{y}_N^T \mathbf{Q}_N \mathbf{y}_N + \sum_{k=0}^{N-1} \left(\mathbf{y}_k^T \mathbf{Q}_k \mathbf{y}_k + \mathbf{u}_k^T \mathbf{R}_k \mathbf{u}_k \right) \right\},$$

where the matrices $\mathbf{Q}_k \in \mathbb{R}^{n\times n}$ and $\mathbf{R}_k \in \mathbb{R}^{m\times m}$ for $k = 0, ..., N-1$ are equivalently equal to the zero matrix, and the terminal cost, $\mathbf{Q}_N \in \mathbb{R}^{n\times n}$, is given by:

$$\mathbf{Q}_N = \begin{pmatrix} \lambda_N & 0 \\ 0 & 0 \end{pmatrix}.$$

Initializing the cumulative return to $r_0 = \frac{\gamma_N}{2\lambda_N}$ produces the desired quadratic objective function defined in Problem P_1. All of the *reward* is attributed to the final stage, as the investor acts so as to maximize his or her expected utility of terminal wealth.

Due to the presence of the term $\Delta\mathbf{x}_{k+1}$, both $\mathbf{A}_k$ and $\mathbf{B}_k$ are time-varying and stochastic, and therefore the standard LQR solution does not apply. As shown in [13], the optimal control law is still a linear function of the state, with form:

$$\mathbf{u}_k = -\left(\mathbf{R}_k + E\left[\mathbf{B}_k^T\mathbf{K}_{k+1}\mathbf{B}_k\right]\right)^{-1} E\left[\mathbf{B}_k^T\mathbf{K}_{k+1}\mathbf{A}_k\right]\mathbf{y}_k = \mathbf{L}_k\mathbf{y}_k,$$

where the matrices $\mathbf{K}_k$ evolve recursively according to:

$$\begin{aligned} \mathbf{K}_N &= \mathbf{Q}_N \\ \mathbf{K}_k &= E\left[\mathbf{A}_k^T\mathbf{K}_{k+1}\mathbf{A}_k\right] - E\left[\mathbf{A}_k^T\mathbf{K}_{k+1}\mathbf{B}_k\right]\left(\mathbf{R}_k + E\left[\mathbf{B}_k^T\mathbf{K}_{k+1}\mathbf{B}_k\right]\right)^{-1} E\left[\mathbf{B}_k^T\mathbf{K}_{k+1}\mathbf{A}_k\right] + \mathbf{Q}_k. \end{aligned}$$

It can be shown that the resulting set of moments needed here are identical to the moments required for the computation of Eqs. 5.17 through 5.19. On the one hand, the LQR framework provides a convenient and well-known representation for the problem of constructing dynamic MVO portfolios under the assumption that the log-prices evolve according to a cointegrated VAR process. However, such a formulation does not eliminate accompanying computational difficulties, and approximations schemes, such as the Monte Carlo based methods proposed in Section 5.1.3, must still be utilized.

5.2 Model Extensions

Portfolio choice problems often include a budget constraint to ensure that all the investor's assets are accounted for among the available investment vehicles. Without a budget constraint, there is an implicit assumption that the investor can borrow additional capital for

free, and as a result, the optimal solution often utilizes leverage to achieve a higher expected return for a given level of risk. In the real world, however, investors must pay interest on borrowed funds, at the so-called risk-free rate, r_f, and this expense must be taken into account when determining the optimal portfolio weights. Therefore, it is common to explicitly incorporate both a budget constraint and a risk-free asset into the model, as discussed in Sections 5.2.1 and 5.2.2, respectively.

5.2.1 Budget Constraint

The dynamic MVO portfolio choice problem considered thus far does not include a budget constraint of the form $\mathbf{w}^T\mathbf{1} = 1$, as discussed in Section 2.2.3. Such a constraint can easily be incorporated into a dynamic portfolio choice setting, as presented here.

Recall that in the static portfolio choice setting, the budget constraint is enforced by representing the portfolio weight vector in a $p-1$ dimensional space, using the affine transformation $\mathbf{w}_N = \mathbf{c}+\mathbf{D}\mathbf{v}_N$, as defined by Eq. 2.7. In a dynamic multiperiod setting, the vector $\mathbf{c}$ must be replaced with the term $\mathbf{c}\,(1+r_k)$, where r_k represents the *cumulative realized* return up to time t_k. This change captures the fact that at the beginning of each stage, the investor must allocate all his *current wealth* among the risky and risk-free assets, not just his *initial wealth*. The value function one stage from the end can now be rewritten as follows:

$$\begin{aligned}
J^*_{N-1}(r_{N-1}) &= \max_{\mathbf{w}_{N-1}} \; E_{t_{N-1}}\left[J_N(r_N)\right] = \max_{\mathbf{w}_N} E_{t_{N-1}}\left[\gamma_N r_N - \lambda_N r_N^2\right] \\
&= \max_{\mathbf{w}_{N-1}} \; E_{t_{N-1}}\left[\gamma_N\left(r_{N-1}+\mathbf{w}_{N-1}^T(\mathbf{x}_N-\mathbf{x}_{N-1})\right) - \lambda_N\left(r_{N-1}+\mathbf{w}_{N-1}^T(\mathbf{x}_N-\mathbf{x}_{N-1})\right)^2\right] \\
&= \max_{\mathbf{v}_{N-1}} \; r_{N-1}\alpha_{N-1} - r_{N-1}^2\beta_{N-1} + \left(\mathbf{c}\,(1+r_{N-1})+\mathbf{D}\mathbf{v}_{N-1}\right)^T\left(\mathbf{m}_{N-1}-2r_{N-1}\mathbf{n}_{N-1}\right) \\
&\qquad -\left(\mathbf{c}\,(1+r_{N-1})+\mathbf{D}\mathbf{v}_{N-1}\right)^T\mathbf{S}_{N-1}\left(\mathbf{c}\,(1+r_{N-1})+\mathbf{D}\mathbf{v}_{N-1}\right) \qquad (5.22)
\end{aligned}$$

where:

$$\begin{aligned}
\alpha_{N-1} &= E_{t_{N-1}}\left[\gamma_N\right], \\
\beta_{N-1} &= E_{t_{N-1}}\left[\lambda_N\right], \\
\mathbf{m}_{N-1} &= E_{t_{N-1}}\left[\gamma_N\left(\mathbf{x}_N - \mathbf{x}_{N-1}\right)\right], \\
\mathbf{n}_{N-1} &= E_{t_{N-1}}\left[\lambda_N\left(\mathbf{x}_N - \mathbf{x}_{N-1}\right)\right], \\
\mathbf{S}_{N-1} &= E_{t_{N-1}}\left[\lambda_N\left(\mathbf{x}_N - \mathbf{x}_{N-1}\right)\left(\mathbf{x}_N - \mathbf{x}_{N-1}\right)^T\right].
\end{aligned}$$

The optimal value for the reduced-dimension portfolio weight vector over the last stage is determined by setting the derivative of the objective function equal to zero, yielding:

$$\mathbf{v}^*_{N-1} = 0.5\left(\mathbf{D}^T\mathbf{S}_{N-1}\mathbf{D}\right)^{-1}\mathbf{D}^T\left(\mathbf{m}_{N-1} - 2r_{N-1}\mathbf{n}_{N-1} - 2\mathbf{D}^T\mathbf{S}_{N-1}\mathbf{c}\left(1 + r_{N-1}\right)\right).$$

The value of $\mathbf{v}^*_{N-1}$ is then substituted back into Eq. 5.22, which can then be rearranged into a quadratic function of r_{N-1}, as follows:

$$J^*_{N-1}\left(r_{N-1}\right) = \gamma_{N-1}r_{N-1} - \lambda_{N-1}r^2_{N-1} + c_{N-1},$$

where:

$$\begin{aligned}
\lambda_{N-1} &= \beta_{N-1} + 2\mathbf{g}^T_{N-1}\mathbf{n}_{N-1} + \mathbf{g}^T_{N-1}\mathbf{S}_{N-1}\mathbf{g}_{N-1}, \\
\gamma_{N-1} &= \alpha_{N-1} + \mathbf{g}^T_{N-1}\mathbf{m}_{N-1} - 2\mathbf{f}^T_{N-1}\left(\mathbf{n}_{N-1} + \mathbf{S}_{N-1}\mathbf{g}_{N-1}\right), \\
c_{N-1} &= \mathbf{f}^T_{N-1}\mathbf{m}_{N-1} + \mathbf{f}^T_{N-1}\mathbf{S}_{N-1}\mathbf{f}_{N-1}, \\
\mathbf{f}_{N-1} &= \mathbf{c} + 0.5\mathbf{D}\left(\mathbf{D}^T\mathbf{S}_{N-1}\mathbf{D}\right)^{-1}\mathbf{D}^T\left(\mathbf{m}_{N-1} - 2\mathbf{S}_{N-1}\mathbf{c}\right), \\
\mathbf{g}_{N-1} &= \mathbf{c} - \mathbf{D}\left(\mathbf{D}^T\mathbf{S}_{N-1}\mathbf{D}\right)^{-1}\mathbf{D}^T\left(\mathbf{n}_{N-1} + \mathbf{S}_{N-1}\mathbf{c}\right).
\end{aligned}$$

The above procedure can then be repeated to solve for $\{\mathbf{v}_{N-2}, \ldots, \mathbf{v}_0\}$. Therefore, when a budget constraint is included, the portfolio weight vector at time t_k is computed as:

$$\mathbf{w}^*_k\left(\mathbf{x}_k, r_k\right) = \mathbf{c} + 0.5\mathbf{D}\left(\mathbf{D}^T\mathbf{S}_k\mathbf{D}\right)^{-1}\mathbf{D}^T\left(\mathbf{m}_k - 2r_k\mathbf{n}_k - 2\mathbf{D}^T\mathbf{S}_k\mathbf{c}\left(1 + r_k\right)\right).$$

The relationship between λ_N and σ^2_0 is again determined as shown in Section 5.1.4.

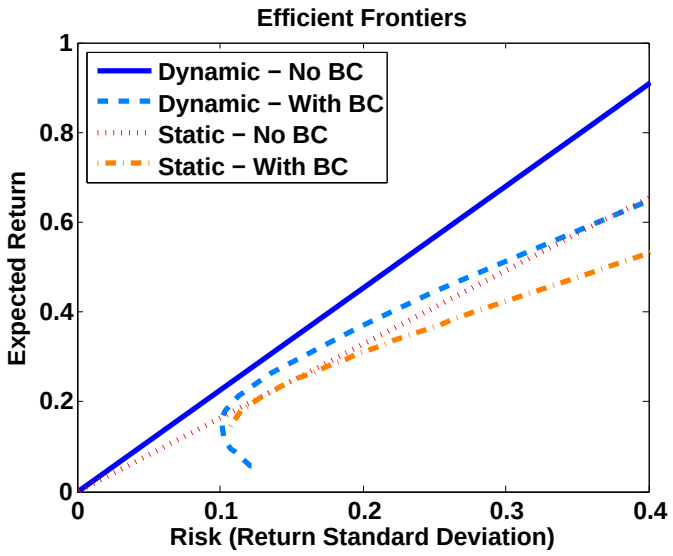


Figure 5-5. Efficient frontiers for static and dynamic MVO portfolios, with and without a budget constraint.

The following example illustrates how the multistage dynamic mean-variance efficient frontier is affected by the presence of a budget constraint.

Example 5.1.

Consider again the system of two risky assets from Ex. 4.2, in which the log-prices are assumed to evolve according to the following cointegrated VAR process:

$$\mathbf{x}_{k+1} = \begin{pmatrix} 0.7878 & 0.0707 \\ 0.2634 & 0.9122 \end{pmatrix} \mathbf{x}_k + \boldsymbol{\epsilon}_k,$$

where $\boldsymbol{\epsilon}_k \sim N(\mathbf{0}, \boldsymbol{\Psi})$ and $\boldsymbol{\Psi} = \begin{pmatrix} 0.0400 & 0 \\ 0 & 0.0049 \end{pmatrix}$. The initial condition is chosen to be $\begin{pmatrix} 1.7500 & 4.3000 \end{pmatrix}^T$. During the training phase, the portfolio weight vector functions are numerically approximated using a grid of $M = 5000$ sample paths. This is followed by a testing phase in order to generate the efficient frontier shown in Fig. 5-5 using a set of $T = 5000$ test paths. Recall that removal of the budget constraint implies the investor can borrow for free, evidenced here by the fact that the corresponding efficient frontier line intercepts the y-axis at the origin. In contrast to the static portfolio choice problem considered in Chapter 4, the set of MVO portfolios for the case where the budget constraint is omitted is no longer tangent to the efficient frontier with the budget constraint. Tangency

implies that the same percentage of wealth is invested in the risk-free asset over both stages, whereas the true dynamic MVO solution is allowed to change the amount allocated to the risk-free asset as needed. Figure 5-5 compares the efficient frontiers for this example using both the dynamic and static MVO solutions, illustrating that the dynamic strategy does indeed outperform its static counterpart, both with and without the budget constraint. ■

5.2.2 Risk-free rate

In addition to including a budget constraint in the portfolio choice problem, it is often desirable to explicitly model the availability of a risk-free asset for borrowing and lending. In this case, the vector of individual asset returns, $\mathbf{r}_k$, defined by Eq. 5.2, is augmented as follows:

$$\mathbf{r}_k = \begin{pmatrix} \Delta \mathbf{x}_k \\ r_f \end{pmatrix} = \begin{pmatrix} \mathbf{x}_k - \mathbf{x}_{k-1} \\ r_f \end{pmatrix}. \tag{5.23}$$

In addition, the length of the portfolio weight vector, $\mathbf{w}_k$, is increased by one to account for the additional asset, and the functions $\mathbf{m}_k$, $\mathbf{n}_k$, and $\mathbf{S}_k$ must be augmented as follows:

$$\begin{aligned}
\mathbf{m}_k &= E_{t_k}\left[\begin{matrix} \gamma_{k+1}\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right) \\ \gamma_{k+1} r_f \end{matrix}\right] = \begin{pmatrix} E\left[\gamma_{k+1}\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right] \\ \gamma_{k+1} r_f \end{pmatrix}, \\
\mathbf{n}_k &= E_{t_k}\left[\begin{matrix} \lambda_{k+1}\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right) \\ \lambda_{k+1} r_f \end{matrix}\right] = \begin{pmatrix} E\left[\lambda_{k+1}\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right] \\ \lambda_{k+1} r_f \end{pmatrix}, \\
\mathbf{S}_k &= E_{t_k}\left[\begin{matrix} \lambda_{k+1}\left(\mathbf{x}_N - \mathbf{x}_k\right)\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)^T \\ \lambda_{k+1} r_f^2 \end{matrix}\right] = \begin{pmatrix} E\left[\lambda_{k+1}\left(\mathbf{x}_N - \mathbf{x}_k\right)\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)^T\right] \\ \lambda_{k+1} r_f^2 \end{pmatrix}.
\end{aligned}$$

Thus the presence of a risk-free asset is easily incorporated into the model through augmentation of the asset return state vector. For a given portfolio weight vector, a risk-free weight of $w_{r_f} \in [0, 1]$ represents investing $w_{r_f}\%$ in the risk-free investment, while a weight of $w_{r_f} < 0$ represents borrowing $w_{r_f}\%$ of the available wealth in order to increase exposure to the set of risky investments. The following example demonstrates how the presence of the risk-free asset influences the shape of the efficient frontier.

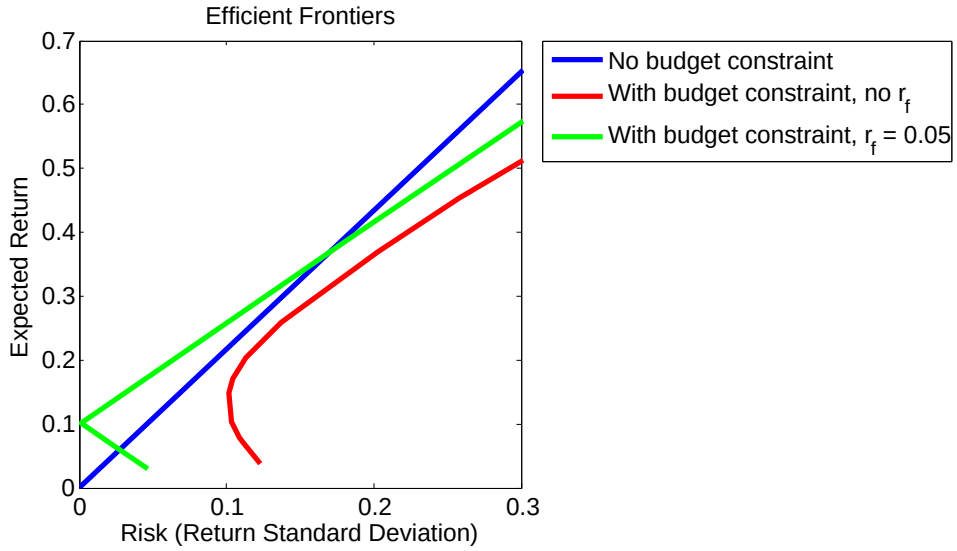


Figure 5-6. Efficient frontier for system in Ex. 5.2 with risk-free rate $r_f = 0.05$. The system with the budget constraint and risk-free asset initially outperforms the alternative schemes. However, at some point the advantage earned from borrowing capital at the higher risk-free rate is outweighed by the additional cost of interest paid to the lender.

Example 5.2.

Consider the system of two risky assets presented in Example 5.1. Now suppose the model is augmented to include a risk-free investment opportunity, with corresponding risk-free rate of $r_f = 5\%$. The resulting efficient frontier is shown in Fig. 5-6, along with the efficient frontiers for the case where the budget constraint is omitted and the case with a budget constraint but no risk-free asset. As expected, the new efficient frontier is a straight line that intercepts the y-axis at the rate earned by keeping all of the available wealth in the risk-free asset for both stages, computed as:

$$r_T = (1.05)^2 - 1 = 0.1025.$$

As Fig. 5-6 reveals, the new efficient frontier for the system with the budget constraint and the risk-free asset initially outperforms the alternative schemes. However, at some point the blue and green lines intersect, corresponding to the regime where any advantage earned from borrowing capital at the higher risk-free rate is outweighed by the additional cost of interest paid to the lender. ■

5.2.3 No short-sale constraint

On September 19, 2008, the Securities and Exchange Commission (SEC) enacted a temporary short-selling ban on the stock of 799 financial companies, in order to "protect the integrity and quality of the securities market and strengthen investor confidence" [1]. While this action was only a short-term emergency measure, it is possible that similar short-selling bans may come into effect permanently as the SEC introduces new regulation in reaction to the ongoing global financial crisis. Therefore, it is of interest to study the impact of such constraints on the trading strategies presented in this thesis. First, brief summaries of how to incorporate the no short-sale constraint into both the optimal static and dynamic MVO strategies are given, followed by the presentation of two illustrative examples.

In Section 4.2, closed-form solutions for the multistage static portfolio choice problem are presented, both with and without a budget constraint. By introducing a no short-sale constraint, Problem P_0 may be rewritten as Problem Q_0:

$$\left.\begin{aligned} \{\mathbf{w}_N^*\} = \arg\max_{\mathbf{w}_N} \ & \mathbf{w}_N^T \boldsymbol{\mu}_{\mathbf{r}_N} \\ \text{s.t. } & \mathbf{w}_N^T \boldsymbol{\Sigma}_{\mathbf{r}_N} \mathbf{w}_N = \sigma_0^2 \\ & \mathbf{w}_N \succeq 0 \end{aligned}\right\} \ Q_0.$$

While Q_0 is still a quadratic program that must be solved once at time t_0, it no longer admits a closed-form solution, and a numerical optimization search routine must be utilized. When using numerical search, the budget constraint is easily incorporated by adding a linear constraint of the form $\mathbf{w}_N^T \mathbf{1} = 1$, rather than using the affine transformation described in Section 4.2.2. The resulting efficient frontier is a proper subset of the efficient frontier for the unconstrained version, and corresponds to a finite range of risk levels (i.e., for some choices of $sigma_0$ no feasible solution to Problem Q_0 exists).

In the dynamic MVO solution derived in Section 5.1.2, closed-form solutions for the optimal portfolio weight vectors $\mathbf{w}_k$ exist as functions of the moments $\mathbf{m}_k(\mathbf{x}_k)$ and $\mathbf{S}_k(\mathbf{x}_k)$. As with the static case, introducing a no short-sale constraint of the form $\mathbf{w}_N \succeq 0$ necessitates the use of numerical search at each stage to solve the resulting quadratic program for the optimal portfolio policy. Specifically, at time t_k, the optimal portfolio policy $\mathbf{w}_k^*$ is found

by solving Problem Q_1, as follows:

$$\left.\begin{array}{r} \{\mathbf{w}_k^*\} = \arg\max_{\mathbf{w}_k} \ (\gamma_N - 2\lambda_N r_k)\,\mathbf{w}_k^T\mathbf{m}_k - \lambda_N \mathbf{w}_k^T\mathbf{S}_k\mathbf{w}_k \\ \text{s.t. } \mathbf{w}_N \succeq 0 \end{array}\right\} \ Q_1.$$

Again, the budget constraint is easily incorporated by adding a linear constraint of the form $\mathbf{w}_k^T\mathbf{1} = 1 + r_k$ to Problem Q_1. It is important to note that the Monte Carlo methods required to compute $\mathbf{m}_k(\mathbf{x}_k)$ and $\mathbf{S}_k(\mathbf{x}_k)$ are unaffected by the presence of the additional constraint; only the manner in which they are used is altered by the short-sale ban.

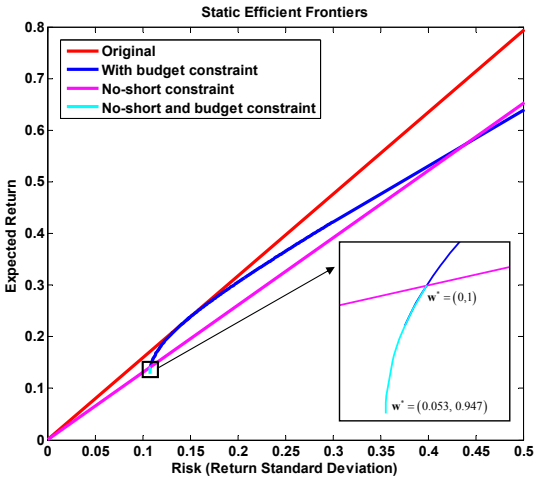
The following example illustrates the impact of a no short-sale constraint on the example system studied throughout this chapter.

Example 5.3.

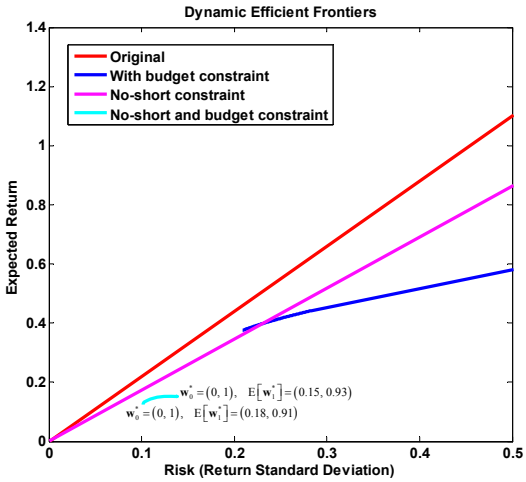
Consider again the system of two risky assets from Ex. 5.1, in which the log-prices are assumed to evolve according to a first-order cointegrated VAR system with parameters:

$$\mathbf{\Pi}_1 = \begin{pmatrix} 0.7878 & 0.0707 \\ 0.2634 & 0.9122 \end{pmatrix}, \qquad \mathbf{\Psi} = \begin{pmatrix} 0.2^2 & 0 \\ 0 & 0.07^2 \end{pmatrix},$$

and arbitrary initial condition $\mathbf{x}_0 = \begin{pmatrix} 1.75 & 4.3 \end{pmatrix}^T$. The efficient frontiers for a two-stage investment horizon are shown in Figure 5-7, for the four scenarios corresponding to the inclusion of a budget and no-short constraint. When the MVO static strategy of Ch. 4 is used, the efficient frontier when both the budget and no-short constraints are active is a subset of the efficient frontier when just the budget constraint is active, as shown in Fig. 5-7(a). In addition, the efficient frontier for the budget/no short-sale case does not include all convex combinations of the portfolios between $\mathbf{w}^* = (1, 0)$ and $\mathbf{w}^* = (0, 1)$. Only operating points between the portfolios $\mathbf{w}^* = (0.053, 0.947)$ and $\mathbf{w}^* = (0, 1)$ are shown, as this segment corresponds to the upper half of the budget constrained efficient frontier parabola. The rest of the operating points between $\mathbf{w}^* = (1, 0)$ and $\mathbf{w}^* = (0.053, 0.947)$ reside on the lower half of the parabola and can be achieved by solving the related mean-variance problem of minimizing the variance subject to an equality constraint on the expected return, with both a budget and no short-sale constraint.



(a) Impact of no short-sale constraint on MVO static portfolio choice for two-stage, two-asset example.

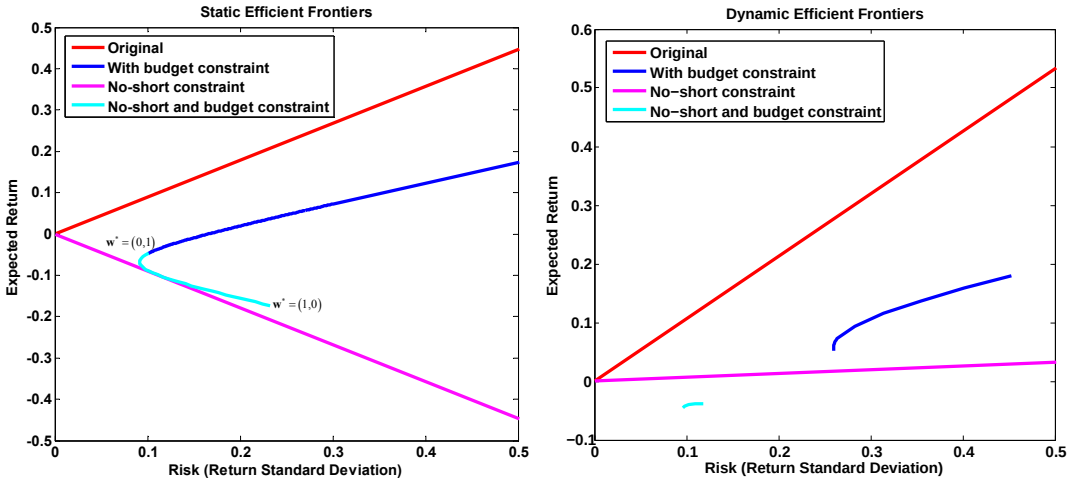


(b) Impact of no short-sale constraint on MVO dynamic portfolio choice for two-stage, two-asset example.

Figure 5-7. Efficient frontiers in two-stage, two-asset example, with no short-sale constraint.

When the dynamic MVO solution of Ch. 5 is used, the efficient frontier for the budget/no-short case is no longer a subset of the efficient frontier with only the budget constraint, as depicted in Fig. 5-7(b), as the two strategies do not form the same portfolio weights for both stages. It is also interesting to note that in the dynamic case, the portfolio weights for the first stage sweep across a wider range of convex combinations of the two assets as compared to the static solution. The remainder of the operating points corresponding to portfolios between $\mathbf{w}_0^* = (1, 0)$ and $\mathbf{w}_0^* = (0.285, 0.715)$ are achievable by solving the related minimum variance problem. ■

As this example illustrates, while a ban on short-sales limits the performance of both the static and dynamic MVO solutions, it does not necessarily imply that no profitable trading strategies exist. For the system presented here, there are still mean-variance efficient operating points that earn a positive expected return and satisfy the no short-sale constraint, both with and without a budget constraint. However, it is possible to construct a system in which there are no MVO operating points that satisfy both the budget and no short-sale constraint and have positive expected return, as illustrated in Ex. 5.4.



(a) In this static MVO example, there is no feasible region with positive expected return when the no short-sale constraint is enforced, with or without the budget constraint.

(b) In contrast to the static case, when using the dynamic MVO solution there is a feasible region corresponding to the case where only the no short-sale constraint is enforced. Without the budget constraint, the investor has the option to sit out during the first stage, waiting for a potential second-stage, long-only opportunity to present.

Figure 5-8. Efficient frontiers in two-stage, two-asset example, with no short-sale constraint, when first stage portfolio vectors are all short.

Example 5.4.

In the system from Ex. 5.3, the matrix $\mathbf{\Pi}$ can be factored according to:

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^T = \begin{pmatrix} -0.37 \\ 0.46 \end{pmatrix} \begin{pmatrix} 0.57 & -0.19 \end{pmatrix}.$$

The system for this new example is generated by replacing $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}' = \begin{pmatrix} -0.58 & -0.15 \end{pmatrix}^T$, so that the direction of expected change is negative for both assets. The new efficient frontiers for a two-stage investment horizon are shown in Figure 5-8, again for the four scenarios corresponding to the inclusion of a budget and no-short constraint, using both the static and dynamic MVO solutions. Unlike, the previous example, here regions of the efficient frontiers corresponding to both positive and negative expected returns are shown. As illustrated in Fig. 5-8(a) for the static case, there are no mean-variance efficient operating points with positive expected return that satisfy the no short-sale constraint, with or without the budget constraint. For the case where only the no short-sale constraint is

enforced, the investor must enter into positions using the risky assets in order to satisfy the budget constraint, even though he will lose money due to his inability to utilize short sales. Sitting out is not an option. On the other hand, in the dynamic case depicted in Fig. 5-8(b), while there are no profitable operating points with both constraints active, there are feasible portfolios when only the short-sale ban is imposed. Solutions along this line correspond to exiting the market for the first stage by choosing $\mathbf{w}_0^* = (0, 0)$, and waiting for a second-stage, long-only opportunity to present. Unlike the static case, the investor can sit out during the first stage, as the variance constraint applies to the total portfolio return, not the per-stage returns. In the dynamic case where both constraints are enforced, the investor is forced to enter into a position for the first stage that is expected to lose money, and unfortunately, any expected gains from the second stage are not enough to cover these losses and produce a net positive mean return. ■

Given the results of these simulations, it may still be possible to construct a profitable dynamic MVO trading strategy using a system of cointegrated assets, even in the presence of a short-sale ban. In the system studied in Ex. 5.4, while there were no efficient operating points corresponding to the use of both a budget and no short-sale constraint, removal of the budget constraint yielded an efficient frontier with positive expected return. Recall from Ch. 2 that removable of the budget constraint corresponds to an implicit assumption that the investor has access to a risk-free asset with rate $r_f = 0$. Hence, the investor was able to sit out of the market for the first stage, effectively allocating all of his money in the risk-free investment, waiting for a long-only investment opportunity to appear.

5.3 Portfolio Properties

The purpose of this section is explore the properties of the dynamic MVO solution derived here, both with and without the budget constraint. The dynamic solution is also compared to the static MVO strategy developed in Chapter 4. First, in Section 5.3.1, the per-stage portfolio return distributions and inter-stage return correlation structure are explored. Second, in Section 5.3.2, the behavior of the dynamic MVO solution as a function of the trading horizon is studied.

5.3.1 Return Statistics

In Section 4.3.1, the per-stage and cumulative portfolio returns for any static asset allocation scheme were shown to be Normally distributed. However, in dynamic asset allocation, the portfolio weight vectors are state dependent random variables, and therefore the resulting per-stage portfolio returns are no longer all equivalently Gaussian, as explored here.

Consider the per-stage portfolio returns in a 2−stage example, with $r_1 = \mathbf{w}_0^T (\mathbf{x}_1 - \mathbf{x}_0)$ and $r_2 = \mathbf{w}_1^T (\mathbf{x}_2 - \mathbf{x}_1)$. While the exact form for $\mathbf{w}_0$ is not known in closed-form, it is not random with respect to the information available at time t_0. Thus, the return for the first stage is distributed as $r_1 \sim N\left(\mathbf{w}_0^T \mathbf{\Pi} \mathbf{x}_0, \mathbf{w}_0^T \mathbf{\Psi} \mathbf{w}_0\right)$. For the second stage, the MVO portfolio weight vector is known in closed-form, and is given by:

$$\begin{aligned} \mathbf{w}_1 &= \frac{1}{2\lambda_N} \left(\gamma_N - 2\lambda_N r_1\right) \mathbf{S}_1^{-1} \left(\mathbf{x}_1\right) \mathbf{m}_1 \left(\mathbf{x}_{N-1}\right) \\ &= \frac{1}{2\lambda_N} \left(\gamma_N - 2\lambda_N \mathbf{w}_0^T \left(\mathbf{x}_1 - \mathbf{x}_0\right)\right) \left(\mathbf{\Psi} + \mathbf{\Pi} \mathbf{x}_1 \mathbf{x}_1^T \mathbf{\Pi}^T\right)^{-1} \mathbf{\Pi} \mathbf{x}_1. \end{aligned}$$

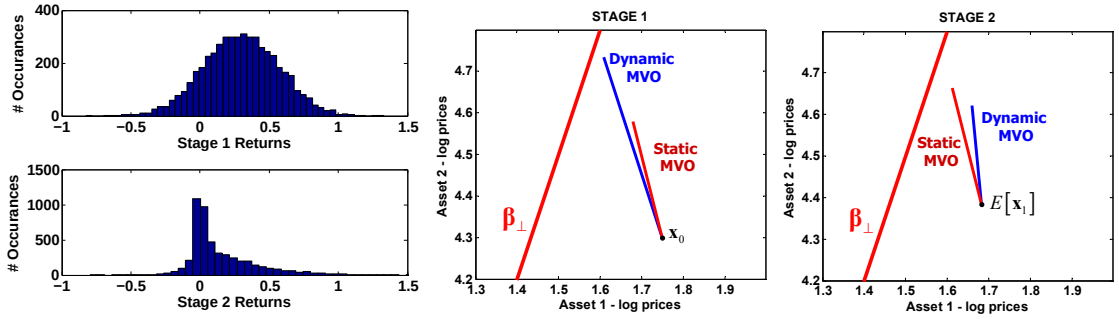
However, despite the fact that $\mathbf{w}_1$ is a known, non-linear function of $\mathbf{x}_1$, it is not possible to determine the exact form of the distribution of r_2, which depends on both $\mathbf{x}_1$ and $\mathbf{x}_2$. In addition, it is not possible to derive an analytic expression for the covariance between r_1 and r_2, as was done in Chapter 4 for the case of static asset allocation. The nature of these statistics must be studied through empirical studies, such as the one presented next.

Example 5.5.

Consider again the system from Ex. 5.1, with a variance budget of $\sigma_0^2 = 0.2^2 = 0.04$. While it is not possible to determine the exact distribution for the portfolio returns over either stage in closed-form, one can examine a histogram of returns, as depicted in Figure 5-9(a).

Strategy	**Stage 1,** r_1					**Stage 2,** r_2					**corr**$[r_1, r_2]$	**Total,** r_T	
	Mean	**Std**	**Weights**			**Mean**	**Std**	**Weights**				**Mean**	**Std**
			w_1	w_2	net lev			w_1	w_2	net lev			
Dynamic: no BC	0.31	0.26	-0.87	2.85	1.98	0.14	0.23	-0.15	1.47	1.32	-0.69	0.44	0.20
Dynamic: with BC	0.24	0.24	-0.99	1.99	1.00	0.14	0.23	-0.38	1.62	1.24	-0.60	0.38	0.20
Static: no BC	0.19	0.16	-0.47	1.86	1.39	0.13	0.20	-0.47	1.86	1.39	-0.39	0.32	0.20
Static: with BC	0.19	0.17	-0.64	1.64	1.00	0.13	0.21	-0.64	1.64	1.00	-0.45	0.31	0.20

Table 5.1. Second-order statistics comparing static vs. dynamic MVO solutions for two-stage example, with total risk budget $\sigma_0 = 0.20$, both with and without a budget constraint.



(a) Empirical distributions for first and second stage portfolio returns in dynamic MVO solution.

(b) In the first stage, the dynamic solution utilizes more leverage, and has a slightly larger component in the $\boldsymbol{\beta}$ direction.

(c) In the second stage, the dynamic solution assumes less leverage, and alters its position away from $\boldsymbol{\beta}$ direction.

Figure 5-9. Per-stage portfolio return empirical distributions and corresponding portfolio weight vectors.

The plots confirm that the returns over the first stage appear Normally distributed, while the returns over the second stage are characterized by a non-symmetric distribution with higher kurtosis (peakedness) than a Gaussian. The corresponding second-order statistics are displayed in Table 5.1. As was true in Ex. 4.2, there is a direct relationship between the total expected return and the degree of negative correlation between the inter-stage portfolio returns. Higher negative correlation between r_1 and r_2 allows the per-stage return variances to assume a larger magnitude, while the total variance remains fixed. The increased amount of per-stage risk is realized through the use of leverage. Whereas the static MVO solution uses a net leverage of 139% across both stages, the dynamic MVO solution first takes on a net leverage of 198%, followed by a total leverage level of 132%.

Figures 5-9(b) and 5-9(c) contrast the portfolio weight vectors for the static and dynamic solutions graphically. The plots highlight not only the different directions taken by each strategy, but also the different degrees of leverage, represented by the relative lengths of the two vectors. In the first stage, the dynamic solution utilizes more leverage than its static counterpart, and has a slightly larger component in the $\boldsymbol{\beta}$ direction. Due to the state dependence of the portfolio policy for the second stage, the figure depicts the average vector direction and length, which utilizes less leverage, and alters its position away from $\boldsymbol{\beta}$ direction, as compared to the static solution. ■

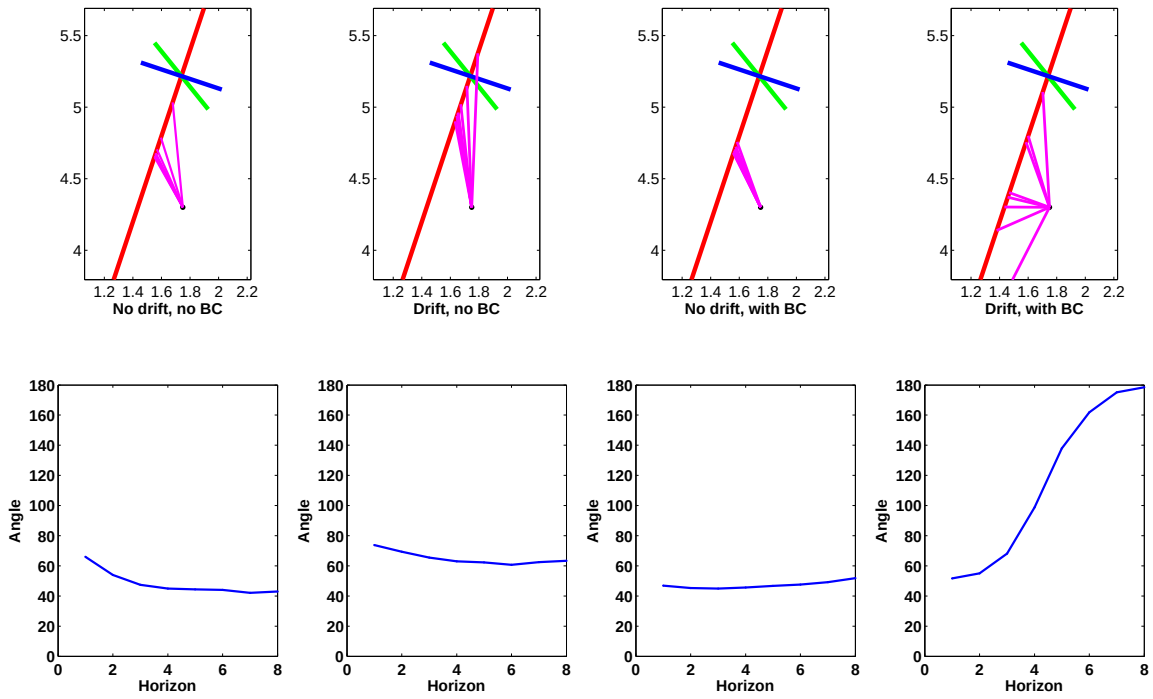


Figure 5-10. Direction of dynamic portfolio weight vectors as a function of investment length.

5.3.2 Asymptotic Analysis

The behavior of the dynamic MVO solution as a function of the total investment horizon is explored. Specifically, for a fixed initial condition, the relationship between the optimal portfolio vector direction over the first stage and the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ subspaces of a cointegrated system is investigated. Recall from Thm. 4.4 of Ch. 4, that when intermediate portfolio rebalancing is prohibited, the weight vector always converges to a steady-state direction, but the value of that direction depends on whether or not drift is present in the log-price model. Only in systems without drift does this vector asymptotically approach the span of the cointegrating space, $\boldsymbol{\beta}$. The availability of a closed-form solution for the static portfolio weight vector facilitated the asymptotic analysis, and the results were confirmed via simulation. In the dynamic setting, analytic solutions for the portfolio weight vectors at each stage are not available, and a numerical approximation scheme is utilized instead. In lieu of closed-form solutions, the corresponding asymptotic analysis can only be performed via Monte Carlo simulation, as illustrated in Example 5.6.

Example 5.6.

Consider the system of two risky assets presented in Ex. 5.1. Figure 5-10 depicts the direction of the dynamic MVO portfolio weight vector as a function of the number of total stages in the problem, for each of the four cases corresponding to the inclusion of a budget constraint and log-price drift. The solutions are represented both graphically, as vectors within the space spanned by the log-prices of the two assets, and numerically, as the angle between the MVO solution and $\boldsymbol{\beta}$ direction. As the figure reveals, in only one of the four cases does the dynamic MVO portflio weight vector asymptotically approach the span of the cointegrating space, $\boldsymbol{\beta}$. One intuitive explanation for this phenomenon is that when rebalancing is permitted, there is some benefit to buying the portfolio with a component in the direction of expected short-term change. ■

5.A Proofs of Chapter 4 Theorems

Proof of Theorem 5.1.

The following is based on the proof of Theorem 1 in [36]. Let $\pi_1^* = \{\mathbf{w}_0^*, ..., \mathbf{w}_{N-1}^*\}$ denote the optimal sequence of portfolio policies for Problem P_1. Assume to the contrary that the set π_1^* does not also solve Problem P_0, the original mean-variance dynamic portfolio choice problem. Furthermore, the objective function of Problem P_0 may be expressed as a convex function of the vector variable $\mathbf{x} = \begin{pmatrix} E[r_N] & E[r_N^2] \end{pmatrix}^T$, as follows:

$$f(\mathbf{x}) = f(E[r_N], E[r_N^2]) = -\lambda_L E[r_N^2] + \left(E[r_N] + \lambda_L E^2[r_N]\right) + \lambda_L \sigma_0^2,$$

where λ_L denotes the Lagrange multiplier for the variance equality constraint. Thus, there exists some other sequence of portfolio policies that is optimal for Problem P_0, denoted by π_0^*, that satisfies the following first-order optimality condition [17]:

$$\begin{aligned} \bigtriangledown f\left(\mathbf{x}|_{\pi_0^*}\right)\left(\mathbf{x}|_{\pi_1^*} - \mathbf{x}|_{\pi_0^*}\right) &> 0 \qquad (5.24) \\ \begin{pmatrix} 1 + 2\lambda_L\, E[r_N]|_{\pi_0^*} & -\lambda_L \end{pmatrix} \begin{pmatrix} E[r_N] \\ E[r_N^2] \end{pmatrix}\Bigg|_{\pi_1^*} &> \begin{pmatrix} 1 + 2\lambda_L\, E[r_N]|_{\pi_0^*} & -\lambda_L \end{pmatrix} \begin{pmatrix} E[r_N] \\ E[r_N^2] \end{pmatrix}\Bigg|_{\pi_0^*}. \end{aligned}$$

Furthermore, since f is convex, the following is also true [17]:

$$f\left(\mathbf{x}|_{\pi_1^*}\right) \geq f\left(\mathbf{x}|_{\pi_0^*}\right) + \bigtriangledown f\left(\mathbf{x}|_{\pi_0^*}\right)\left(\mathbf{x}|_{\pi_1^*} - \mathbf{x}|_{\pi_0^*}\right).$$

However, since the term $\bigtriangledown f\left(\mathbf{x}|_{\pi_0^*}\right)\left(\mathbf{x}|_{\pi_1^*} - \mathbf{x}|_{\pi_0^*}\right)$ is positive by Eq. 5.24, the above implies that $f\left(\mathbf{x}|_{\pi_1^*}\right) \geq f\left(\mathbf{x}|_{\pi_0^*}\right)$. Hence, if π_0^* is optimal for Problem P_0, it must be true that π_1^* is also optimal, which contradicts the original assumption.

Chapter 6

Approximate Dynamic Portfolio Choice

The solution to the dynamic portfolio choice problem presented in Chapter 5 is mean-variance efficient with respect to final wealth. The optimal sequence of portfolio policies is derived by invoking the Principle of Separable Embedding [36], in which the original mean-variance objective function is embedded into an auxiliary quadratic utility problem, which is separable in the sense required by dynamic programming. While an exact solution to the auxiliary problem can be found, the resulting portfolio weight vectors depend on a set of moments that cannot be computed in closed-form. This necessitates the use of Monte Carlo methods, which in turn, introduces approximation error, and the overall method becomes intractable as the number of assets and stages increases.

To address these computational limitations, a set of four approximate dynamic schemes for mean-variance portfolio choice are considered. The name *approximate dynamic* does not refer to a set of techniques for approximating the optimal dynamic solution derived in Chapter 5, but rather refers to the fact that each approximation scheme relaxes at least one of the assumptions of the original dynamic MVO problem in order to derive a suboptimal, yet tractable, trading scheme. As of result of these simplifications, three out of four of the methods admit closed-form solutions, while the fourth only requires the use of a single numerical search routine.

All of the approximation schemes presented here are compared using a series of simulations based on synthetic data. It is shown that the performance of these trading strategies cannot all be ranked on an absolute scale, as some solutions have a regime over which it may

be preferred to alternative methods. In addition, the per-stage statistics of the portfolio returns are studied, and the performance of each scheme is explained in terms of use of leverage and the inter-stage return correlations.

The organization of this chapter is as follows. First, Section 6.1 presents the separable embedding certainty equivalence approximation scheme, in which the stochastic parameters of the per-stage return distributions are replaced with their time t_0 conditional expectations. Next, Section 6.2 details the sequential rescaling approach, in which it is assumed that the solution to an $(N+1)$-stage problem is found by modifying only the scale (i.e., degree of net leverage), not the direction (i.e., relative asset proportions) of the N-stage optimal solution. The third approximation strategy parametrizes the portfolio policy at each stage using a linear function of the log-prices, as derived in Section 6.3. This is followed by a discussion of a semi-myopic approach in Section 6.4, in which the N-stage problem is solved as a series of consecutive single-stage problems. Finally, Section 6.5 concludes with an empirical study of the risk-reward characteristics of all the approximate dynamic portfolio asset allocation schemes considered.

6.1 Separable Embedding Certainty Equivalence

The first relaxation scheme considered in this chapter applies the principle of certainty equivalence to the separable embedding technique utilized in Chapter 5, in order to derive a sequence of optimal portfolio policies that can be computed without the use of numerical approximation methods. By replacing the stochastic parameters of the per-stage return distributions with their expectations, the portfolio weight vectors admit closed-form solutions that depend only on the cumulative realized return.

6.1.1 Problem Formulation

Recall from Chapter 5 that in the dynamic multistage mean-variance portfolio choice framework, the optimal sequence of portfolio policies is computed according to Problem P_0, as follows:

$$\left.\begin{array}{l} \left\{\mathbf{w}_0^*, \mathbf{w}_1^*, \ldots, \mathbf{w}_{N-1}^*\right\} = \underset{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{N-1}}{\arg\max} \; E_{t_0}\left[\sum_{k=0}^{N-1} \mathbf{w}_k^T \mathbf{r}_{k+1}\right] \\ \qquad \text{s.t. } \operatorname{var}_{t_0}\left[\sum_{k=0}^{N-1} \mathbf{w}_k^T \mathbf{r}_{k+1}\right] = \sigma_0^2, \end{array}\right\} \; P_0,$$

where the per-period vector of individual asset returns, $\mathbf{r}_k$, is defined by:

$$\mathbf{r}_k = \Delta\mathbf{x}_k = \mathbf{x}_k - \mathbf{x}_{k-1}, \tag{6.1}$$

and the vector $\mathbf{x}_k$ denotes the log-prices of the assets at time t_k. Under the assumption that the log-prices evolve according to the first-order cointegrated VAR process of Eq. 5.1, the per-stage asset returns are multivariate Gaussian random variables with mean $\mathbf{\Pi}\mathbf{x}_{k-1}$ and covariance matrix $\mathbf{\Psi}$. Thus, the expected value of the per-stage returns depends explicitly on the value of the asset log-prices at the beginning of the period, and is therefore itself a random variable with respect to the information available at the start of the investment horizon. It is precisely this stochastic nature of the return distribution parameters that necessitates the use of Monte Carlo methods to determine the sequence of optimal portfolios in Chapter 5.

Here, a relaxation of Problem P_0 is considered in which the true, state-dependent return statistics are replaced by a set of deterministic proxies. By invoking the Principle of Certainty Equivalence [13], the stochastic means are approximated by their time t_0 conditional expectations. Letting $\mathbf{x}_0$ denote the initial log-prices, the return vector over the first stage is computed as:

$$\mathbf{r}_1 = \mathbf{x}_1 - \mathbf{x}_0 = (\mathbf{\Pi}_1 - \mathbb{I})\,\mathbf{x}_0 + \boldsymbol{\phi} + \boldsymbol{\epsilon}_1 = \mathbf{\Pi}\mathbf{x}_0 + \boldsymbol{\phi} + \boldsymbol{\epsilon}_1.$$

Under the assumption that $\boldsymbol{\epsilon}_1$ is Normally distributed with zero-mean and covariance matrix $\mathbf{\Psi}$, the return $\mathbf{r}_1$ is also multivariate Gaussian with mean $\mathbf{\Pi}\mathbf{x}_0 + \boldsymbol{\phi}$ and covariance $\mathbf{\Psi}$. In a similar manner, the stage k return is computed as:

$$\mathbf{r}_k = \mathbf{x}_k - \mathbf{x}_{k-1} = (\mathbf{\Pi}_1 - \mathbb{I})\,\mathbf{x}_{k-1} + \boldsymbol{\epsilon}_k = \mathbf{\Pi}\mathbf{x}_{k-1} + \boldsymbol{\phi} + \boldsymbol{\epsilon}_k.$$

In order the remove the dependence of $\mathbf{r}_k$ on any time t_{k-1} information, the log-prices $\mathbf{x}_{k-1}$ are represented by their equivalent moving-average form, as follows:

$$\mathbf{x}_{k-1} = \mathbf{\Pi}_1^{k-1}\mathbf{x}_0 + \sum_{i=1}^{k-1} \mathbf{\Pi}_1^{k-1-i}\left(\boldsymbol{\phi} + \boldsymbol{\epsilon}_i\right).$$

The asset returns over stage k is now given by:

$$\mathbf{r}_k = \mathbf{\Pi}\mathbf{\Pi}_1^{k-1}\mathbf{x}_0 + \mathbf{\Pi}\sum_{i=1}^{k-1}\mathbf{\Pi}_1^{k-1-i}\left(\boldsymbol{\phi}+\boldsymbol{\epsilon}_i\right) + \boldsymbol{\phi} + \boldsymbol{\epsilon}_k.$$

Therefore, conditioned on the information available at time t_0, $\mathbf{r}_k$ is Gaussian with mean:

$$\boldsymbol{\mu}_k = \mathbf{\Pi}\mathbf{\Pi}_1^{k-1}\mathbf{x}_0 + \mathbf{\Pi}\sum_{i=1}^{k-1}\mathbf{\Pi}_1^{k-1-i}\boldsymbol{\phi} + \boldsymbol{\phi}, \tag{6.2}$$

and covariance matrix:

$$\mathbf{\Sigma}_k = \text{var}\left[\mathbf{\Pi}\sum_{i=1}^{k-1}\mathbf{\Pi}_1^{k-1-i}\boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_k\right] = \sum_{i=1}^{k-1}\mathbf{\Pi}\mathbf{\Pi}_1^{k-1-i}\mathbf{\Psi}\left(\mathbf{\Pi}_1^{k-1-i}\right)^T\mathbf{\Pi}^T + \mathbf{\Psi}. \tag{6.3}$$

While the principle of certainty equivalence is useful in order to remove the stochastic nature of the return statistics, the downside is that the return statistics are not updated as new information becomes available over the lifetime of the trading strategy. In addition, due to the cointegration assumption, the matrix $\mathbf{\Pi}_1$ has at least one eigenvalue at unity, and therefore the covariance matrix $\mathbf{\Sigma}_k$ diverges at the number of stages increases, as stated in Thm. 4.3. Hence, the estimate of the per-stage asset return means becomes less reliable as time progresses.

6.1.2 Optimal Solution for Separable Embedding Certainty Equivalence Scheme

The per-stage asset return distributions are now characterized by a set of deterministic moments, given by Eqs. 6.2 and 6.3. Again, the Principle of Separable Embedding (Thm. 5.1) can be applied so that Problem P_0 can be solved by first computing the optimal sequence of portfolio policies for Problem P_1, according to:

$$\left\{\mathbf{w}_0^*, \mathbf{w}_1^*, \ldots, \mathbf{w}_{N-1}^*\right\} = \underset{\mathbf{w}_0,\mathbf{w}_1,\ldots,\mathbf{w}_{N-1}}{\arg\max} \quad E_{t_0}\left[\gamma_N \sum_{k=0}^{N-1}\mathbf{w}_k^T\mathbf{r}_{k+1} - \lambda_N\left(\sum_{k=0}^{N-1}\mathbf{w}_k^T\mathbf{r}_{k+1}\right)\right] \quad \Bigg\} \; P_1.$$

The resulting quadratic utility problem is solved by applying the dynamic programming algorithm, as derived next.

To begin the derivation, the value function one stage from the end, at time t_{N-1}, is given

by:

$$\begin{aligned}
J^*_{N-1}(r_{N-1}) &= \max_{\mathbf{w}_{N-1}} E_{t_{N-1}}\left[J_N(r_N)\right] = \max_{\mathbf{w}_{N-1}} E_{t_{N-1}}\left[\gamma_N r_N - \lambda_N r_N^2\right], \\
&= \max_{\mathbf{w}_{N-1}} E_{t_{N-1}}\left[\gamma_N\left(r_{N-1} + \mathbf{w}_{N-1}^T\mathbf{r}_N\right) - \lambda_N\left(r_{N-1} + \mathbf{w}_{N-1}^T\mathbf{r}_N\right)^2\right], \\
&= \max_{\mathbf{w}_{N-1}} E_{t_{N-1}}\left[\gamma_N\left(r_{N-1} + \mathbf{w}_{N-1}^T\mathbf{r}_N\right) - \lambda_N\left(r_{N-1}^2 + 2r_{N-1}\mathbf{w}_{N-1}^T\mathbf{r}_N + \mathbf{w}_{N-1}^T\mathbf{r}_N\mathbf{r}_N^T\mathbf{w}_{N-1}\right)\right], \\
&= \max_{\mathbf{w}_{N-1}} \gamma_N\left(r_{N-1} + \mathbf{w}_{N-1}^T\boldsymbol{\mu}_N\right) - \lambda_N\left(r_{N-1}^2 + 2r_{N-1}\mathbf{w}_{N-1}^T\boldsymbol{\mu}_N + \mathbf{w}_{N-1}^T\left(\boldsymbol{\Sigma}_N + \boldsymbol{\mu}_N\boldsymbol{\mu}_N^T\right)\mathbf{w}_{N-1}\right)
\end{aligned} \tag{6.4}$$

where $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$ are non-random and are defined by Eqs. 6.2 and 6.3, respectively. Differentiating with respect to $\mathbf{w}_{N-1}$ and setting the result equal to zero yields the following portfolio weight vector for the last stage:

$$\mathbf{w}^*_{N-1} = \frac{1}{2\lambda_N}\left(\gamma_N - 2\lambda_N r_{N-1}\right)\left(\boldsymbol{\Sigma}_N + \boldsymbol{\mu}_N\boldsymbol{\mu}_N^T\right)^{-1}\boldsymbol{\mu}_N. \tag{6.5}$$

While Eq. 6.5 does not depend on the current log-prices of the system, it does depend on the cumulative portfolio return, r_{N-1}, and therefore $\mathbf{w}^*_{N-1}$ is therefore *state-dependent*. In this sense, the strategy is still dynamic, taking advantage of some of the new information that becomes available over time.

The expression for $\mathbf{w}^*_{N-1}$ can now be plugged back into the objective function given by Eq. 6.4, and after some algebra, it is possible to show that $J^*_{N-1}(r_{N-1})$ is also a quadratic function of r_{N-1}, as follows:

$$J^*_{N-1}(r_{N-1}) = \gamma_{N-1}r_{N-1} - \lambda_{N-1}r_{N-1}^2 + c_{N-1},$$

where:

$$\begin{aligned}
\lambda_{N-1} &= \lambda_N\left(1 - \boldsymbol{\mu}_N^T\left(\boldsymbol{\Sigma}_N + \boldsymbol{\mu}_N\boldsymbol{\mu}_N^T\right)^{-1}\boldsymbol{\mu}_N\right) \\
\gamma_{N-1} &= \gamma_N\left(1 - \boldsymbol{\mu}_N^T\left(\boldsymbol{\Sigma}_N + \boldsymbol{\mu}_N\boldsymbol{\mu}_N^T\right)^{-1}\boldsymbol{\mu}_N\right) \\
c_{N-1} &= 0.25\frac{\gamma_N}{\lambda_N}\boldsymbol{\mu}_N^T\left(\boldsymbol{\Sigma}_N + \boldsymbol{\mu}_N\boldsymbol{\mu}_N^T\right)^{-1}\boldsymbol{\mu}_N.
\end{aligned}$$

In contrast to Eqs. 5.8 through 5.9 in which λ_{N-1} and γ_{N-1} are functions of the asset log-prices $\mathbf{x}_{N-1}$, here the coefficients of the cost function are deterministic quantities that can be computed in closed-form.

The above procedure is repeated for each successive stage. At time t_k, the value function is given as:

$$\begin{aligned} J_k^*(r_k) &= \max_{\mathbf{w}_k} E_{t_k}\left[J_{k+1}(r_{k+1})\right] = \max_{\mathbf{w}_k} E_{t_k}\left[\gamma_{k+1} r_{k+1} - \lambda_{k+1} r_{k+1}^2\right] + \sum_{i=k}^{N-1} c_i, \\ &= \max_{\mathbf{w}_k} \gamma_{k+1}\left(r_k + \mathbf{w}_k^T \boldsymbol{\mu}_{k+1}\right) - \lambda_{k+1}\left(r_k^2 + 2 r_k \mathbf{w}_k^T \boldsymbol{\mu}_{k+1} + \mathbf{w}_k^T\left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right)\mathbf{w}_k\right) \\ &\quad + \sum_{i=k}^{N-1} c_i. \end{aligned} \tag{6.6}$$

Differentiating with respect to $\mathbf{w}_k$ and setting the result equal to zero yields the following portfolio weight vector for stage k:

$$\mathbf{w}_k^* = \frac{1}{2\lambda_{k+1}}\left(\gamma_{k+1} - 2\lambda_{k+1} r_k\right)\left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right)^{-1}\boldsymbol{\mu}_{k+1}. \tag{6.7}$$

The scalars λ_k, γ_k, and c_k evolve according to:

$$\lambda_k = \lambda_{k+1}\left(1 - \boldsymbol{\mu}_{k+1}^T\left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right)^{-1}\boldsymbol{\mu}_{k+1}\right) \tag{6.8}$$

$$\gamma_k = \gamma_{k+1}\left(1 - \boldsymbol{\mu}_{k+1}^T\left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right)^{-1}\boldsymbol{\mu}_{k+1}\right) \tag{6.9}$$

$$c_k = 0.25\frac{\gamma_{k+1}}{\lambda_{k+1}}\boldsymbol{\mu}_{k+1}^T\left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right)^{-1}\boldsymbol{\mu}_{k+1} \tag{6.10}$$

Thus, as shown here, by replacing the state-dependent per-stage return distribution parameters with their time t_0 expectations, a closed-form solution for the optimal sequence of portfolio weight vectors can be determined using the separable embedding technique of [36]. While the resulting solution does not depend on the new asset log-prices observed over time, the portfolio policies do depend on the cumulative realized return, and therefore the overall strategy is still dynamic.

6.1.3 Budget Constraint

The addition of a budget constraint of the form $\mathbf{w}^T\mathbf{1} = 1$ to the certainty-equivalence approximation scheme is now considered. As discussed in Section 2.2.3, the constraint is

enforced by representing the portfolio weight vector by $\mathbf{w}_k = \mathbf{c}\,(1 + r_k) + \mathbf{D}\mathbf{v}_k$, where:

$$\mathbf{c} = \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}^T, \quad \mathbf{D} = \begin{pmatrix} \mathbb{I}_{p-1} \\ -\mathbf{1}^T \end{pmatrix}.$$

The expression $\mathbf{c}\,(1 + r_k)$ represents the investor's *current wealth* budget, which varies stage to stage. Equation 6.6 is now given by:

$$\begin{aligned} J_k^*(r_k) &= \max_{\mathbf{v}_k} \gamma_{k+1} r_k + \gamma_{k+1} \left(\mathbf{c}(1 + r_k) + \mathbf{D}\mathbf{v}_k\right)^T \boldsymbol{\mu}_{k+1} \\ &- \lambda_{k+1} r_k^2 - 2\lambda_{k+1} r_k \left(\mathbf{c}(1 + r_k) + \mathbf{D}\mathbf{v}_k\right)^T \boldsymbol{\mu}_{k+1} \\ &- \lambda_{k+1} \left(\mathbf{c}(1 + r_k) + \mathbf{D}\mathbf{v}_k\right)^T \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \left(\mathbf{c}(1 + r_k) + \mathbf{D}\mathbf{v}_k\right) + \sum_{i=k}^{N-1} c_i \end{aligned} \tag{6.11}$$

Differentiating Eq. 6.11 with respect to $\mathbf{v}_k$ and setting the result equal to zero, yields the following portfolio weight vector for time t_k:

$$\begin{aligned} \mathbf{v}_k^* &= \frac{1}{2\lambda_{k+1}} \left(\mathbf{D}^T \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \mathbf{D}\right)^{-1} \mathbf{D}^T \cdot \\ &\qquad \left(\left(\gamma_{k+1} - 2\lambda_{k+1} r_k\right) \boldsymbol{\mu}_{k+1} - 2\lambda_{k+1}(1 + r_k) \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \mathbf{c}\right), \\ &= \mathbf{f}_k + \mathbf{g}_k r_k, \end{aligned}$$

where:

$$\begin{aligned} \mathbf{f}_k &= \frac{1}{2\lambda_{k+1}} \left(\mathbf{D}^T \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \mathbf{D}\right)^{-1} \mathbf{D}^T \left(\gamma_{k+1}\boldsymbol{\mu}_{k+1} - 2\lambda_{k+1} \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \mathbf{c}\right), \\ \mathbf{g}_k &= -\left(\mathbf{D}^T \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \mathbf{D}\right)^{-1} \mathbf{D}^T \left(\boldsymbol{\mu}_{k+1} + \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \mathbf{c}\right). \end{aligned}$$

The scalars λ_k, γ_k, and c_k now evolve according to:

$$\begin{aligned} \lambda_k &= \lambda_{k+1} \left(1 + \left(\mathbf{c} + \mathbf{D}\mathbf{g}_k\right)^T \left(2\boldsymbol{\mu}_{k+1} + \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \left(\mathbf{c} + \mathbf{D}\mathbf{g}_k\right)\right)\right) \\ \gamma_k &= \gamma_{k+1} \left(1 + \left(\mathbf{c} + \mathbf{D}\mathbf{g}_k\right)^T \boldsymbol{\mu}_{k+1}\right) - 2\lambda_{k+1} \left(\mathbf{c} + \mathbf{D}\mathbf{f}_k\right)^T \left(\boldsymbol{\mu}_{k+1} + \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \left(\mathbf{c} + \mathbf{D}\mathbf{g}_k\right)\right) \\ c_k &= \gamma_{k+1} \left(\mathbf{c} + \mathbf{D}\mathbf{f}_k\right)^T \boldsymbol{\mu}_{k+1} - \lambda_{k+1} \left(\mathbf{c} + \mathbf{D}\mathbf{f}_k\right)^T \left(\boldsymbol{\Sigma}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T\right) \left(\mathbf{c} + \mathbf{D}\mathbf{f}_k\right) + \sum_{i=k}^{N-1} c_i \end{aligned}$$

Again, the scalars λ_k, γ_k, and c_k are non-random, as is the case in the work of [36]. An in depth look at the characteristics and performance of the certainty equivalence separable

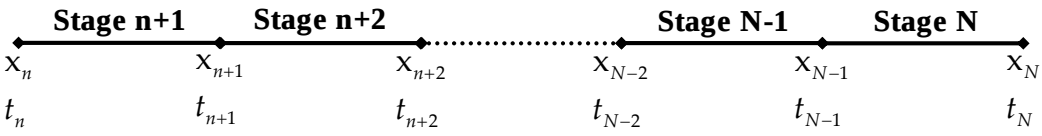


Figure 6-1. Timeline for Sequential Rescaling Algorithm.

embedding technique presented here, both with and without a budget constraint, is given in Section 6.5.

6.2 The Sequential Rescaling Algorithm

A second relaxation of problem P_0, which utilizes the backwards induction algorithm of dynamic programming in a novel way, is given here. Given the optimal solution to an N-stage problem, it is assumed that the solution to an $(N+1)$-stage problem is found by modifying only the magnitude (i.e., degree of net leverage), not the direction (i.e., relative asset proportions) of the existing portfolio weight vectors. The optimal portfolio policies are shown to be linear functions of the asset log-prices, which are modified by a constant scale factor as new stages are added to the problem.

6.2.1 Problem Formulation

Consider the N-stage dynamic mean-variance portfolio selection task defined by Problem P_0 of Section 6.1.1. The dynamic programming algorithm, often used to solve sequential decision making problems, cannot be directly applied due to the fact that the objective function is not additive in the number of stages due to the presence of non-zero covariance terms. However, a modified backwards recursive algorithm may still be used, as described here. Assume that at time t_{n+1}, the optimal sequence of portfolio policies $\pi^*_{n+1} = \{\mathbf{w}^*_{n+1}, ..., \mathbf{w}^*_{N-1}\}$ is known exactly, corresponding to a problem with $N-n+1$ stages and a risk budget of σ_0^2, as depicted in Figure 6-1. Furthermore, each policy is known to be a linear function of the asset log-prices (i.e., $\mathbf{w}_k = \mathbf{W}_k \mathbf{x}_k$ for all $k = n+1, ..., N-1$). Given π^*_{n+1}, the optimal portfolio policy at time t_n, $\mathbf{w}^*_n$, is found by solving Problem A_2, defined as follows:

$$\left.\begin{array}{c} \left\{\mathbf{w}^*_n, a^*_{n+1}, .., a^*_{N-1}\right\} = \underset{\mathbf{w}_n, a_{n+1}, ..., a_{N-1}}{\arg\max} E\left[\sum_{k=n+1}^{N} r_k\right] \\ \text{s.t. var}\left[\sum_{k=n+1}^{N} r_k\right] = \sigma_0^2 \end{array}\right\} A_2.$$

where

$$\begin{aligned} r_k &= \mathbf{w}_{k-1}^T \left(\mathbf{x}_k - \mathbf{x}_{k-1}\right), \\ \mathbf{w}_k &= a_k \mathbf{W}_k \mathbf{x}_k. \end{aligned}$$

Thus, in determining π_n^*, the augmented set of mean-variance optimal policies for an $N-n$ stage problem with risk budget σ_0^2, the assumption is imposed that each portfolio from π_{n+1}^* maintains the same *direction*, and only undergoes a *proportional scaling* by the constant a_k. For notational convenience, let $\mathbf{z}_k = \mathbf{x}_k^T \mathbf{W}_k^T \left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)$ for $k = n+1 \ldots N-1$, and define the vectors $\mathbf{z}$ and $\mathbf{a}$ as:

$$\mathbf{z} = \begin{pmatrix} \mathbf{x}_{n+1} - \mathbf{x}_n \\ \mathbf{z}_{n+1} \\ \vdots \\ \mathbf{z}_{N-1} \end{pmatrix}, \qquad \mathbf{a} = \begin{pmatrix} \mathbf{w}_n \\ a_{n+1} \\ \vdots \\ a_{N-1} \end{pmatrix}. \tag{6.12}$$

The resulting linear-quadratic program, denoted as Problem $A_2^{'}$, is given by:

$$\mathbf{a}^*, \lambda_L^* = \arg\max_{\mathbf{a}, \lambda_L} \mathbf{a}^T \boldsymbol{\mu}_{\mathbf{z}} - \lambda_L \left(\mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{z}} \mathbf{a} - \sigma_0^2\right) \Big\} A_2^{'},$$

with well-known solution:

$$\mathbf{a}^* = \begin{pmatrix} \mathbf{w}_n^* \\ a_{n+1}^* \\ \vdots \\ a_{N-1}^* \end{pmatrix} = \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}_{\mathbf{z}}^T \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\mu}_{\mathbf{z}}}} \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\mu}_{\mathbf{z}}. \tag{6.13}$$

The required moments of $\mathbf{z}$, $\boldsymbol{\mu}_{\mathbf{z}} = E\left[\mathbf{z}\right]$ and $\boldsymbol{\Sigma}_{\mathbf{z}} = \mathrm{var}\left[\mathbf{z}\right]$, are computed in Section 6.2.2.

Problem A_2 is solved for each successive stage until all N portfolio policies have been computed, denoted by the set π_0^*. As stated in Theorem 6.1, this backwards recursive approach, referred to as *the sequential rescaling algorithm*, produces a sequence of portfolio weight vectors that are all linear functions of the state.

Theorem 6.1.

The optimal sequence of portfolio policies $\pi_0^* = \{\mathbf{w}_0^*, ..., \mathbf{w}_{N-1}^*\}$*, computed according to the*

sequential rescaling algorithm of Problem A_2, are each linear functions of the asset log-prices, $\mathbf{x}_k$, and have the form $\mathbf{w}_k = \mathbf{W}_k \mathbf{x}_k$.

The proof of Theorem 6.1 is given in Appendix 6.A.

The sequential rescaling algorithm can be motivated from two perspectives. First, the set π^*_{n+1} corresponds to a solution that uses *all* the allowed variance budget, σ_0^2. Thus, when a new stage is added at each iteration of the backwards recursion, the existing solutions must be rescaled so that the risk can be optimally repartitioned across all the resulting stages. In this way, the variance constraint is viewed as a shared resource that must be allocated across the stages according to some scheduling scheme. Second, while a change to the portfolio size can be achieved using a single scale factor, a modification to a p-dimensional portfolio weight vector direction consists of multiplication by a rotation matrix with $p-1$ parameters. In addition, the rotation matrices introduce a source of non-linearity into the problem and make the resulting optimization problem intractable.

6.2.2 Optimal Solution for Sequential Rescaling Scheme

Analytic expressions for $\boldsymbol{\mu}_{\mathbf{z}}$ and $\boldsymbol{\Sigma}_{\mathbf{z}}$ are derived here. All expectation operators are conditioned on time t_n information. While the vector $\mathbf{z}$ is not Gaussian, only its first and second order statistics are required in order to solve Problem A_2.

First, consider $\boldsymbol{\mu}_{\mathbf{z}} = E[\mathbf{z}]$. The expected value for the first entry is equal to:

$$E[\mathbf{x}_{n+1} - \mathbf{x}_n] = E[\boldsymbol{\Pi}\mathbf{x}_n + \boldsymbol{\epsilon}_{n+1}] = \boldsymbol{\Pi}\mathbf{x}_n,$$

which is deterministic with respect to the information available at time t_n. In order to compute the remaining terms of $\boldsymbol{\mu}_{\mathbf{z}}$, given by $E[\mathbf{z}_k]$, observe that it is possible to equivalently represent $\mathbf{x}_k$ as:

$$\mathbf{x}_k = \boldsymbol{\Pi}_1^{k-n}\mathbf{x}_n + \sum_{i=n+1}^{k} \boldsymbol{\Pi}_1^{k-i}\boldsymbol{\epsilon}_i.$$

Accordingly, the quantity $E[\mathbf{z}_k]$ for $k = n+1,\ ...,\ N-1$ is computed as follows:

$$\begin{aligned} E[\mathbf{z}_k] &= E\left[\mathbf{x}_k^T \mathbf{W}_k^T \left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right] = E\left[\mathbf{x}_k^T \mathbf{W}_k^T \left(\mathbf{\Pi}\mathbf{x}_k + \boldsymbol{\epsilon}_{k+1}\right)\right] = E\left[\mathbf{x}_k^T \mathbf{W}_k^T \mathbf{\Pi}\mathbf{x}_k\right] \\ &= \mathbf{m}_k^T \mathbf{W}_k^T \mathbf{\Pi}\mathbf{m}_k + \text{trace}\left[\mathbf{W}_k^T \mathbf{\Pi}\mathbf{\Sigma}_k\right], \end{aligned} \tag{6.14}$$

where:

$$\mathbf{m}_k = \mathbf{\Pi}_1^{k-n}\mathbf{x}_n \tag{6.15}$$

$$\mathbf{\Sigma}_k = \sum_{i=n+1}^{k} \mathbf{\Pi}_1^{k-i} \mathbf{\Psi} \left(\mathbf{\Pi}_1^T\right)^{k-i}. \tag{6.16}$$

Second, consider the covariance matrix of z, $\mathbf{\Sigma}_\mathbf{z}$. The first entry is computed as:

$$\text{var}\left[\mathbf{x}_{n+1} - \mathbf{x}_n\right] = \text{var}\left[\mathbf{\Pi}\mathbf{x}_n + \boldsymbol{\epsilon}_{n+1}\right] = \mathbf{\Psi}.$$

In order to compute all of the remaining terms on the diagonal, the law of total variance is invoked, as follows:

$$\text{var}[\mathbf{z}_k] = \text{var}\left[E\left[\mathbf{z}_k \middle| \mathbf{x}_k\right]\right] + E\left[\text{var}\left[\mathbf{z}_k \middle| \mathbf{x}_k\right]\right] = \text{var}\left[\mathbf{x}_k^T \mathbf{W}_k^T \mathbf{\Pi}\mathbf{x}_k\right] + E\left[\mathbf{x}_k^T \mathbf{\Pi}^T \mathbf{\Psi}\mathbf{\Pi}\mathbf{x}_k\right].$$

Letting $\mathbf{A} = \mathbf{W}_k^T \mathbf{\Pi}$ and $\mathbf{B} = \mathbf{\Pi}^T \mathbf{\Psi}\mathbf{\Pi}$, the above expression can be rewritten as:

$$\text{var}[\mathbf{z}_k] = 4\mathbf{m}_k^T \left(\frac{\mathbf{A} + \mathbf{A}^T}{2}\right) \mathbf{\Sigma}_k \left(\frac{\mathbf{A} + \mathbf{A}^T}{2}\right) \mathbf{m}_k + 2\text{trace}\left[\mathbf{A}\mathbf{\Sigma}_k \mathbf{A}\mathbf{\Sigma}_k\right] + \mathbf{m}_k^T \mathbf{B}\mathbf{m}_k + \text{trace}\left[\mathbf{B}\mathbf{\Sigma}_k\right], \tag{6.17}$$

where $\mathbf{m}_k$ and $\mathbf{\Sigma}_k$ are defined by Eqs. 6.15 and 6.16 above.

Next, consider the off-diagonal terms of the covariance matrix. The entries of the first row and column, which correspond to the covariance of $\mathbf{x}_{n+1} - \mathbf{x}_n$ with $\mathbf{z}_k$ for $k = n+1,\ ...,\ N-1$,

are given by:

$$\begin{aligned}
\mathrm{cov}[\mathbf{z}_n, \mathbf{z}_k] &= \mathrm{cov}\left[(\mathbf{x}_{n+1} - \mathbf{x}_n), (\mathbf{W}_k \mathbf{x}_k)^T (\mathbf{x}_{k+1} - \mathbf{x}_k)\right], \\
&= E\left[(\mathbf{x}_{n+1} - \mathbf{x}_n)(\mathbf{W}_k \mathbf{x}_k)^T (\mathbf{x}_{k+1} - \mathbf{x}_k)\right] - E\left[\mathbf{x}_{n+1} - \mathbf{x}_n\right] E\left[(\mathbf{W}_k \mathbf{x}_k)^T (\mathbf{x}_{k+1} - \mathbf{x}_k)\right], \\
&= E\left[\boldsymbol{\epsilon}_{n+1} \mathbf{x}_k^T \mathbf{W}_k^T (\mathbf{x}_{k+1} - \mathbf{x}_k)\right], \\
&= E\left[\boldsymbol{\epsilon}_{n+1} \left(\boldsymbol{\Pi}_1^k \mathbf{x}_n + \sum_{i=n+1}^{k} \boldsymbol{\Pi}_1^{k-i} \boldsymbol{\epsilon}_i\right)^T \mathbf{W}_k^T \left(\boldsymbol{\Pi}\left(\boldsymbol{\Pi}_1^k \mathbf{x}_n + \sum_{i=n+1}^{k} \boldsymbol{\Pi}_1^{k-i} \boldsymbol{\epsilon}_i\right) + \boldsymbol{\epsilon}_{k+1}\right)\right], \\
&= \boldsymbol{\Psi}\left(\boldsymbol{\Pi}_1^{k-n+1}\right)^T \left(\boldsymbol{\Pi}^T \mathbf{W}_k + \mathbf{W}_k^T \boldsymbol{\Pi}\right) \boldsymbol{\Pi}_1^{k-n} \mathbf{x}_n. \qquad (6.18)
\end{aligned}$$

The remainder of the off-diagonal terms in $\boldsymbol{\Sigma}_{\mathbf{z}}$, $\mathrm{cov}[\mathbf{z}_i, \mathbf{z}_j]$, are computed in Appendix 6.B.

Given $\boldsymbol{\mu}_{\mathbf{z}}$ and $\boldsymbol{\Sigma}_{\mathbf{z}}$, the solution to Problem A_2 is given by Eq. 6.13. In addition, as discussed in the Proof of Theorem 6.1 in Appendix 6.A, the optimal portfolio policy at time t_n is equal to the following linear function of the asset log-prices:

$$\mathbf{w}_n^* = \left(\frac{1}{2\lambda_L} \boldsymbol{\Psi}^{-1} \boldsymbol{\Pi} - \sum_{k=n+1}^{N-1} a_k^* \left(\boldsymbol{\Pi}_1^{k-1}\right)^T \left(\boldsymbol{\Pi}^T \mathbf{W}_k + \mathbf{W}_k^T \boldsymbol{\Pi}\right) \boldsymbol{\Pi}_1^{k-n}\right) \mathbf{x}_n = \mathbf{W}_n \mathbf{x}_n = \mathbf{W}_n \mathbf{x}_n \qquad (6.19)$$

with $\lambda_L = \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}_{\mathbf{z}}^T \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\mu}_{\mathbf{z}}}}$. Examination of Eq. 6.19 reveals that the optimal policy can be divided into two distinct components. The first term, $\boldsymbol{\Psi}^{-1} \boldsymbol{\Pi} \mathbf{x}_n$, is proportional to the optimal solution for a single stage problem beginning at time t_n, and thus can be thought of as the "myopic" component. In this light, the second term of Eq. 6.19 can be viewed as a correction factor or "inter-temporal hedging" component that modifies the myopic solution to account for the uncertainty in the new log-price information that becomes available in the future. The hedging component explicitly depends on both the values the $\{a_k^*\}$ and $\{\mathbf{W}_k\}$ factors, and is directly due to the non-zero correlation between the stage n and all future portfolio returns, as defined by Eq. 6.18. A detailed performance analysis of the sequential rescaling dynamic approximation scheme is given in Section 6.5.

6.3 Linear Portfolio Parametrization

In the sequential rescaling scheme of Section 6.2, a sequence of portfolio policies with closed-form solutions is computed using a backwards recursion algorithm based on the methodology

of dynamic programming. As additional stages are considered, it is assumed that the existing portfolio policies are modified only by a constant scale factor. The resulting portfolios are linear functions of the asset log-prices, which naturally raises the question of how these policies relate to the *optimal linear* sequence of portfolio policies. Here, a linear parametric form is imposed on the portfolio weight vectors, and the dynamic mean-variance problem defined in Problem P_0 is solved directly, without the need for dynamic programming or other iterative methods.

6.3.1 Problem Formulation

Consider the N-stage dynamic mean-variance portfolio choice framework defined by Problem P_0 of Section 6.1.1. Furthermore, assume that each portfolio weight vector is a linear function of the log-prices, as follows:

$$\begin{aligned} \mathbf{w}_0 &= \mathbf{W}_0\mathbf{x}_0 \\ &\vdots \\ \mathbf{w}_{N-1} &= \mathbf{W}_{N-1}\mathbf{x}_{N-1}. \end{aligned}$$

Hence, Problem P_0 can be reformulated as Problem A_3, according to:

$$\left.\begin{array}{l} \left\{\mathbf{W}_0^*, ..., \mathbf{W}_{N-1}^*\right\} = \underset{\mathbf{W}_0,...,\mathbf{W}_{N-1}}{\arg\max}\ E\left[\sum_{k=0}^{N-1}\mathbf{x}_k^T\mathbf{W}_k^T\left(\mathbf{x}_{k+1}-\mathbf{x}_k\right)\right] \\ \qquad\qquad \text{s.t. } \operatorname{var}\left[\sum_{k=0}^{N-1}\mathbf{x}_k^T\mathbf{W}_k^T\left(\mathbf{x}_{k+1}-\mathbf{x}_k\right)\right] = \sigma_0^2 \end{array}\right\}\ A_3.$$

For notational convenience, define the stacked vector $\mathbf{z}_0$ and matrix $\mathbf{W}$, as follows:

$$\mathbf{z}_0 = \begin{pmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_{N-1} \end{pmatrix}, \qquad \mathbf{W} = \begin{pmatrix} \mathbf{W}_0 & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{W}_{N-1} \end{pmatrix}.$$

Problem A_3 can be equivalently expressed by problem $A_3^{'}$, as:

$$\left.\left\{\mathbf{W}^*, \lambda_L\right\} = \underset{\mathbf{W},\lambda_L}{\arg\max}\ E_{t_0}\left[\mathbf{z}_0^T\mathbf{W}^T\left(\mathbf{z}_1-\mathbf{z}_0\right)\right] - \lambda_L\left\{\operatorname{var}_{t_0}\left[\mathbf{z}_0^T\mathbf{W}^T\left(\mathbf{z}_1-\mathbf{z}_0\right)\right] - \sigma_0^2\right\}\right\}\ A_3^{'},$$

with Lagrange multiplier λ_L and where:

$$\mathbf{z}_1 = \mathbf{A}\mathbf{z}_0 + \boldsymbol{\epsilon}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbb{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbb{I} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Pi}_1 \end{pmatrix}, \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \boldsymbol{\epsilon}_{N+1} \end{pmatrix}.$$

All of the required moments of the objective function of Problem $A_3^{'}$ are computed next.

6.3.2 Optimal Solution for Linear Scheme

Before a solution to Problem $A_3^{'}$ is computed, expressions for the mean and variance of $\mathbf{z}_0^T \mathbf{W}^T (\mathbf{z}_1 - \mathbf{z}_0)$ are required. The first moment is computed using the law of iterated expectations, as follows:

$$\begin{aligned} E_{t_0}\left[\mathbf{z}_0^T \mathbf{W}^T (\mathbf{z}_1 - \mathbf{z}_0)\right] &= E_{\mathbf{z}_0}\left[E_{\mathbf{z}_1}\left[\mathbf{z}_0^T \mathbf{W}^T (\mathbf{z}_1 - \mathbf{z}_0)\middle|\, \mathbf{z}_0\right]\right] = E_{\mathbf{z}_0}\left[\mathbf{z}_0^T \mathbf{W}^T (\mathbf{A} - \mathbb{I})\, \mathbf{z}_0\right] \\ &= \text{trace}\left[\mathbf{W}^T (\mathbf{A} - \mathbb{I}) \left(\mathbf{\Sigma}_{\mathbf{z}} + \mathbf{m}_{\mathbf{z}} \mathbf{m}_{\mathbf{z}}^T\right)\right], \end{aligned}$$

where $\mathbf{m}_{\mathbf{z}}$ and $\mathbf{\Sigma}_{\mathbf{z}}$ are defined by:

$$\begin{aligned} \mathbf{m}_{\mathbf{z}} &= E[\mathbf{z}_0] = \begin{pmatrix} \mathbf{x}_0 & \mathbf{\Pi}_1 \mathbf{x}_0 & \cdots & \mathbf{\Pi}_1^{N-1} \mathbf{x}_0 \end{pmatrix}^T \\ \mathbf{\Sigma}_{\mathbf{z}} &= \text{var}[\mathbf{z}_0] = E\left[\mathbf{z}_0 \mathbf{z}_0^T\right] - \mathbf{m}_{\mathbf{z}} \mathbf{m}_{\mathbf{z}}^T \\ &= \begin{pmatrix} E\left[\mathbf{x}_0 \mathbf{x}_0^T\right] & E\left[\mathbf{x}_0 \mathbf{x}_1^T\right] & \cdots & E\left[\mathbf{x}_0 \mathbf{x}_{N-1}^T\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[\mathbf{x}_{N-1} \mathbf{x}_0^T\right] & E\left[\mathbf{x}_{N-1} \mathbf{x}_1^T\right] & \cdots & E\left[\mathbf{x}_{N-1} \mathbf{x}_{N-1}^T\right] \end{pmatrix} - \mathbf{m}_{\mathbf{z}} \mathbf{m}_{\mathbf{z}}^T \\ E\left[\mathbf{x}_i \mathbf{x}_j^T\right] &= E\left[\mathbf{x}_j \mathbf{x}_i^T\right] = \mathbf{m}_i^T \mathbf{\Pi}_1^{(j-i)T} \mathbf{m}_i + \text{trace}\left[\mathbf{\Pi}_1^{(j-i)T} \mathbf{\Sigma}_i\right] (\text{ for } i < j) \\ \mathbf{m}_i &= \mathbf{\Pi}_1^i \mathbf{x}_0 \\ \mathbf{\Sigma}_i &= \sum_{j=0}^{i-1} \mathbf{\Pi}_1^j \mathbf{\Psi} \left(\mathbf{\Pi}_1^T\right)^j. \end{aligned}$$

Next, consider the variance of $\mathbf{z}_0^T \mathbf{W}^T (\mathbf{z}_1 - \mathbf{z}_0)$, which is computing using the law of total variance, as follows:

$$\begin{aligned} \text{var}[\mathbf{z}_0^T \mathbf{W}^T (\mathbf{z}_1 - \mathbf{z}_0)] &= E_{\mathbf{z}_0}\left[\text{var}_{\mathbf{z}_1}[\mathbf{z}_0^T \mathbf{W}^T (\mathbf{z}_1 - \mathbf{z}_0)\middle|\, \mathbf{z}_0]\right] + \text{var}_{\mathbf{z}_0}\left[E_{\mathbf{z}_1}[\mathbf{z}_0^T \mathbf{W}^T (\mathbf{z}_1 - \mathbf{z}_0)\middle|\, \mathbf{z}_0]\right] \\ &= E_{\mathbf{z}_0}\left[\mathbf{z}_0^T \mathbf{W}^T \mathbf{\Sigma}_{\boldsymbol{\epsilon}} \mathbf{W} \mathbf{z}_0\right] + \text{var}_{\mathbf{z}_0}\left[\mathbf{z}_0^T \mathbf{W}^T (\mathbf{A} - \mathbb{I})\, \mathbf{z}_0\right] \\ &= \text{trace}\left[\mathbf{W}^T \mathbf{\Sigma}_{\boldsymbol{\epsilon}} \mathbf{W} \left(\mathbf{\Sigma}_{\mathbf{z}} + \mathbf{m}_{\mathbf{z}} \mathbf{m}_{\mathbf{z}}^T\right) + \mathbf{B} \mathbf{\Sigma}_{\mathbf{z}} \mathbf{B} \left(0.5 \mathbf{\Sigma}_{\mathbf{z}} + \mathbf{m}_{\mathbf{z}} \mathbf{m}_{\mathbf{z}}^T\right)\right] \end{aligned}$$

where:

$$\mathbf{B} = \mathbf{W}^T (\mathbf{A} - \mathbb{I}) + (\mathbf{A} - \mathbb{I})^T \mathbf{W}.$$

The first-order optimality condition is computed by taking the derivative of Problem $A_3^{'}$ with respect to $\mathbf{W}$, and setting the result equal to zero, yielding:

$$\begin{aligned} 0 \quad = \quad & (\mathbf{A} - \mathbb{I}) (\mathbf{\Sigma_z} + \mathbf{m_z}\mathbf{m_z}^T) \\ & -2\lambda_L \left\{ \mathbf{\Sigma}_{\boldsymbol{\epsilon}} \mathbf{W} \left(\mathbf{\Sigma_z} + \mathbf{m_z}\mathbf{m_z}^T \right) \right. \\ & + (\mathbf{A} - \mathbb{I}) \left(0.5\mathbf{\Sigma_z} + \mathbf{m_z}\mathbf{m_z}^T \right) \mathbf{W}^T (\mathbf{A} - \mathbb{I})^T \mathbf{\Sigma_z} + (\mathbf{A} - \mathbb{I}) \mathbf{\Sigma_z} \mathbf{W}^T (\mathbf{A} - \mathbb{I}) \left(0.5\mathbf{\Sigma_z} + \mathbf{m_z}\mathbf{m_z}^T \right) \\ & \left. + (\mathbf{A} - \mathbb{I}) \mathbf{\Sigma_z} (\mathbf{A} - \mathbb{I})^T \mathbf{W} \left(0.5\mathbf{\Sigma_z} + \mathbf{m_z}\mathbf{m_z}^T \right) + (\mathbf{A} - \mathbb{I}) \left(0.5\mathbf{\Sigma_z} + \mathbf{m_z}\mathbf{m_z}^T \right) (\mathbf{A} - \mathbb{I})^T \mathbf{W} \mathbf{\Sigma_z} \right\} \end{aligned}$$

Unfortunately, this condition cannot be used to derive a closed-form expression for $\mathbf{W}$, and numerical search methods must be utilized instead. The performance of this relaxation scheme is considered in Section 6.5.

6.4 Semi-Myopic

In the semi-myopic approach to the dynamic mean-variance portfolio choice problem, the $N-$stage problem is solved as N consecutive single-stage problems. The direction of the per-stage portfolio weight vectors are determined by solving the corresponding single-stage MVO problem, with no consideration given to past or future stages. Once these directions are computed, the amount of per-stage leverage is determined so that the total expected return is mean-variance efficient.

6.4.1 Problem Formulation

Consider the N-stage dynamic mean-variance portfolio selection problem given by Problem P_0 of Section 6.1.1. Furthermore, assume that the portfolio weight vector is given by a scaled version of the corresponding single-stage MVO solution starting at time t_k, as follows:

$$\begin{aligned} \mathbf{w}_0 \quad &= \quad a_0 \mathbf{\Psi}^{-1} \mathbf{\Pi} \mathbf{x}_0 \\ & \vdots \\ \mathbf{w}_{N-1} \quad &= \quad a_{N-1} \mathbf{\Psi}^{-1} \mathbf{\Pi} \mathbf{x}_{N-1}. \end{aligned}$$

Like the sequential rescaling and optimal linear approaches of Sections 6.2 and 6.3, respectively, the semi-myopic portfolio weight vectors are linear functions of the log-prices. The name semi-myopic refers to the fact that the portfolio directions are determined according to the single-stage MVO solution, but the vectors are jointly scaled so that the total solution is mean-variance efficient. Problem P_0 can be reformulated as problem A_4, according to:

$$\left.\begin{array}{r} \left\{a_0^*, ..., a_{N-1}^*\right\} = \underset{a_0,...,a_{N-1}}{\arg\max}\ E\left[\sum_{k=0}^{N-1} a_k \mathbf{x}_k^T \mathbf{\Pi}^T \mathbf{\Psi}^{-1}\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right] \\ \text{s.t. var}\left[\sum_{k=0}^{N-1} a_k \mathbf{x}_k^T \mathbf{\Pi}^T \mathbf{\Psi}^{-1}\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)\right] = \sigma_0^2 \end{array}\right\}\ A_4.$$

Now define the stacked vectors $\mathbf{s} \in \mathbb{R}^{pN\times 1}$, with k^{th} element $\mathbf{x}_k^T \mathbf{\Pi}^T \mathbf{\Psi}^{-1}\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)$, and $\mathbf{a} \in \mathbb{R}^{N\times 1}$, with k^{th} element a_k. Futhermore, let $\boldsymbol{\mu}_{\mathbf{s}}$ denote the mean vector of $\mathbf{s}$ and let $\mathbf{\Sigma}_{\mathbf{s}}$ denote the corresponding covariance matrix. Subsequentially, Problem A_4 can be equivalently expressed by problem $A_4^{'}$, given by:

$$\left.\begin{array}{r} \left\{\mathbf{a}^*, \lambda_L\right\} = \underset{\mathbf{a},\lambda_L}{\arg\max}\ \mathbf{a}^T \boldsymbol{\mu}_{\mathbf{s}} - \lambda_L\left\{\mathbf{a}^T \mathbf{\Sigma}_{\mathbf{s}} \mathbf{a} - \sigma_0^2\right\} \end{array}\right\}\ A_4^{'},$$

with Lagrange multiplier λ_L. The required moments are given next.

6.4.2 Optimal Solution for Semi-Myopic Scheme

Problem $A_4^{'}$ is a linear-quadratic program with well-known solution, given by:

$$\mathbf{a}^* = \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}_{\mathbf{s}}^T \mathbf{\Sigma}_{\mathbf{s}}^{-1} \boldsymbol{\mu}_{\mathbf{s}}}} \mathbf{\Sigma}_{\mathbf{s}}^{-1} \boldsymbol{\mu}_{\mathbf{s}}.$$

Letting $\mathbf{W}_k = \mathbf{\Psi}^{-1}\mathbf{\Pi}$, observe that the definition for the vector $\mathbf{s}$ used in Problem $A_4^{'}$ is equivalent to the definition for $\mathbf{z}$ used in the sequential rescaling algorithm, as defined by Eq. 6.12. Therefore, the mean vector, $\boldsymbol{\mu}_{\mathbf{s}}$, is computed according to Eq. 6.14 and the covariance matrix, $\mathbf{\Sigma}_{\mathbf{s}}$, is computed according to Eq. 6.17 and Appendix 6.B. The performance of all the approximate dynamic schemes presented in this chapter are considered next.

6.5 Comparison of Approximation Strategies

The four approximation strategies developed in this chapter are now contrasted in order to determine the regimes in which each strategy performs best. For completeness, the

strategies are also compared to the static portfolios of Chapter 4 and the dynamic MVO solution of Chapter 5. All of the analysis presented here is based on synthetic data.

6.5.1 Portfolio Return Statistics

First, consider the per-stage portfolio return statistics for each relaxation scheme presented in this chapter, in the context of a *two-stage portfolio choice* problem. Like the static and dynamic MVO solutions, all four approximate dynamic strategies produce first stage portfolio returns that are Normally distributed, with means $\mathbf{w}_0^T \mathbf{\Pi} \mathbf{x}_0$ and variances $\mathbf{w}_0^T \mathbf{\Psi} \mathbf{w}_0$. On the other hand, the distribution for the second stage portfolio return depends on the exact functional form of $\mathbf{w}_1$, which is itself a random variable.

In the certainty equivalence scheme of Section 6.1, the portfolio weight vector over the second stage may be expressed as an affine function of the log-prices, as follows:

$$\begin{aligned}
\mathbf{w}_1 &= \frac{1}{2\lambda_2} \left(\gamma_2 - 2\lambda_2 r_1\right) \left(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T\right)^{-1} \boldsymbol{\mu}_2 \\
&= \frac{1}{2\lambda_2} \left(\gamma_2 - 2\lambda_2 \mathbf{w}_0^T \left(\mathbf{x}_1 - \mathbf{x}_0\right)\right) \left(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T\right)^{-1} \boldsymbol{\mu}_2 \\
&= \frac{\gamma_2 \left(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T\right)^{-1} \boldsymbol{\mu}_2}{2\lambda_2} - \left(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T\right)^{-1} \boldsymbol{\mu}_2 \mathbf{w}_0^T \left(\mathbf{x}_1 - \mathbf{x}_0\right) \\
&= \mathbf{f} + \mathbf{L} \mathbf{x}_1,
\end{aligned}$$

where $\mathbf{f} = \frac{\gamma_2 \left(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T\right)^{-1} \boldsymbol{\mu}_2}{2\lambda_2} + \left(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T\right)^{-1} \boldsymbol{\mu}_2 \mathbf{w}_0^T \mathbf{x}_0$ and $\mathbf{L} = -\left(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T\right)^{-1} \boldsymbol{\mu}_2 \mathbf{w}_0^T$. The corresponding return over the second stage is equal to:

$$r_2 = \mathbf{w}_1^T \left(\mathbf{x}_2 - \mathbf{x}_1\right) = \mathbf{f}^T \left(\mathbf{x}_2 - \mathbf{x}_1\right) + \mathbf{x}_1^T \mathbf{L}^T \left(\mathbf{x}_2 - \mathbf{x}_1\right), \tag{6.20}$$

which is a Gaussian quadratic form with no known distribution [1].

In the remaining three approximation strategies, the portfolio weight vectors over the second stage are all linear functions of the log-prices (i.e., $\mathbf{w}_1 = \mathbf{L} \mathbf{x}_1$). Thus, the corresponding second-stage returns are also Gaussian quadratic forms with no known distributions. The

[1] Only certain Gaussian quadratic forms have known distributions. For example, if $\mathbf{z} \sim N\left(\mathbf{0}, \sigma^2 \mathbb{I}\right)$ and the matrix $\mathbf{A}$ is idempotent ($\mathbf{A}^2 = \mathbf{A}$), then the quadratic form $\frac{\mathbf{z}^T \mathbf{A} \mathbf{z}}{\sigma^2}$ follows a chi-squared distribution. The quadratic form given in Eq. 6.20 is a sum of random variables with Gaussian, chi-squared, and Gaussian-product distributions.

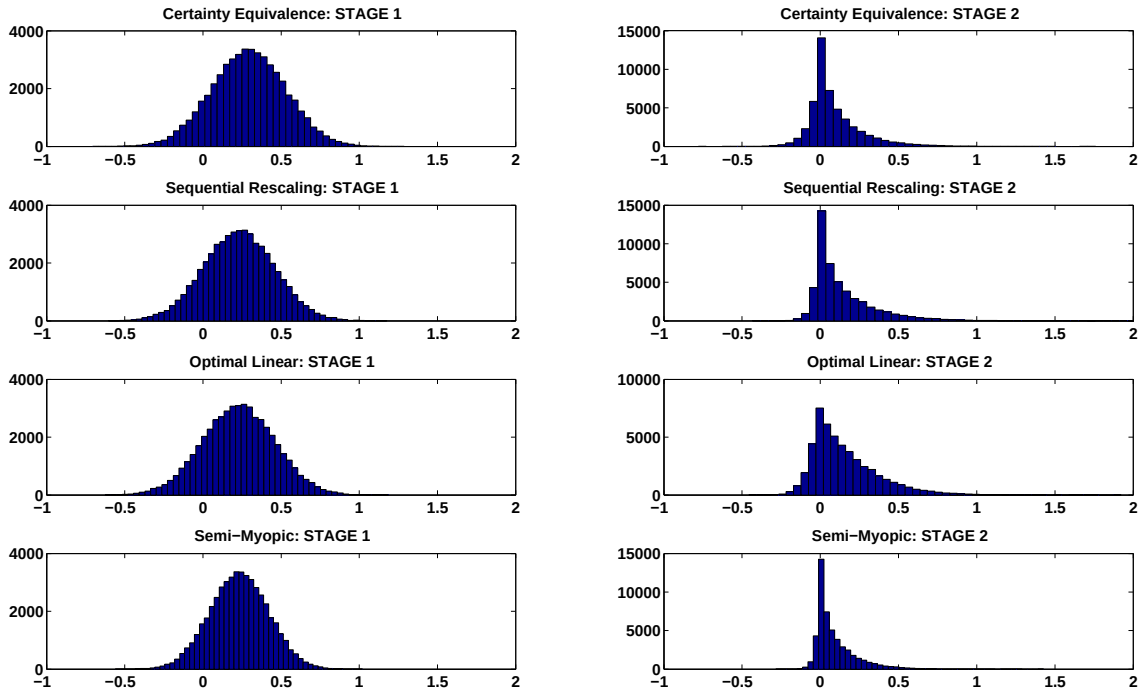


Figure 6-2. Per-stage portfolio return histograms for approximate dynamic strategies.

distributions corresponding to the first and second stages of a two-stage example are characterized experimentally in Ex. 6.1.

Example 6.1.

Consider again the system of two risky assets presented in Example 5.1. Using $M = 5 \times 10^4$ test sample paths, the per-stage returns were computed using each of the four dynamic approximation schemes presented in this chapter. The corresponding histograms are depicted in Fig. 6-2. As predicted, the returns over the first stage appear to be Normally distributed, while the returns over the second stage have no known distribution, but possess higher kurtosis (peakedness) and skewness (asymmetry) than the Gaussian distribution. ■

Second, consider the relationship among the per-stage returns. Throughout this thesis, the performance of each investment strategy is measured using the cumulative portfolio log-return, which accumulates in an additive manner. This enables the variance of the total return to be readily computed as the sum of the per-stage variances and inter-stage covariances. As demonstrated by Example 6.2 below, the strategies that achieve the greatest levels of negative inter-stage portfolio return correlation can in turn assume the highest

degree of per-stage risk, thus earning the highest expected total return.

Example 6.2.

Table 6.1 displays the per-stage return statistics for all seven of the strategies considered in this thesis under the synthetic system presented in Ex. 5.1, with a total portfolio return risk of $\sigma_0 = 0.20$. Here, the overall performance of each scheme is measured by the Sharpe ratio (ratio of the total expected return to standard deviation), as defined by Eq. 2.8. This statistic measures the quantity of reward (expected profit) earned per unit of risk (standard deviation). As the table reveals, the dynamic MVO solution outperforms all other strategies, and all four approximate dynamic solutions outperform both static schemes.

Recall from examples 4.2 and 5.5 that there was a direct correspondence between the total expected portfolio return and the magnitude of the inter-stage return correlation. One might surmise that correlation alone explains the success of each strategy; however, a counterexample to this hypothesis is shown in Ex. 4.4, as the static MVO portfolio vector did not correspond to the direction with minimum correlation (i.e., maximum negative correlation). The simulation results presented here confirm that both *correlation and leverage* determines the relative success of each scheme. For example, while the dynamic semi-myopic scheme yields a slightly higher Sharpe ratio than the static MVO strategy (1.62 vs. 1.59), the semi-myopic scheme uses a significantly smaller correlation coefficient (-0.26 vs. -0.40). On the other hand, while the static strategy uses a much larger net leverage of 139% of initial wealth in each stage, the semi-myopic scheme utilizes a net leverage of 258% of initial wealth in the first stage, followed by 73% of time t_1 wealth in the second stage.

It is also interesting to note that all five dynamic schemes allocate more risk (standard deviation) to the first stage as compared to the second. In all but the optimal linear approach, this corresponds to a higher degree of first stage net leverage. This implies that net leverage is only part of the story, and both individual asset and total portfolio leverage levels must be examined in order to understand the behavior of each strategy. ■

6.5.2 Asymptotic Properties

Recall from the discussion of the static MVO solution of Chapter 4 that the time horizon of the investor matters. For short investment horizons, the investor can earn additional profit by constructing a portfolio with components both in the $\boldsymbol{\beta}$ direction, which produces a portfolio value process that is AWSS, and in the direction of expected change, which produces a portfolio value process that is integrated of order one. However, as the investment length increases and the variance along the direction of expected change grows, the variance constraint placed on the terminal portfolio return limits the amount allocated to the integrated component. Ultimately, in the limit of an infinite number of stages, all of the wealth is allocated to a portfolio purely in the $\boldsymbol{\beta}$ direction (assuming the no drift, i.e., $\boldsymbol{\phi} = \mathbf{0}$).

In the dynamic MVO solution of Chapter 5, the direction of the optimal portfolio weight vector does not converge to the $\boldsymbol{\beta}$ direction. Due to the ability of an investor to rebalance the portfolio at the beginning of each stage, there is always some incentive to capitalize on the short-term predictive power of the cointegrated VAR process. This logic should carry over to the four approximate dynamic schemes, as explored in Example 6.3.

Example 6.3.

For the system of two risky assets in Example 5.1, the direction of the portfolio weight vector

Strategy	Stage 1, r_1					Stage 2, r_2					corr	Total, r_T		
	Mean	Std	Weights			Mean	Std	Weights				Mean	Std	Sharpe Ratio
			w_1	w_2	Net Lev			w_1	w_2	Net Lev	$[r_1, r_2]$			
Dynamic: MVO	0.30	0.26	-0.88	2.94	2.06	0.14	0.23	-0.15	1.54	1.39	-0.69	0.44	0.20	2.22
Dynamic: Linear	0.23	0.23	-0.92	2.00	1.09	0.16	0.21	-0.17	1.75	1.58	-0.60	0.39	0.20	1.92
Dynamic: CE	0.30	0.24	-0.32	3.25	2.93	0.09	0.17	-0.09	0.96	0.87	-0.56	0.38	0.20	1.90
Dynamic: SR	0.23	0.23	-0.87	2.07	1.20	0.14	0.20	-0.12	1.21	1.09	-0.57	0.37	0.20	1.85
Dynamic: Myopic	0.23	0.19	-0.26	2.58	2.32	0.09	0.13	-0.07	0.80	0.73	-0.26	0.32	0.20	1.62
Static: MVO	0.19	0.16	-0.47	1.86	1.39	0.13	0.20	-0.47	1.86	1.39	-0.40	0.32	0.20	1.59
Static: Beta	0.08	0.16	-0.81	0.27	-0.54	0.05	0.17	-0.81	0.27	-0.54	-0.29	0.13	0.20	0.66

Table 6.1. Second-order return statistics for two-stage example. The success of each trading scheme, measured in terms of the Sharpe ratio (ratio of expected return to standard deviation), depends both on the inter-stage correlation and per-stage leverage. The dynamic MVO solution outperforms all other strategies, and all four approximate dynamic solutions outperform both static schemes.

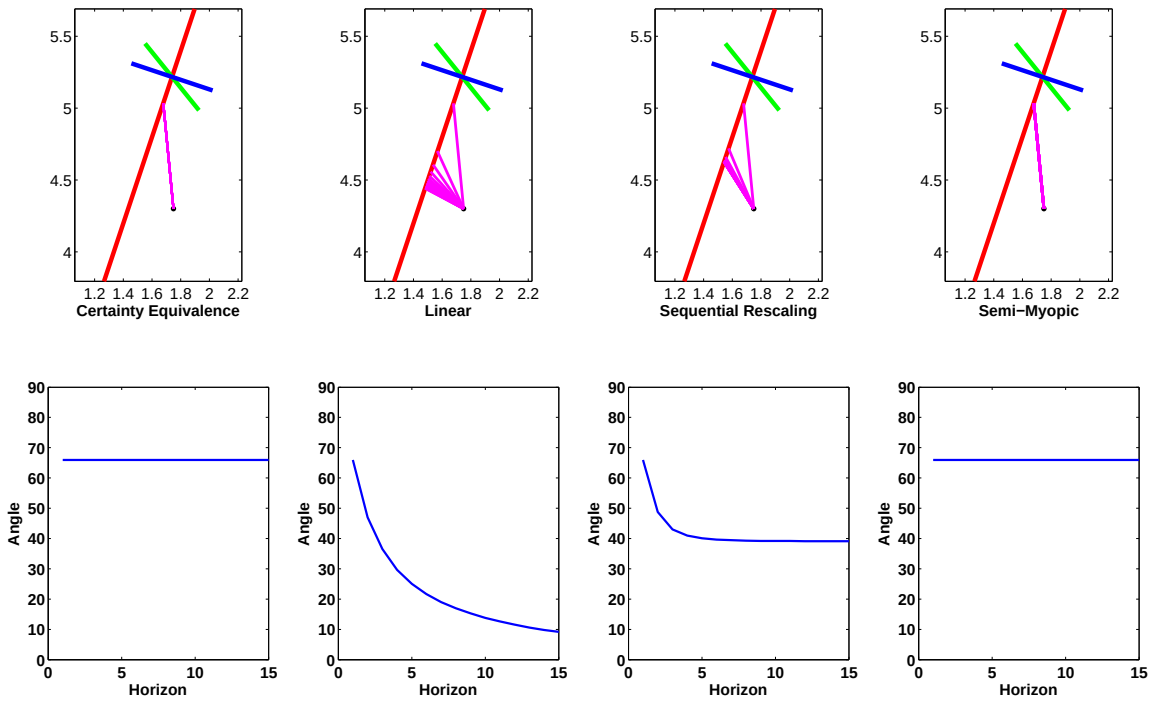
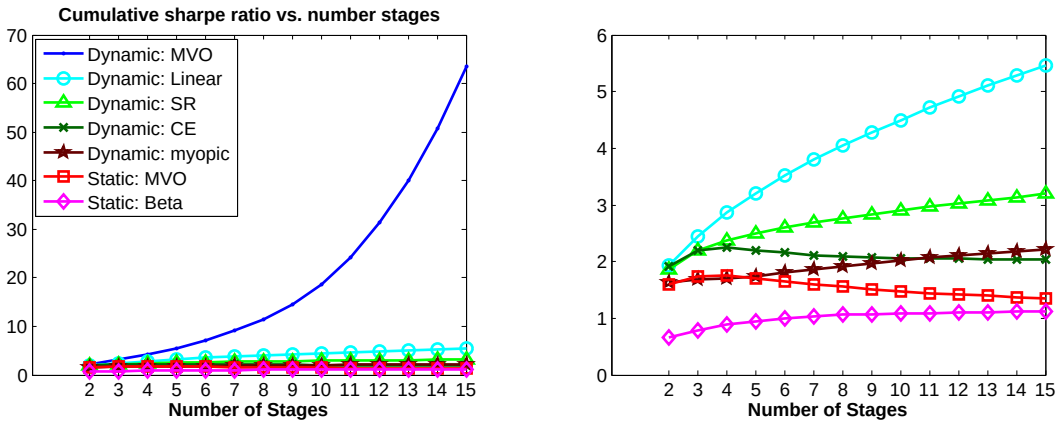


Figure 6-3. Direction of approximate dynamic portfolio weight vectors as a function of investment length.

over the first stage is plotted as the length of the investment horizon is varied over the set $N = \{2, \ldots, 15\}$, as depicted in Figure 6-3. For clarity, the angle between the portfolio weight vector and the direction of the Beta portfolio is also shown. As this series of figures illustrates, only in the case corresponding to the optimal linear approximation scheme does the direction of the portfolio weight vector converge towards the $\boldsymbol{\beta}$ direction as the number of stages increases. In the certainty equivalence and semi-myopic schemes, the direction of the portfolio weight vector is independent of the number of stages, and thus remains constant over time. Furthermore, due to the fact that their corresponding portfolio vectors are proportional, the two strategies admit equivalent angles. In the sequential-rescaling strategy, the vector converges to a direction approximately 40 degrees away from $\boldsymbol{\beta}$. ■

Thus, the preceding example demonstrates that all of the strategies do not exploit short-term predictability. In some cases, such as the optimal linear strategy, sacrificing immediate rewards in the early stages actually leads to a higher net expected return. However, it is important for the reader to remember that a closed-form solution does not exist for the portfolio weight vectors under the optimal linear strategy, and numerical search methods are utilized to determine the sequence of portfolio policies. Therefore, the results from Ex.



(a) Cumulative Sharpe ratio vs. investment time horizon.

(b) Cumulative Sharpe ratio vs. investment time horizon for all but dynamic MVO strategy.

Figure 6-4. Cumulative Sharpe ratio.

6.3 represent one example problem, and should not be assumed to hold in general.

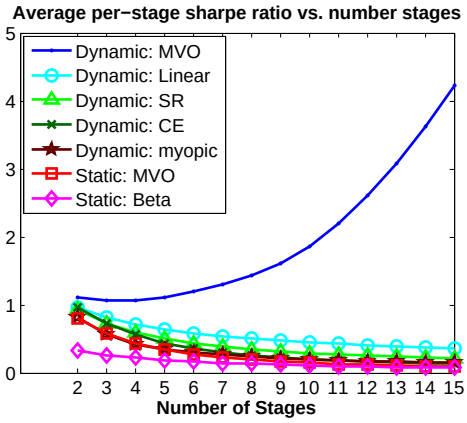
Having now established the relationship between the direction of the optimal portfolio weight vector and the investment time horizon, the relative performance of the four approximation schemes over time is now studied. Performance is again defined in terms of the Sharpe ratio, defined by Eq. 2.8.

Example 6.4.

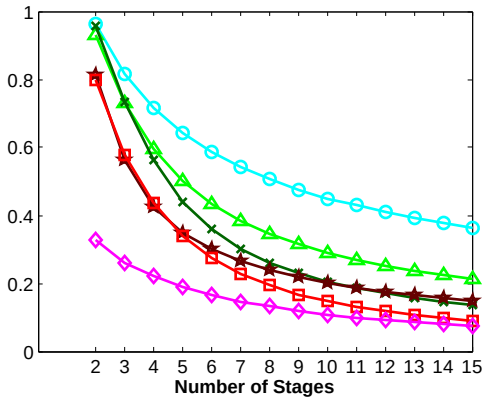
Consider once more the system of two risky assets from Example 5.1. The performance of each strategy is accessed using the overall and normalized Sharpe ratios, denoted by S and S_n, respectively, and defined as follows:

$$S = \frac{E[r_N]}{\sigma_N}, \qquad S_n = \frac{E[r_N]}{N\sigma_N} \tag{6.21}$$

For each strategy, S and S_n are computed over the set of investment horizon lengths $N = \{2, ..., 15\}$, as depicted in Figures 6-4 and 6-5. In each figure, the first plot depicts the Sharpe ratio curves for all seven strategies, while the second plot displays all but the dynamic MVO solution, so that the shape of each curve can be clearly discerned.



(a) Normalized Sharpe ratio vs. investment time horizon.



(b) Normalized Sharpe ratio vs. investment time horizon for all but dynamic MVO solution.

Figure 6-5. Normalized Sharpe ratio.

As Fig. 6-4 reveals, the dynamic MVO, optimal linear, sequential rescaling, and semi-myopic approaches all have cumulative Sharpe ratios that grow over time, while the remaining three converge to steady-state levels. The cumulative Sharpe ratio for the dynamic MVO solution appears to be growing at an exponential rate, dramatically out performing all other strategies. In both the static MVO and dynamic certainty equivalence solutions, the cumulative Sharpe ratio initially grows, before falling back and appearing to converg to a smaller steady-state. It is also interesting to note that the semi-myopic strategy is initially out performed by all but the Beta strategy. However, in the end, the semi-myopic scheme ranks fourth out of seven, eventually overtaking both the static MVO and dynamic certainty equivalence solutions. Finally, consistent with the results from Chapter 4, the Sharpe ratio for the static MVO solution approaches the Sharpe ratio for the Beta portfolio as time increases, as the direction of the static solutions converges to a vector in the span of the $\boldsymbol{\beta}$ subspace.

The normalized Sharpe ratio curves are generated by dividing the cumulative Sharpe ratios by their corresponding investment horizons. The result is that for all but the dynamic MVO solution, the Sharpe ratios decline over time before appearing to converge to steady-state solutions. Again, the curves for the sequential rescaling and semi-myopic schemes cross

around $N = 11$, and the curve for the static MVO solution approaches the curve for the static Beta scheme. ■

As examples 6.2 through 6.4 demonstrate, the performance of some of the strategies can be ranked on an absolute scale, while others have regimes under which they may outperform alternative strategies. First and foremost, the dynamic scheme developed in Chapter 5 is the mean-variance optimal strategy for a multistage cointegrated asset allocation problem. However, the algorithm to compute the optimal portfolio policy at each stage requires the use of Monte Carlo numerical approximation methods and therefore may not be preferred over alternative, computationally efficient trading schemes. Second, of the set of strategies characterized by linear portfolio policies (i.e., the sequential rescaling, the optimal linear, and the semi-myopic schemes), the optimal linear scheme of Section 6.3 is mean-variance optimal by construction. However, the algorithm to compute these portfolio weight vectors require the use of numerical search and therefore, like the true optimal dynamic solution, may not be preferred over alternative, computationally efficient methods. In addition, the sequential rescaling scheme always outperforms the semi-myopic scheme, as the portfolio weight vector for the former strategy contains both an explicit myopic component, and an inter-temporal hedging component that accounts for uncertainty in the future asset log-prices. However, the performance of the certainty equivalence scheme relative to the sequential rescaling and semi-myopic schemes changes over time. Initially, the certainty equivalence scheme may earn a higher Sharpe ratio than either alternative. However, due to the fact that the variance of the estimator used to characterize the per-stage asset returns grows linearly with the length of the investment horizon, the performance of the strategy degrades over time. In the limit as the number of stages increases, both sequential rescaling and semi-myopic asset allocation rules outperform the certainty equivalence scheme. Lastly, while the static MVO solution is mean-variance optimal by design and therefore must outperform the static Beta portfolio, it is not necessarily true that all of the dynamic schemes always outperform all of the static schemes. In some cases corresponding to a short investment length, the static MVO solution may earn a higher Sharpe ratio than the semi-myopic scheme.

6.A Proofs of Chapter 6 Theorems

Proof of Theorem 6.1.

Consider the following proof by induction.

Base Case: At time t_{N-1}, the investor faces a single-stage mean-variance portfolio choice problem. According to Eq. 4.2, for an investment length of one period, the optimal portfolio weight vector is given by:

$$\mathbf{w}_{N-1}^{*} = \frac{\sigma_0}{\sqrt{\mathbf{x}_{N-1}^T \mathbf{\Pi}^T \mathbf{\Psi}^{-1} \mathbf{\Pi} \mathbf{x}_{N-1}}} \mathbf{\Psi}^{-1} \mathbf{\Pi} \mathbf{x}_{N-1}. \tag{6.22}$$

The direction of the portfolio vector is determined by the quantity $\mathbf{\Psi}^{-1}\mathbf{\Pi}\mathbf{x}_{N-1}$, which is a linear function of the asset log-prices, $\mathbf{x}_{N-1}$.

Assume true for M stage problem: Assume $\pi_{N-M}^{*} = \{\mathbf{w}_{N-M}^{*}, ..., \mathbf{w}_{N-1}^{*}\}$ denotes the optimal sequence of portfolio policies for a dynamic mean-variance portfolio choice problem with M stages, where each portfolio weight vector is a linear function of the asset log-prices, i.e., $\mathbf{w}_k^{*} = \mathbf{W}_k \mathbf{x}$ for $k = N-M, \ldots, N-1$.

Show true for $M+1$ stage problem: In order to derive an explicit expression for $\mathbf{w}_{N-M-1}^{*}$, substitute $\mathbf{a} = \begin{pmatrix} \mathbf{w}_{N-M-1} & a_{N-M} & \cdots & a_{N-1} \end{pmatrix}^T$ into Problem A_2, and differentiate with respect to $\mathbf{w}_n$, as follows:

$$\mathbf{w}_{N-M-1}^{*} = \left(\frac{1}{2\lambda_L} \mathbf{\Psi}^{-1}\mathbf{\Pi} - \sum_{k=N-M}^{N-1} a_k \left(\mathbf{\Pi}_1^{k-1} \right)^T \left(\mathbf{\Pi}^T \mathbf{W}_k + \mathbf{W}_k^T \mathbf{\Pi} \right) \mathbf{\Pi}_1^{k-(N-M-1)} \right) \mathbf{x}_{N-M-1} \tag{6.23}$$

with $\lambda_L = \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}_{\mathbf{z}}^T \mathbf{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\mu}_{\mathbf{z}}}}$. Therefore, $\mathbf{w}_{N-M-1}$ is again a linear function of the state, $\mathbf{x}_n$.

6.B Derivation of terms in sequential rescaling algorithm

In the sequential rescaling approximation scheme, the off-diagonal terms in $\mathbf{\Sigma_z}$, denoted as $\text{cov}[\mathbf{z}_i, \mathbf{z}_j]$, are computed as follows:

$$\begin{aligned}
\text{cov}[\mathbf{z}_i, \mathbf{z}_j] &= E\left[\mathbf{z}_i\mathbf{z}_j\right] - E\left[\mathbf{z}_i\right] E\left[\mathbf{z}_j\right] \\
&= E\left[\left(\mathbf{x}_i^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{x}_i + \mathbf{x}_i^T\mathbf{W}_i^T\boldsymbol{\epsilon}_{i+1}\right)\left(\mathbf{x}_j^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{x}_j + \mathbf{x}_j^T\mathbf{W}_j^T\boldsymbol{\epsilon}_{j+1}\right)\right] \\
&\quad -E\left[\mathbf{x}_i^T\mathbf{W}_i^T\mathbf{\Pi}x_i + \mathbf{x}_i^T\mathbf{W}_i^T\boldsymbol{\epsilon}_{i+1}\right] E\left[\mathbf{x}_j^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{x}_j + \mathbf{x}_j^T\mathbf{W}_j^T\boldsymbol{\epsilon}_{j+1}\right] \\
&= E\left[\mathbf{x}_i^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{x}_i\mathbf{x}_j^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{x}_j + \mathbf{x}_i^T\mathbf{W}_i^T\boldsymbol{\epsilon}_{i+1}\mathbf{x}_j^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{x}_j\right. \\
&\quad \left.+\mathbf{x}_i^T\mathbf{W}_i^T\mathbf{\Pi}x_i\mathbf{x}_j^T\mathbf{W}_j^T\boldsymbol{\epsilon}_{j+1} + \mathbf{x}_i^T\mathbf{W}_i^T\boldsymbol{\epsilon}_{i+1}\mathbf{x}_j^T\mathbf{W}_j^T\boldsymbol{\epsilon}_{j+1}\right] \\
&\quad - \left(\mathbf{m}_i^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{m}_i + \text{trace}\left[\mathbf{W}_i^T\mathbf{\Pi}\mathbf{\Sigma}_i\right]\right)\left(\mathbf{m}_j^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{m}_j + \text{trace}\left[\mathbf{W}_j^T\mathbf{\Pi}\mathbf{\Sigma}_j\right]\right)
\end{aligned}$$

Assuming $i > j$ implies that $\mathbf{x}_i^T\mathbf{W}_i^T\boldsymbol{\epsilon}_{i+1}\mathbf{x}_j^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{x}_j$ and $\mathbf{x}_i^T\mathbf{W}_i^T\boldsymbol{\epsilon}_{i+1}\mathbf{x}_j^T\mathbf{W}_j^T\boldsymbol{\epsilon}_{j+1}$ are both equal to zero. The last remaining two terms are evaluated as:

$$\begin{aligned}
E\left[\mathbf{x}_i^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{x}_i\mathbf{x}_j^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{x}_j\right] &= E\left[\left(\mathbf{\Pi}_1^{i-n}\mathbf{x}_n + \sum_{\ell=n+1}^{i}\mathbf{\Pi}_1^{i-\ell}\boldsymbol{\epsilon}_\ell\right)\mathbf{W}_i^T\mathbf{\Pi}\left(\mathbf{\Pi}_1^{i-n}\mathbf{x}_n + \sum_{\ell=n+1}^{i}\mathbf{\Pi}_1^{i-\ell}\boldsymbol{\epsilon}_\ell\right)\cdot\right. \\
&\qquad \left.\left(\mathbf{\Pi}_1^{j-n}\mathbf{x}_n + \sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{j-\ell}\boldsymbol{\epsilon}_\ell\right)\mathbf{W}_j^T\mathbf{\Pi}\left(\mathbf{\Pi}_1^{j-n}\mathbf{x}_n + \sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{j-\ell}\boldsymbol{\epsilon}_\ell\right)\right] \\
&= \mathbf{x}_n^T\left(\mathbf{\Pi}_1^{i-n}\right)^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{\Pi}_1^{i-n}\mathbf{x}_n\mathbf{x}_n^T\left(\mathbf{\Pi}_1^{j-n}\right)^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{\Pi}_1^{j-n}\mathbf{x}_n \\
&+\mathbf{x}_n^T\left(\mathbf{\Pi}_1^{i-n}\right)^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{\Pi}_1^{i-n}\mathbf{x}_n\sum_{\ell=n+1}^{j}\text{trace}\left[\left(\mathbf{\Pi}_1^{j-\ell}\right)^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{\Pi}_1^{j-\ell}\mathbf{\Psi}\right] \\
&+\mathbf{x}_n^T\left(\mathbf{\Pi}_1^{i-n}\right)^T\left(\mathbf{W}_i^T\mathbf{\Pi} + \mathbf{\Pi}^T\mathbf{W}_i\right)\left(\sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{j-\ell}\mathbf{\Psi}\left(\mathbf{\Pi}_1^{j-\ell}\right)^T\right)\left(\mathbf{W}_j^T\mathbf{\Pi} + \mathbf{\Pi}^T\mathbf{W}_j\right)\mathbf{\Pi}_1^{j-n}\mathbf{x}_n \\
&+\sum_{\ell=n+1}^{i}\text{trace}\left[\left(\mathbf{\Pi}_1^{i-\ell}\right)^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{\Pi}_1^{i-\ell}\mathbf{\Psi}\right]x_n^T\left(\mathbf{\Pi}_1^{j-n}\right)^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{\Pi}_1^{j-n}\mathbf{x}_n \\
&+\sum_{\ell=i+1}^{j}\text{trace}\left[\left(\mathbf{\Pi}_1^{i-\ell}\right)^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{\Pi}_1^{i-\ell}\mathbf{\Psi}\right]\sum_{\ell=n+1}^{j}\text{trace}\left[\left(\mathbf{\Pi}_1^{j-\ell}\right)^T\mathbf{W}_j^T\mathbf{\Pi}\mathbf{\Pi}_1^{j-\ell}\mathbf{\Psi}\right] \\
&\quad +E\left[\left(\sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{i-\ell}\boldsymbol{\epsilon}_\ell\right)^T\mathbf{W}_i^T\mathbf{\Pi}\sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{i-\ell}\boldsymbol{\epsilon}_\ell\left(\sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{j-\ell}\boldsymbol{\epsilon}_\ell\right)^T\mathbf{W}_j^T\mathbf{\Pi}\sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{j-\ell}\boldsymbol{\epsilon}_\ell\right]
\end{aligned}$$

The quartic term above is also computable in closed-form, but is omitted here for brevity.

$$
\begin{aligned}
E\left[\mathbf{x}_i^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{x}_i\mathbf{x}_j^T\mathbf{W}_j^T\boldsymbol{\epsilon}_{j+1}\right] &= E\left[\left(\mathbf{\Pi}_1^{i-n}\mathbf{x}_n+\sum_{\ell=n+1}^{i}\mathbf{\Pi}_1^{i-\ell}\boldsymbol{\epsilon}_\ell\right)^T\mathbf{W}_i^T\mathbf{\Pi}\left(\mathbf{\Pi}_1^{i-n}\mathbf{x}_n+\sum_{\ell=n+1}^{i}\mathbf{\Pi}_1^{i-\ell}\boldsymbol{\epsilon}_\ell\right)\cdot\right.\\
&\qquad\left.\left(\mathbf{\Pi}_1^{j-n}\mathbf{x}_n+\sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{j-\ell}\boldsymbol{\epsilon}_\ell\right)^T\mathbf{W}_j^T\boldsymbol{\epsilon}_{j+1}\right]\\
&= \operatorname{trace}\left[\left(\mathbf{\Pi}_1^{i-j-1}\right)^T\mathbf{W}_i^T\mathbf{\Pi}\mathbf{\Pi}_1^{i-n}\mathbf{x}_n\mathbf{x}_n^T\left(\mathbf{\Pi}_1^{j-n}\right)^T\mathbf{W}_j^T\mathbf{\Psi}\right]\\
&\quad+\operatorname{trace}\left[\mathbf{W}_i^T\mathbf{\Pi}\mathbf{\Pi}_1^{i-j-1}\mathbf{\Psi}\mathbf{W}_j\left(\sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{j-\ell}\mathbf{\Psi}\left(\mathbf{\Pi}_1^{j-\ell}\right)^T\right)\right]\\
&\quad+\operatorname{trace}\left[\left(\mathbf{\Pi}_1^{i-j-1}\right)^T\mathbf{\Pi}^T\mathbf{W}_i\mathbf{\Pi}_1^{i-n}\mathbf{x}_n\mathbf{x}_n^T\left(\mathbf{\Pi}_1^{j-n}\right)^T\mathbf{W}_j^T\mathbf{\Psi}\right]\\
&\quad+\operatorname{trace}\left[\left(\sum_{\ell=n+1}^{j}\mathbf{\Pi}_1^{i-\ell}\mathbf{\Psi}\left(\mathbf{\Pi}_1^{j-\ell}\right)^T\right)\mathbf{W}_i^T\mathbf{\Pi}\mathbf{\Pi}_1^{i-j-1}\mathbf{\Psi}\mathbf{W}_j\right]
\end{aligned}
$$

Appendix A

Vector Autoregressive Models

A vector autoregressive (VAR) process is an example of a linear multiple-input, multiple-output (MIMO) system. This appendix explores the properties of VAR processes, providing the necessary framework for the study of cointegrated VAR systems in Chapter 3. In Section A.1, the canonical and state-space forms of a VAR model are presented. The transformation to modal coordinates is derived in Section A.2. Section A.3 details the total system response, which is comprised of the zero-input and zero-state responses. The maximum likelihood estimation procedure for the process parameters is given in Section A.4.

A.1 State Space Representation

Let $\mathbf{x}[n] \in \mathbb{R}^p$ denote a random vector, which evolves according to an order k discrete-time vector autoregressive (VAR(k)) process, defined by:

$$\mathbf{x}[n] = \mathbf{\Pi}_1\mathbf{x}[n-1] + \mathbf{\Pi}_2\mathbf{x}[n-2] + \cdots + \mathbf{\Pi}_k\mathbf{x}[n-k] + \mathbf{\Phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n], \tag{A.1}$$

where:

- $\mathbf{\Pi}_i \in \mathbb{R}^{p\times p}$ are matrices of coupling coefficients,
- $\mathbf{d} \in \mathbb{R}^{r\times 1}$ is a vector of deterministic inputs,
- $\mathbf{\Phi} \in \mathbb{R}^{p\times r}$ matrix of coefficients relating the deterministic terms to the elements of $\mathbf{x}$,
- $\boldsymbol{\epsilon} \in \mathbb{R}^{p\times 1}$ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{\Psi}$.

The canonical difference equation form explicitly highlights the dependence of the current state vector on the k previous states. Note, however, that the model does not include lags of the inputs, known as moving average terms. More general vector autoregressive, moving-average (VARMA) models do exist [27], but are not considered here.

The system given in Eq. A.1 can be rewritten in state-space form by augmenting the state vector, as:

$$\underbrace{\begin{pmatrix} \mathbf{x}[n] \\ \mathbf{x}[n-1] \\ \vdots \\ \mathbf{x}[n-(k-1)] \end{pmatrix}}_{\mathbf{q}[n+1]} = \underbrace{\begin{pmatrix} \mathbf{\Pi}_1 & \cdots & \mathbf{\Pi}_{k-1} & \mathbf{\Pi}_k \\ \mathbb{I}_p & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbb{I}_p & \mathbf{0} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} \mathbf{x}[n-1] \\ \mathbf{x}[n-2] \\ \vdots \\ \mathbf{x}[n-k] \end{pmatrix}}_{\mathbf{q}[n]} + \underbrace{\begin{pmatrix} \mathbb{I}_p \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}}_{\mathbf{b}} \underbrace{\Big(\mathbf{\Phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n]\Big)}_{\mathbf{u}_n}$$

$$\underbrace{\mathbf{x}[n]}_{\mathbf{y}[n]} = \underbrace{\begin{pmatrix} \mathbf{\Pi}_1 & \cdots & \mathbf{\Pi}_{k-1} & \mathbf{\Pi}_k \end{pmatrix}}_{\mathbf{c}^T} \underbrace{\begin{pmatrix} \mathbf{x}[n-1] \\ \mathbf{x}[n-2] \\ \vdots \\ \mathbf{x}[n-k] \end{pmatrix}}_{\mathbf{q}_n} + \underbrace{\begin{pmatrix} \mathbb{I}_p \end{pmatrix}}_{\mathbf{d}[n]} \underbrace{\Big(\mathbf{\Phi}\mathbf{d}[n] + \boldsymbol{\epsilon}[n]\Big)}_{\mathbf{u}[n]}, \tag{A.2}$$

where $\mathbf{A} \in \mathbb{R}^{pk \times pk}$, $\mathbf{b} \in \mathbb{R}^{pk \times p}$, $\mathbf{c}^T \in \mathbb{R}^{p \times pk}$, $\mathbf{d} \in \mathbb{R}^{p \times p}$, $\mathbf{u} \in \mathbb{R}^{p \times p}$, $\mathbf{q} \in \mathbb{R}^{pk \times 1}$, and $\mathbb{I}_p$ is the $p \times p$ identity matrix. Beginning at time n_0, the output of the system is uniquely determined through knowledge of the sequence of inputs beginning with $\mathbf{u}[n_0]$, and the initial value of the state vector, $\mathbf{q}[n_0]$. Typically, the initial time is taken to be $n_0 = 0$.

A random process $x[n]$ is *wide-sense stationary* (WSS) if both the mean and the variance of the process remain constant over time, and the autocorrelation function, $E[x[n]x[m]]$, only depends on the time difference, $|n-m|$. A process that starts at time $n_0 = 0$ with arbitrarily initial conditions is *asymptotically WSS* (AWSS) if any initial condition response decays to zero as $n \to \infty$ *and* the process' statistics approach those of a WSS process [67]. A process that is WSS is also AWSS. It is possible to impose restrictions on the state transition matrix $\mathbf{A}$ so that the output process is AWSS, as stated in Theorem A.1 below.

Theorem A.1.

Let the vector random process denoted by $\mathbf{x}[n]$ *evolve according to the* $VAR(k)$ *model given in Eq. A.2, with no deterministic inputs (i.e.,* $\mathbf{d}[n] = 0$*) and WSS stochastic input* $\boldsymbol{\epsilon}[n]$*. Furthermore, the process begins at time* $n_0 = 0$ *with initial condition* $\mathbf{q}[0]$*. The output*

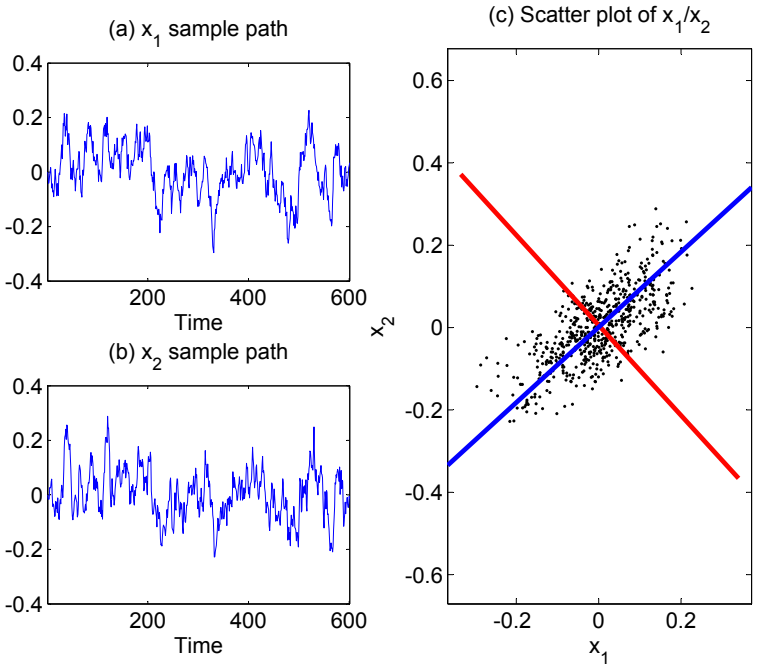


Figure A-1. Two views of an AWSS VAR process.

process is AWSS if the eigenvalues of the state transition matrix, $\mathbf{A}$*, satisfy* $|\lambda_i| < 1,\ \forall\ i$*.*

Proof. Since $|\lambda_i| < 1,\ \forall\ i$, the system is bounded-input, bounded-output (BIBO) stable. Putting a WSS process through a BIBO stable system produces an output process that is AWSS. □

When the stochastic input process is assumed to be i.i.d. with $\boldsymbol{\epsilon}[n] \sim N(\boldsymbol{\mu}, \boldsymbol{\Psi})$ and the deterministic input is given by $\mathbf{d}[n] = 0$, the asymptotic or steady-state distribution of $\mathbf{x}[n]$ is also Gaussian, with the mean vector $\boldsymbol{\mu}_\infty$ and covariance matrix $\boldsymbol{\Sigma}_\infty$ computed as follows:

$$\boldsymbol{\mu}_\infty = \mathbf{C}\left(\sum_{i=0}^{\infty} \mathbf{A}^i \widetilde{\boldsymbol{\mu}}\right) = \mathbf{0}, \quad \boldsymbol{\Sigma}_\infty = \mathbf{C}\left(\sum_{i=0}^{\infty} \mathbf{A}^i \widetilde{\boldsymbol{\Psi}} \left(\mathbf{A}^i\right)^T\right) \mathbf{C}^T,$$

where:

$$\widetilde{\boldsymbol{\mu}} = \begin{pmatrix} \boldsymbol{\mu} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}^T, \quad \widetilde{\boldsymbol{\Psi}} = \operatorname{diag}\begin{pmatrix} \boldsymbol{\Psi} & \mathbf{0}_p & \cdots & \mathbf{0}_p \end{pmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \mathbb{I}_p & \mathbf{0}_{p \times kp-p} \end{pmatrix}.$$

Convergence of the series is guaranteed by the assumption that all the eigenvalues of $\mathbf{A}$ are inside the unit circle. An example of an AWSS VAR process, along with a derivation of the corresponding asymptotic distribution, is presented next.

Example A.1.

Consider the following VAR(1) model, with no deterministic inputs:

$$\mathbf{x}[n] \;=\; \mathbf{\Pi}_1\mathbf{x}[n-1]+\boldsymbol{\epsilon}[n] = \begin{pmatrix} 0.9912 & -0.0684 \\ 0.2549 & 0.7088 \end{pmatrix}\mathbf{x}[n-1]+\boldsymbol{\epsilon}[n],$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{\Psi})$ and $\mathbf{\Psi} = 0.001\mathbb{I}_2$. The state transition matrix is given by $\mathbf{A} = \mathbf{\Pi}_1$, and here $\mathbf{C} = \mathbb{I}$. The covariance matrix of the limiting distribution of $\mathbf{x}[n]$ is found through an eigendecomposition of $\mathbf{A}$, as follows:

$$\mathbf{A} = \begin{pmatrix}\mathbf{m}_1 & \mathbf{m}_2\end{pmatrix}\begin{pmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{pmatrix}\begin{pmatrix}\mathbf{n}_1^T \\ \mathbf{n}_2^T\end{pmatrix} = \begin{pmatrix}0.3000 & 0.1684 \\ 0.4000 & 0.4708\end{pmatrix}\begin{pmatrix}0.9 & 0 \\ 0 & 0.8\end{pmatrix}\begin{pmatrix}6.3724 & -2.2793 \\ -5.4142 & 4.0606\end{pmatrix},$$

which results in:

$$\begin{aligned}\mathbf{\Sigma}_\infty \;&=\; \sum_{i=1}^{\infty}\mathbf{A}^i\widetilde{\mathbf{\Psi}}\left(\mathbf{A}^i\right)^T = 10^{-3}\sum_{i=1}^{\infty}\mathbf{M}\mathbf{D}^i\mathbf{M}^{-1}\left(\mathbf{M}^{-1}\right)^T\mathbf{D}^i\mathbf{M}^T, \\ &=\; 10^{-3}\mathbf{M}\left(\sum_{i=1}^{\infty}\mathbf{D}\mathbf{M}^{-1}\left(\mathbf{M}^{-1}\right)^T\mathbf{D}\right)\mathbf{M}^T = 10^{-3}\mathbf{M}\left(\sum_{i=1}^{\infty}\begin{pmatrix}\lambda_1^2\mathbf{n}_1^T\mathbf{n}_1 & \lambda_1\lambda_2\mathbf{n}_1^T\mathbf{n}_2 \\ \lambda_2\lambda_1\mathbf{n}_1^T\mathbf{n}_2 & \lambda_2^2\mathbf{n}_2^T\mathbf{n}_2\end{pmatrix}\right)\mathbf{M}^T, \\ &=\; \begin{pmatrix}0.0095 & 0.0064 \\ 0.0064 & 0.0079\end{pmatrix}.\end{aligned}$$

Figure A-1 contains two views of this process in the time-domain for a single sample path realization. On the left, the two underlying signals are plotted independently in time, while the right side contains a scatter plot of the two series, which clearly highlights their correlation structure. The eigenvectors or principal axes of $\mathbf{\Sigma}_\infty$ are also shown. ■

A.2 Coordinate Transformations

It is often easier to analyze the properties of a state-space system in an alternate coordinate system, obtained through a linear transformation of the state variables. For example, one may be interested to know how each component of the input process impacts the current state and output process, and whether or not the system is bounded-input, bounded-output (BIBO) stable. One such representation is known as *modal coordinates*, which decouples the system behavior along each eigenvalue or *modal frequency* of the system. This alternate form relies on the spectral decomposition of the matrix $\mathbf{A}$, discussed in Section A.2.1, which

is followed by a description of the coordinate transformation in Section A.2.2. Section A.2.3 presents a second decomposition known as Jordan Canonical Form, which is used when the matrix $\mathbf{A}$ is not diagonalizable.

A.2.1 Spectral Analysis

When the linear operator $\mathbf{A}$ is diagonalizable, it may be expressed as:

$$\mathbf{A} = \mathbf{M}\mathbf{D}\mathbf{M}^{-1}, \tag{A.3}$$

where $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathbf{A}$, and $\mathbf{M}$ is a matrix whose columns consist of the corresponding eigenvectors, which are all linearly independent, though not necessarily orthogonal. Linear independence of the eigenvectors is required in order to guarantee that the inverse matrix, $\mathbf{M}^{-1}$, exists. Diagonalization is possible when either:

- $\mathbf{A}$ has n distinct eigenvalues,
- $\mathbf{A}$ has repeated eigenvalues, but the geometric multiplicity of each (dimension of the associated eigenspace) equals the algebraic multiplicity (number of times the repeated eigenvalue appears).

In cases when the geometric multiplicity of an eigenvalue is less than the algebraic, the matrix $\mathbf{A}$ can be factored into Jordan Canonical Form, as described in Section A.2.3.

The eigenvalues of $\mathbf{A}$ can be computed by determining the roots of $\det(\mathbf{A} - \lambda\mathbb{I}_p) = \mathbf{0}$. Using a cofactor expansion of the last row of $\mathbf{A}$ from Eq. A.2, it is possible to show that:

$$\det(\mathbf{A} - \lambda\mathbb{I}_p) \quad = \quad \det\left(\mathbb{I}_p - \lambda^{-1}\mathbf{\Pi}_1 - \lambda^{-2}\mathbf{\Pi}_2 - \cdots - \lambda^{-k}\mathbf{\Pi}_k\right), \tag{A.4}$$

whereby the eigenvalues of $\mathbf{A}$ are the pk roots of Eq. A.4.

Once the matrix of eigenvalues $\mathbf{D}$ is determined, the corresponding eigenvector matrix $\mathbf{M}$ can be computed. Let λ_i denote the i^{th} eigenvalue and let $\mathbf{m}_i \in \mathbb{R}^{pk}$ denote the corresponding eigenvector, which satisfy $\mathbf{A}\mathbf{m}_i = \lambda_i\mathbf{m}_i$. Expressing $\mathbf{m}_i$ in block form to match

the block structure of the matrix $\mathbf{A}$, gives:

$$\begin{pmatrix} \mathbf{\Pi}_1 & \cdots & \mathbf{\Pi}_{k-1} & \mathbf{\Pi}_k \\ \mathbb{I}_p & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbb{I}_p & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{m}_{1,i} \\ \mathbf{m}_{2,i} \\ \vdots \\ \mathbf{m}_{k,i} \end{pmatrix} = \lambda_i \begin{pmatrix} \mathbf{m}_{1,i} \\ \mathbf{m}_{2,i} \\ \vdots \\ \mathbf{m}_{k,i} \end{pmatrix}.$$

Multiplying through results in the following system of matrix equations:

$$\begin{aligned} \mathbf{\Pi}_1\mathbf{m}_{1,i} + \mathbf{\Pi}_2\mathbf{m}_{2,i} + \cdots + \mathbf{\Pi}_k\mathbf{m}_{k,i} &= \lambda_i\mathbf{m}_{1,i}, \\ \mathbf{m}_{1,i} &= \lambda_i\mathbf{m}_{2,i}, \\ &\vdots \\ \mathbf{m}_{k-1,i} &= \lambda_i\mathbf{m}_{k,i}. \end{aligned} \tag{A.5}$$

Thus, each eigenvector satisfies:

$$\mathbf{m}_i = \begin{pmatrix} \mathbf{m}_{1,i} & \lambda_i^{-1}\mathbf{m}_{1,i} & \cdots & \lambda_i^{-k+1}\mathbf{m}_{1,i} \end{pmatrix}^T. \tag{A.6}$$

Furthermore, substituting Eq. A.6 back into Eq. A.5 yields:

$$\begin{aligned} \mathbf{\Pi}_1\mathbf{m}_{1,i} + \lambda_i^{-1}\mathbf{\Pi}_2\mathbf{m}_{1,i} + \cdots + \lambda_i^{-k+1}\mathbf{\Pi}_k\mathbf{m}_{1,i} &= \lambda_i\mathbf{m}_{1,i}, \\ \lambda_i^{-1}\mathbf{\Pi}_1\mathbf{m}_{1,i} + \lambda_i^{-2}\mathbf{\Pi}_2\mathbf{m}_{1,i} + \cdots + \lambda_i^{-k}\mathbf{\Pi}_k\mathbf{m}_{1,i} - \mathbf{m}_{1,i} &= \mathbf{0}, \\ \underbrace{\left(\lambda_i^{-1}\mathbf{\Pi}_1 + \lambda_i^{-2}\mathbf{\Pi}_2 + \cdots + \lambda_i^{-k}\mathbf{\Pi}_k - \mathbb{I}_p\right)}_{\mathbf{B}_i}\mathbf{m}_{1,i} &= \mathbf{0}. \end{aligned} \tag{A.7}$$

The vector $\mathbf{m}_{1,i}$, subsequently referred to as the "base" vector of the eigenvector $\mathbf{m}_i$, is a non-trivial member of the nullspace of the matrix polynomial denoted by $\mathbf{B}_i$, or equivalently, $\mathbf{m}_{1,i}$ is the eigenvector of $\mathbf{B}_i$ corresponding to a zero eigenvalue. In summary, the

eigendecomposition of the state transition matrix is given by $\mathbf{A} = \mathbf{M}\mathbf{D}\mathbf{M}^{-1}$, where:

$$\mathbf{M} = \begin{pmatrix} \mathbf{m}_{1,1} & \mathbf{m}_{1,2} & \cdots & \mathbf{m}_{1,pk} \\ \lambda_1^{-1}\mathbf{m}_{1,1} & \lambda_2^{-1}\mathbf{m}_{1,2} & \cdots & \lambda_{pk}^{-1}\mathbf{m}_{1,pk} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{-k+1}\mathbf{m}_{1,1} & \lambda_2^{-k+1}\mathbf{m}_{1,2} & \cdots & \lambda_{pk}^{-k+1}\mathbf{m}_{1,pk} \end{pmatrix}, \tag{A.8}$$

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{pk} \end{pmatrix}, \tag{A.9}$$

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{n}_{1,1}^T & \mathbf{n}_{1,2}^T & \cdots & \mathbf{n}_{1,k}^T \\ \mathbf{n}_{2,1}^T & \mathbf{n}_{2,2}^T & \cdots & \mathbf{n}_{2,k}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{n}_{pk,1}^T & \mathbf{n}_{pk,2}^T & \cdots & \mathbf{n}_{pk,k}^T \end{pmatrix}, \tag{A.10}$$

with the inverse eigenvector matrix also written in block form in order to match the overall block structure of the decomposition. In Section A.3.2, the zero-state response of the overall system is expressed as a function of the base vectors $\mathbf{m}_{1,i}$ and $\mathbf{n}_{i,1}^T$, and thus it is useful to understand in greater detail the relationship between them. Since $\mathbf{M}^{-1}\mathbf{A} = \mathbf{D}\mathbf{M}^{-1}$, each $\mathbf{n}_i^T$ vector is a left eigenvector of $\mathbf{A}$, and satisfies:

$$\mathbf{n}_i^T\mathbf{A} = \lambda_i\mathbf{n}_i^T. \tag{A.11}$$

Expressing $\mathbf{n}_i$ in block form gives:

$$\begin{pmatrix} \mathbf{n}_{i,1}^T & \mathbf{n}_{i,2}^T & \cdots & \mathbf{n}_{i,k}^T \end{pmatrix} \begin{pmatrix} \mathbf{\Pi}_1 & \cdots & \mathbf{\Pi}_{k-1} & \mathbf{\Pi}_k \\ \mathbb{I}_p & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbb{I}_p & \mathbf{0} \end{pmatrix} = \lambda_i \begin{pmatrix} \mathbf{n}_{i,1}^T & \mathbf{n}_{i,2}^T & \cdots & \mathbf{n}_{i,k}^T \end{pmatrix}.$$

Multiplying through produces the following system of matrix equations:

$$\begin{aligned}
\mathbf{n}_{i,1}^T\mathbf{\Pi}_1 + \mathbf{n}_{i,2}^T &= \lambda_i\mathbf{n}_{i,1}^T, \\
\mathbf{n}_{i,1}^T\mathbf{\Pi}_2 + \mathbf{n}_{i,3}^T &= \lambda_i\mathbf{n}_{i,2}^T, \\
&\vdots \\
\mathbf{n}_{i,1}^T\mathbf{\Pi}_{k-1} + \mathbf{n}_{i,k}^T &= \lambda_i\mathbf{n}_{i,k-1}^T, \\
\mathbf{n}_{i,1}^T\mathbf{\Pi}_k &= \lambda_i\mathbf{n}_{i,k}^T.
\end{aligned} \tag{A.12}$$

Each $\mathbf{n}_i^T$ vector satisfies:

$$\mathbf{n}_i^T = \begin{pmatrix} \mathbf{n}_{i,1} \\ \left(\sum_{j=2}^{k} \lambda_i^{-(k-j-1)}\mathbf{\Pi}_j^T\right)\mathbf{n}_{i,1} \\ \vdots \\ \left(\sum_{j=k}^{k} \lambda_i^{-(k-j+1)}\mathbf{\Pi}_j^T\right)\mathbf{n}_{i,1} \end{pmatrix}^T. \tag{A.13}$$

Whereas each $\mathbf{m}_{1,i}$ is in the right nullspace of $\mathbf{B}_i$, substituting the above block representation for $\mathbf{n}_i$ into Equation A.12 shows that the corresponding base vector $\mathbf{n}_{i,1}^T$ is in the left nullspace of $\mathbf{B}_i$, as:

$$\mathbf{n}_{i,1}^T \underbrace{\left(\lambda_i^{-1}\mathbf{\Pi}_1 + \lambda_i^{-2}\mathbf{\Pi}_2 + \cdots + \lambda_i^{-k}\mathbf{\Pi}_k - \mathbb{I}_p\right)}_{\mathbf{B}_i} = \mathbf{0}. \tag{A.14}$$

All that remains is to determine the appropriate scale factors for $\mathbf{m}_{1,i}$ and $\mathbf{n}_{i,1}^T$. Since $\mathbf{M}^{-1}\mathbf{M} = \mathbb{I}$, the following must hold:

$$\begin{aligned}
\mathbf{n}_i^T\mathbf{m}_i &= 1, \\
\mathbf{n}_{i,1}^T\left(\sum_{j=2}^{k}\sum_{k=1}^{j-1}\lambda_i^{-k}\mathbf{\Pi}_j\right)\mathbf{m}_{1,i} &= 1.
\end{aligned} \tag{A.15}$$

A.2.2 Modal Form

Once the state transition matrix has been factored according to $\mathbf{A} = \mathbf{M}\mathbf{D}\mathbf{M}^{-1}$, the state-space system given in Eq. A.2 can be rewritten as:

$$\begin{aligned} \mathbf{q}[n+1] &= \mathbf{M}\mathbf{D}\mathbf{M}^{-1}\mathbf{q}[n] + \mathbf{b}\mathbf{u}[n], \\ \mathbf{y}[n] &= \mathbf{c}^T\mathbf{q}[n] + \mathbf{d}\mathbf{u}[n]. \end{aligned}$$

Multiplying the top equation by $\mathbf{M}^{-1}$ and letting $\mathbf{r} = \mathbf{M}^{-1}\mathbf{q}$, yields the following state-space model in the transformed coordinate system:

$$\begin{aligned} \mathbf{r}[n+1] &= \underbrace{\mathbf{D}}_{\tilde{\mathbf{A}}}\mathbf{r}[n] + \underbrace{\mathbf{M}^{-1}\mathbf{b}}_{\tilde{\mathbf{b}}}\mathbf{u}[n], \\ \mathbf{y}[n] &= \underbrace{\mathbf{c}^T\mathbf{M}}_{\tilde{\mathbf{c}}^T}\mathbf{r}[n] + \underbrace{\mathbf{d}}_{\tilde{\mathbf{d}}}\mathbf{u}[n]. \end{aligned} \tag{A.16}$$

The new state transition matrix, $\tilde{\mathbf{A}} = \mathbf{D}$, is a diagonal matrix, indicating that the evolution of each state or *mode* depends only on its own past, having been completely decoupled from all other states or modes. This transformation separates the states according to the eigenspectrum of $\mathbf{A}$, and the new state-space system is said to be in *modal form*; each eigenvalue is called a *modal frequency*, with each corresponding eigenvector a *modal shape*.

One of the main benefits of expressing the system in modal form is that many of its properties can be inferred by inspection. For example, examining the i^{th} row and j^{th} column of the matrix $\tilde{\mathbf{b}}$, given by:

$$\tilde{\mathbf{b}} = \mathbf{M}^{-1}\mathbf{b} = \begin{pmatrix} \mathbf{n}_{1,1}^T \\ \mathbf{n}_{2,1}^T \\ \vdots \\ \mathbf{n}_{pk,1}^T \end{pmatrix}$$

reveals whether mode j is *reachable* from input i. Furthermore, examination of the corresponding entries in the matrix $\tilde{\mathbf{c}}^T$, defined by:

$$\begin{aligned}
\tilde{\mathbf{c}}^T &= \mathbf{c}^T\mathbf{M} \\
&= \begin{pmatrix} \mathbf{\Pi}_1 & \cdots & \mathbf{\Pi}_{k-1} & \mathbf{\Pi}_k \end{pmatrix} \begin{pmatrix} \mathbf{m}_{1,1} & \mathbf{m}_{1,2} & \cdots & \mathbf{m}_{1,pk} \\ \lambda_1^{-1}\mathbf{m}_{1,1} & \lambda_2^{-1}\mathbf{m}_{1,2} & \cdots & \lambda_{pk}^{-1}\mathbf{m}_{1,pk} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{-k+1}\mathbf{m}_{1,1} & \lambda_2^{-k+1}\mathbf{m}_{1,2} & \cdots & \lambda_{pk}^{-k+1}\mathbf{m}_{1,pk} \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1\mathbf{m}_{1,1} & \lambda_2\mathbf{m}_{1,2} & \cdots & \lambda_{pk}\mathbf{m}_{1,pk} \end{pmatrix}
\end{aligned}$$

indicates whether mode i is *observable* at output j, where the last line follows from Equation A.7. The asymptotic stability of the system (i.e., whether or not the zero-input response of the system, to be discussed in Section A.3.1, decays to zero) can be determined by verifying that all the eigenvalues lie strictly inside the unit circle of the complex plane. Lastly, this modal form representation is utilized in Section A.3.2, in order to determine the overall transfer function of the VAR system.

A.2.3 Jordan Canonical Form

Consider the case in which the state transition matrix $\mathbf{A}$ contains at least one eigenvalue whose geometric multiplicity is less than its algebraic multiplicity. While diagonalization is no longer possible, the matrix $\mathbf{A}$ can be put into Jordan Canonical Form, which also takes the form $\mathbf{A} = \mathbf{MDM}^{-1}$. The matrix $\mathbf{D}$ is no longer diagonal, but is rather block diagonal, where each of the *Jordan blocks*, J_i, has the form:

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}. \tag{A.17}$$

Under this representation, the algebraic multiplicity of an eigenvalue corresponds to the number of repeated roots of the characteristic equation, while the geometric multiplicity corresponds to the total number of Jordan blocks. Diagonalization is a special case of the Jordan Form where all of the Jordan blocks are 1×1 matrices.

The columns of the matrix $\mathbf{M}$ now contain the corresponding *generalized eigenvectors*, described in Definition A.1.

Definition A.1.

Let $\mathbf{A}$ *be a linear operator with eigenvalue* λ *of rank* r *(i.e., the size of the corresponding Jordan block is* $r \times r$*). Then* $\mathbf{m}_i \in \mathbb{R}^{pk}$ *is a* generalized eigenvector *of* $\mathbf{A}$ *if* $(\mathbf{A} - \lambda\mathbb{I})^r \mathbf{m}_i = \mathbf{0}$*, but* $(\mathbf{A} - \lambda\mathbb{I})^{r-1} \mathbf{m}_i \neq \mathbf{0}$*. The set of vectors* $\left\{\mathbf{m}^1, \cdots, \mathbf{m}^r\right\}$*, defined as:*

$$\begin{aligned} \mathbf{m}^r &\triangleq \mathbf{m}_i \\ \mathbf{m}^{r-1} &\triangleq (\mathbf{A} - \lambda\mathbb{I})\,\mathbf{m}^r \\ &\vdots \\ \mathbf{m}^1 &\triangleq (\mathbf{A} - \lambda\mathbb{I})\,\mathbf{m}^2 \end{aligned}$$

is then referred to as a chain of generalized eigenvectors.

The remainder of this section contains a derivation of the general eigenvector structure in a VAR process corresponding to an eigenvalue with rank 2. These eigenvectors are subsequently used in Section 3.3.5 to obtain a condition that guarantees that the underlying processes of a cointegrated VAR are at most integrated of order one.

Consider a matrix $\mathbf{A}$ with eigenvalue λ_i, with algebraic multiplicity two and geometric multiplicity one, so that the corresponding Jordan block J_i has the form:

$$J_i = \begin{pmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{pmatrix}.$$

The two generalized eigenvectors $\{\mathbf{m}_1, \mathbf{m}_2\}$ satisfy:

$$(\mathbf{A} - \lambda_i\mathbb{I})^2\,\mathbf{m}_2 = 0, \tag{A.18}$$

$$(\mathbf{A} - \lambda_i\mathbb{I})\,\mathbf{m}_2 = \mathbf{m}_1. \tag{A.19}$$

Combining the two equations gives:

$$(\mathbf{A} - \lambda_i \mathbb{I})\, \mathbf{m}_1 \;=\; 0,$$

which implies that the vector $\mathbf{m}_1$ has the same block structure given by Equation A.6.

The explicit form of the generalized eigenvector $\mathbf{m}_2$ can be found using Equation A.19. Writing each eigenvector in block form yields:

$$(\mathbf{A} - \lambda_i \mathbb{I})\, \mathbf{m}_2 \;=\; \mathbf{m}_1,$$

$$\begin{pmatrix} \mathbf{\Pi}_1 - \lambda_i \mathbb{I} & \mathbf{\Pi}_2 & \cdots & \mathbf{\Pi}_{k-1} & \mathbf{\Pi}_k \\ \mathbb{I}_p & -\lambda_i \mathbb{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\lambda_i \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbb{I}_p & -\lambda_i \mathbb{I} \end{pmatrix} \begin{pmatrix} \mathbf{m}_{1,2} \\ \mathbf{m}_{2,2} \\ \vdots \\ \mathbf{m}_{k-1,2} \\ \mathbf{m}_{k,2} \end{pmatrix} = \begin{pmatrix} \mathbf{m}_{1,1} \\ \lambda_i^{-1} \mathbf{m}_{1,1} \\ \vdots \\ \lambda_i^{k-2} \mathbf{m}_{1,1} \\ \lambda_i^{k-1} \mathbf{m}_{1,1} \end{pmatrix}.$$

Multiplying out the last $k-1$ rows gives the following system of equations:

$$\begin{aligned} \mathbf{m}_{1,2} - \lambda_i \mathbf{m}_{2,2} &= \lambda_i^{-1} \mathbf{m}_{1,1}, \\ &\vdots \\ \mathbf{m}_{k-2,2} - \lambda_i \mathbf{m}_{k-1,2} &= \lambda_i^{-(k-2)} \mathbf{m}_{1,1}, \\ \mathbf{m}_{k-1,2} - \lambda_i \mathbf{m}_{k,2} &= \lambda_i^{-(k-1)} \mathbf{m}_{1,1}, \end{aligned}$$

producing the following overall block structure for $\mathbf{m}_2$:

$$\mathbf{m}_2 = \left(\mathbf{m}_{1,2} \quad \tfrac{\lambda_i \mathbf{m}_{1,2} - \mathbf{m}_{1,1}}{\lambda_i^2} \quad \cdots \quad \tfrac{\lambda_i \mathbf{m}_{1,2} - (k-1)\mathbf{m}_{1,1}}{\lambda_i^k} \right)^T. \tag{A.20}$$

The rank 2 eigenvector is therefore a function of its own base vector, $\mathbf{m}_{1,2}$, and the base vector for the eigenvector of rank 1, $\mathbf{m}_{1,1}$.

A.3 System Response

The overall response of a linear system can always be decomposed into two components corresponding to the *zero-input response* and the *zero-state response*, as described next.

A.3.1 Zero Input Response

Let $\mathbf{u}[n] = \mathbf{0}$ for all $n \geq 0$, and let $\mathbf{q}[0]$ denote the initial state of the VAR system. The initial state vector can be expressed in the oblique (non-orthogonal) basis given by the pk linearly independent eigenvectors of the diagonalizable state transition matrix $\mathbf{A}$, as:

$$\mathbf{q}[0] = \sum_{i=1}^{pk} g_i \mathbf{m}_i = \begin{pmatrix} \mathbf{m}_1 & \cdots & \mathbf{m}_{pk} \end{pmatrix} \begin{pmatrix} g_1 \\ \vdots \\ g_{pk} \end{pmatrix} = \mathbf{M}\mathbf{g},$$

where the g_i are the expansion coefficients defined by $\mathbf{g} = \mathbf{M}^{-1}\mathbf{q}[0]$. Using this representation, the state vector at time n is given by:

$$\mathbf{q}[n] = \mathbf{A}^n \mathbf{q}[0] = \sum_{i=1}^{N=pk} g_i \mathbf{A}^n \mathbf{m}_i = \sum_{i=1}^{N=pk} g_i \lambda_i^n \mathbf{m}_i,$$

and the system output is equal to:

$$\mathbf{y}_{zir}[n] = \mathbf{c}^T \mathbf{q}[n] = \begin{pmatrix} \mathbf{\Pi}_1 & \cdots & \mathbf{\Pi}_{k-1} & \mathbf{\Pi}_k \end{pmatrix} \sum_{i=1}^{pk} g_i \lambda_i^n \mathbf{m}_i, \tag{A.21}$$

Combining Equations A.6 and A.21 gives:

$$\begin{aligned}
\mathbf{y}_{zir}[n] &= \sum_{i=1}^{pk} g_i \lambda_i^n \begin{pmatrix} \mathbf{\Pi}_1 & \mathbf{\Pi}_2 & \cdots & \mathbf{\Pi}_k \end{pmatrix} \begin{pmatrix} \mathbf{m}_{1,i} \\ \lambda_i^{-1}\mathbf{m}_{1,i} \\ \vdots \\ \lambda_i^{-(k-1)}\mathbf{m}_{1,i} \end{pmatrix}, \\
&= \sum_{i=1}^{pk} g_i \lambda_i^n \left(\mathbf{\Pi}_1 \mathbf{m}_{1,i} + \mathbf{\Pi}_2 \lambda_i^{-1} \mathbf{m}_{1,i} + \cdots + + \mathbf{\Pi}_k \lambda_i^{-(k-1)} \mathbf{m}_{1,i} \right), \\
&= \sum_{i=1}^{pk} g_i \lambda_i^n \left(\mathbf{\Pi}_1 + \mathbf{\Pi}_2 \lambda_i^{-1} + \cdots + + \mathbf{\Pi}_k \lambda_i^{-(k-1)} \right) \mathbf{m}_{1,i}, \\
&= \sum_{i=1}^{pk} g_i \lambda_i^{n+1} \left(\lambda_i^{-1} \mathbf{\Pi}_1 + \lambda_i^{-2} \mathbf{\Pi}_2 + \cdots + + \lambda_i^{-(k-1)} \mathbf{\Pi}_k \right) \mathbf{m}_{1,i}, \\
&= \sum_{i=1}^{pk} g_i \lambda_i^{n+1} \mathbf{m}_{1,i},
\end{aligned} \tag{A.22}$$

where the last line follows from Equation A.7. Thus, the system output due to the zero-input response can be expressed as a linear combination of the base vectors of the eigenvectors of $\mathbf{A}$, denoted as $\mathbf{m}_{1,i}$.

A.3.2 Zero State Response

The zero-state response describes the behavior of the system when $\mathbf{q}[0] = \mathbf{0}$, and is equal to the matrix convolution of the system inputs, $\mathbf{u}[n]$, with the system transfer function matrix, $\mathbf{h}[n]$. The corresponding $\mathcal{Z}-$domain transfer function matrix is denoted by $\mathbf{H}(z)$. Once this matrix of transfer functions from each input to each output has been computed and expressed in a simplified form using a partial fraction expansion, $\mathbf{h}[n]$ is found by taking the inverse $\mathcal{Z}-$transform.

The $\mathcal{Z}-$transform matrix for the system defined by Eq. A.1 is:

$$\mathbf{H}(z) = \left(\mathbb{I} - z^{-1}\mathbf{\Pi}_1 + \ldots + z^{-(k-1)}\mathbf{\Pi}_{k-1} + z^{-k}\mathbf{\Pi}_k\right)^{-1} = \mathbf{F}(z)^{-1} = \frac{\text{adj}(\mathbf{F}(z))}{\det(\mathbf{F}(z))}, \quad \text{(A.23)}$$

where adj$(\cdot)$ denotes the adjugate matrix and det$(\cdot)$ denotes the determinant. The adjugate is the transpose of the matrix of co-factors of $\mathbf{F}(z)$, where each co-factor, $\mathbf{C}_{ij}$, is the determinant of the matrix formed by deleting row i and column j from $\mathbf{F}(z)$. The adjugate matrix can also be computed using the characteristic polynomial of $\mathbf{F}(z)$, defined by:

$$p(t) = \det\left(\mathbf{F}(z) - t\mathbb{I}\right) = p_0 + p_1 t + \cdots + p_n t^n.$$

Then the adjugate is given as:

$$\text{adj}(\mathbf{F}(z)) = -\left(p_1\mathbb{I} + p_2\mathbf{F}(z) + p_3\mathbf{F}^2(z) + \cdots + p_n\mathbf{F}^{n-1}(z)\right) \quad \text{(A.24)}$$

The determinant of $\mathbf{F}(z)$ is a polynomial in z^{-1} of degree pk, which is identical to the polynomial in Equation A.4, used to compute the eigenvalues of $\mathbf{A}$. Since the denominator of $\mathbf{H}(z)$ is a polynomial in z, Equation A.23 can be expressed using a partial fraction

expansion, as:

$$\mathbf{H}(z) = \frac{adj(\mathbf{F}(z))}{\det(\mathbf{F}(z))} = \sum_{i=1}^{p^2} \frac{\mathbf{R}_i}{1 - \lambda_i z^{-1}}, \tag{A.25}$$

where each residue matrix, $\mathbf{R}_i$, is given by:

$$\mathbf{R}_i = \left. \frac{\text{adj}(\mathbf{F}(z))(1 - \lambda_i z^{-1})}{\det(\mathbf{F}(z))} \right|_{z=\lambda} = \frac{\text{adj}(\mathbf{F}(z = \lambda_i))}{\prod_{\lambda_j \neq \lambda_i}^{p^2} (1 - \lambda_j)}$$

It is also possible to derive Equation A.25 using the transformed state-space system in modal form, given in Equation A.16, as:

$$\mathbf{H}(z) = \tilde{\mathbf{c}}^T \left(z\mathbb{I} - \tilde{\mathbf{A}} \right)^{-1} \tilde{\mathbf{b}} + \tilde{\mathbf{d}}, \tag{A.26}$$

$$= \begin{pmatrix} \lambda_1 \mathbf{m}_{1,1} & \lambda_2 \mathbf{m}_{1,2} & \cdots & \lambda_{pk} \mathbf{m}_{1,pk} \end{pmatrix} \begin{pmatrix} z - \lambda_1 & 0 & \cdots & 0 \\ 0 & z - \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z - \lambda_{pk} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{n}_{1,1}^T \\ \mathbf{n}_{2,1}^T \\ \vdots \\ \mathbf{n}_{pk,1}^T \end{pmatrix} + \mathbb{I},$$

$$= \sum_{i=1}^{pk} \frac{\lambda_i \mathbf{m}_{1,i} \mathbf{n}_{i,1}^T}{z - \lambda_i} + \mathbb{I},$$

$$= \sum_{i=1}^{pk} \frac{\mathbf{m}_{1,i} \mathbf{n}_{i,1}^T}{1 - \lambda_i z^{-1}}. \tag{A.27}$$

Each residue is also given by the outer product of the base vectors from the eigendecomposition of $\mathbf{A}$, as $\mathbf{R}_i = \mathbf{m}_{1,i} \mathbf{n}_{i,1}^T$. Taking the inverse $\mathcal{Z}-$transform, the unit impulse response is given by:

$$\mathbf{h}[n] = \sum_{i=1}^{pk} \lambda_i^n \mathbf{R}_i \mathbf{s}[n]$$

where $\mathbf{s}[n]$ is the vector form of the unit-step function, where each component series $\mathbf{s}_i[n]$ has the form:

$$\mathbf{s}_i[n] = \begin{cases} 1, & \text{if } n \geq 0 \\ 0, & \text{otherwise} \end{cases}. \tag{A.28}$$

The system output due to the zero-state response can be expressed as:

$$\begin{aligned}\mathbf{y}_{zsr}[n] &= \mathbf{h}[n] * \mathbf{u}[n] = \sum_{i=1}^{pk} \lambda_i^n \mathbf{R}_i \mathbf{s}[n] * \mathbf{u}[n] \\ &= \sum_{i=1}^{pk} \mathbf{R}_i \sum_{j=1}^{n} \lambda_i^j \mathbf{u}[n-j] = \sum_{i=1}^{pk} \mathbf{m}_{1,i} \mathbf{n}_{i,1}^T \sum_{j=1}^{n} \lambda_i^j \mathbf{u}[n-j].\end{aligned}$$

A.3.3 Total System Response

The overall response of the vector autoregressive system presented here is the sum of the components corresponding to the *zero-input response* and the *zero-state response*, as:

$$\begin{aligned}\mathbf{y}[n] &= \mathbf{y}_{zir}[n] + \mathbf{y}_{zsr}[n], \\ &= \sum_{i=1}^{pk} g_i \lambda_i^{n+1} \mathbf{m}_{1,i} + \sum_{i=1}^{pk} \mathbf{m}_{1,i} \mathbf{n}_{i,1}^T \sum_{j=1}^{n} \lambda_i^j \mathbf{u}[n-j], \\ &= \sum_{i=1}^{pk} \mathbf{m}_{1,i} \left(g_i \lambda_i^{n+1} + \sum_{j=1}^{n} \lambda_i^j \mathbf{n}_{i,1}^T \mathbf{u}[n-j] \right). \end{aligned} \tag{A.29}$$

Hence, the overall system response may be expressed as a time-varying linear combination of the $\mathbf{m}_{1,i}$ vectors.

A.4 Estimation

The maximum likelihood (ML) estimation procedure for the unrestricted VAR process given by Eq. A.1 is summarized here. It is assumed that the stochastic input, $\boldsymbol{\epsilon}_n$, is independently and identically distributed (i.i.d.) according to a Normal distribution with zero mean and covariance matrix $\boldsymbol{\Psi}$, with known deterministic input, $\mathbf{d}[n]$. Therefore, the set of parameters to be estimated is given by $\{\boldsymbol{\Pi}_1, ..., \boldsymbol{\Pi}_k, \boldsymbol{\Phi}, \boldsymbol{\Psi}\}$. It is easiest to derive the ML estimators by rewriting Eq. A.1 in linear regression form, as follows:

$$\mathbf{x}[n] = \boldsymbol{\theta}^T \mathbf{z}[n] + \boldsymbol{\epsilon}[n],$$

where $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \cdots & \boldsymbol{\Pi}_k & \boldsymbol{\Phi} \end{pmatrix}^T$ and $\mathbf{z}[n] = \begin{pmatrix} \mathbf{x}[n-1] & \cdots & \mathbf{x}[n-k] & \mathbf{d}[n] \end{pmatrix}^T$. Given a random sample of data $\left\{\mathbf{x}^1, ..., \mathbf{x}^M\right\}$, the likelihood of the data given the parameters $\boldsymbol{\theta}$ and

$\boldsymbol{\Psi}$ is given by:

$$L\left(\boldsymbol{\theta},\boldsymbol{\Psi}\right)=Pr_{\boldsymbol{\theta},\boldsymbol{\Psi}}\left(\mathbf{x}^{1},...,\mathbf{x}^{M}\right)=\prod_{i=1}^{M}\frac{1}{(2\pi)^{p/2}\left|\boldsymbol{\Psi}\right|^{1/2}}\exp\left[-0.5\left(\mathbf{x}^{i}-\boldsymbol{\theta}^{T}\mathbf{z}^{i}\right)^{T}\boldsymbol{\Psi}^{-1}\left(\mathbf{x}^{i}-\boldsymbol{\theta}^{T}\mathbf{z}^{i}\right)\right],$$

with corresponding log-likelihood function:

$$\ell\left(\boldsymbol{\theta},\boldsymbol{\Psi}\right)=-\frac{p}{2}M\log(2\pi)-\frac{1}{2}M\log\left(\left|\boldsymbol{\Psi}\right|\right)-\frac{1}{2}\sum_{i=1}^{M}\left(\mathbf{x}^{i}-\boldsymbol{\theta}^{T}\mathbf{z}^{i}\right)^{T}\boldsymbol{\Psi}^{-1}\left(\mathbf{x}^{i}-\boldsymbol{\theta}^{T}\mathbf{z}^{i}\right).$$

The ML estimators for $\boldsymbol{\theta}$ and $\boldsymbol{\Psi}$ are defined as the solutions to the following optimization problem:

$$\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\Psi}}=\underset{\boldsymbol{\theta},\boldsymbol{\Psi}}{\arg\max}\ \ell\left(\boldsymbol{\theta},\boldsymbol{\Psi}\right).$$

Differentiating the log-likelihood function with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\Psi}$ gives:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \left(\sum_{i=1}^{M}\mathbf{z}^{i}\mathbf{z}^{iT}\right)^{-1}\left(\sum_{i=1}^{M}\mathbf{z}^{i}\mathbf{x}^{iT}\right),\\ \hat{\boldsymbol{\Psi}} &= \frac{1}{M}\sum_{i=1}^{M}\left(\mathbf{x}^{i}-\hat{\boldsymbol{\theta}}^{T}\mathbf{z}^{i}\right)\left(\mathbf{x}^{i}-\hat{\boldsymbol{\theta}}^{T}\mathbf{z}^{i}\right)^{T}.\end{aligned}$$

Due to the assumption that $\boldsymbol{\epsilon}_{n}$ is Normally distributed, the ML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\Psi}$ for an unrestricted VAR correspond to the familiar least-squares estimators.

Appendix B

Importance Sampling

Let $\mathbf{x} \in \mathbb{R}^p$ denote a random vector distributed according to $f(\mathbf{x})$. The expectation of a function of $\mathbf{x}$, denoted as $g(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^q$, is defined as:

$$E\left[g(\mathbf{x})\right] = \int_{\mathbf{x}} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \tag{B.1}$$

where the integral is evaluated over each dimension of $\mathbf{x}$. Depending on the choice of f and g, it may or may not be possible to evaluate this integral in closed-form. The expectation can be approximated using a sample mean, as:

$$E\left[g(\mathbf{x})\right] \approx \sum_{i=1}^{N} g(\mathbf{x}_i),$$

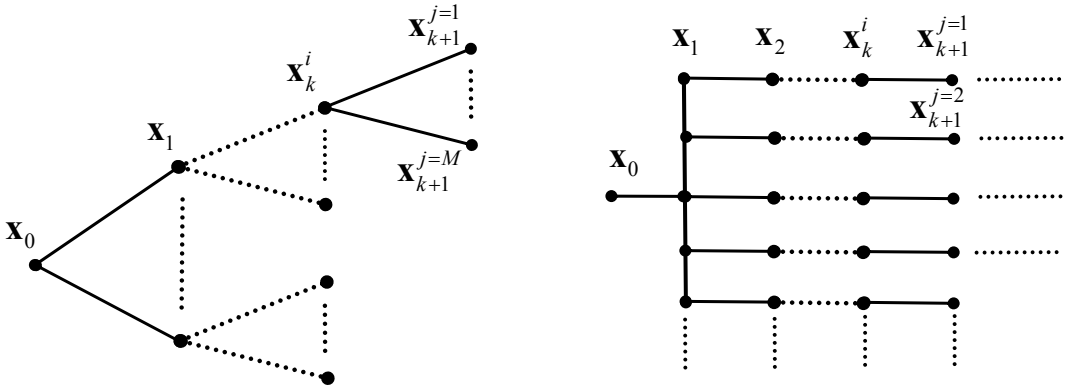
where each i.i.d. sample, $\mathbf{x}_i$, is drawn from the density $f(\mathbf{x})$. According to the Strong Law of Large Numbers, this approximation converges *almost surely* to the true expected value [50].

When it is not possible to sample directly from $f(\mathbf{x})$, the principle of importance sampling may be used. First, Eq. B.1 is equivalently represented as:

$$E\left[g(\mathbf{x})\right] = \int_{\mathbf{x}} g(\mathbf{x}) \frac{f(\mathbf{x})}{h(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}, \tag{B.2}$$

where the distribution $h(\mathbf{x})$ is chosen so that one can easily generate samples of $\mathbf{x}$ from it. This implies that $E\left[g(\mathbf{x})\right]$ can also be approximated as:

$$E\left[g(\mathbf{x})\right] \approx \sum_{i=1}^{N} g(\mathbf{x}_i) w(\mathbf{x}_i),$$



(a) Sample paths of log-price process are generated according to a tree. The size of the tree grows exponentially with the number of stages.

(b) In order to avoid the curse of dimensionality, a fixed number of sample paths of the log-price process are generated.

Figure B-1. Sample path simulation schemes.

where each i.i.d. sample, $\mathbf{x}_i$, is now drawn from the density $h(\mathbf{x})$, and the value of $g(\mathbf{x}_i)$ is weighted by the scale factor $w(\mathbf{x}_i) = \frac{f(\mathbf{x})}{h(\mathbf{x})}$.

While the importance sampling method applied to Monte Carlo integration is typically used in cases where the target distribution ($f(\mathbf{x})$) is "hard" to sample from, it can also be used when sampling from f is "easy", but computationally intractable. The latter scenario often occurs in a sequential setting, where one may need to generate samples of a random process over time. These samples could be generated according to an M-ary tree, as depicted in the left panel of Figure B-1. However, this naive approach suffers from the curse of dimensionality, as the size of the tree grows exponentially with the number of stages. This limitation may be overcome by using the simulation scheme depicted in the right panel of Figure B-1, in which a fixed number of price paths are generated. While the new grid is more efficient, samples from the conditional target distribution $f(\mathbf{x}_{k+1}|\mathbf{x}_k)$ are no longer available, and the importance sampling techniques described here must be used instead, as described in Section 5.1.3.

Bibliography

[1] SEC halts short selling of financial stocks to protect investors and markets, September 2008. `http://www.sec.gov/news/press/2008/2008-211.htm`.

[2] C. Alexander and A. Dimitriu. Index and statistical arbitrage. *Journal of Portfolio Management*, 31(2):50–63, 2005.

[3] T. W. Anderson. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, 22:327–351, 1951.

[4] K. Arrow. *Essays in the Theory of Risk Bearing*, chapter The Theory of Risk Aversion. Markham Economics Series, 1970.

[5] R. Baillie and T. Bollerslev. Common stochastic trends in a system of exchange rates. *The Journal of Finance*, 44(1):167–181, 1989.

[6] R. Baillie and T. Bollerslev. Cointegration, fractional cointegration, and exchange rate dynamics. 49(2):737–745, 1994.

[7] S. Basak and G. Chabakauri. Dynamic mean-variance asset allocation. *Working Paper, London Business School*, 2007.

[8] D. Bauer and M. Wagner. Estimating cointegrated systems using subspace algorithms. *Journal of Econometrics*, 111(1):47–84, 2002.

[9] D. Bauer and M. Wagner. A canonical form for unit root processes in the state space framework. *Working Paper, University of Bern*, 2003.

[10] D. Bauer and M. Wagner. The performance of subspace algorithm cointegration analysis: A simulation study. *Working Paper, University of Bern*, 2003.

[11] V. S. Bawa. Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2:95–121, 1975.

[12] S. E. Beck. Cointegration and market efficiency in commodities futures markets. *Applied Economics*, 26:249–257, 1994.

[13] D. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, MA, 2000.

[14] R. Bewley and M. Yang. Tests for cointegration based on canonical correlation analysis. *Journal of the American Statistical Association*, 90:990–996, 1995.

[15] H. Bierens. Nonparametric cointegration analysis. *Journal of Econometrics*, 77:379–404, 1997.

[16] P. Bossaerts. Common nonstationary components of asset prices. *Journal of Economic Dynamics and Control*, 12:348–364, 1988.

[17] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[18] A. N. Burgess. *Applied Quantitative Methods for Trading and Investment*, chapter Using Cointegration to Hedge and Trade International Equities. Wiley, 2003.

[19] J. Campbell, A. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, 1997.

[20] J. Campbell and L. Viceira. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press, 2002.

[21] C. T. Chen. *Introduction to Linear System Theory*. Holt, Rinhart and Winston, 1970.

[22] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association,*, 74:427–431, 1979.

[23] F. Diebold, J. Gardeazbal, and K. Yilmaz. On cointegration and exchange rate dynamics. *The Journal of Finance*, 49(2):727–735, 1994.

[24] R. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276, 1987.

[25] C. P. Fishburn. Mean-risk analysis with risk associated with below-target returns. *American Economic Review*, 67(5):116–126, 1977.

[26] A. Hall, H. Anderson, and C. W. J. Granger. A cointegration analysis of treasury bill yields. *The Review of Economics and Statistics*, 74:116–126, 1992.

[27] J. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton NJ, 1994.

[28] Peter R. Hansen. Granger's representation theorem: A closed-form expression for I(1) processes. *Econometrics Journal*, 8:23–38, 2005.

[29] S. Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, 1988.

[30] S. Johansen. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, New York, 1995.

[31] Y. Kawasaki, S. Tachiki, H. Udaka, and T. Hirano. A characterization of long-short trading strategies based on cointegration. *Proceedings of the 2003 International Conference on Computational Intelligence for Financial Engineering (CIFEr2003)*, pages 411–416, 2003.

[32] F. Kleibergen and R. Paap. Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics*, 111:223–249, 2002.

[33] G. Koop, R. Strachan, H. K. van Dijk, and M. Villani. *Palgrave Handbook of Econometrics, Volume 1*, chapter Bayesian approaches to cointegration. Palgrave Macmillan, 2006.

[34] M. Lettau and S. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, 56(3):815–849, 2001.

[35] D. Li, T. Chan, and W. Ng. Safety-first dynamic portfolio selection. *Dynamic of Continuous, Discrete, and Impulsive Systems*, 4:858–600, 1998.

[36] D. Li and W. Ng. Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical Finance*, 10(3):387–406, 2000.

[37] J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965.

[38] S. Lohr. In modeling risk, the human factor was left out. The New York Times, 2008. http://www.nytimes.com/2008/11/05/business/05risk.html?partner=permalink & exprod=permalink.

[39] R. Lowenstein. *When Genious Failed: The Rise and Fall of Long-Term Capital Management*. Random House, 2000.

[40] A. Lucas. Strategic and tactical asset allocation and the effect of long-run equilibrium relations. Research Memorandum 1997-42, Vrije Universiteit, Amsterdam, August 1997.

[41] S. Ludvigson and C. Steindel. How important is the stock market effect on consumption? *Economic Policy Review, Federal Reserve Bank of New York*.

[42] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[43] H. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1959.

[44] R. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51:139–162, 1969.

[45] J. Mossin. Optimal multiperiod portfolio policies. *The Journal of Business*, 41(2):215–229, 1968.

[46] M. Murray. A drunk and her dog: An illustration of cointegration and error correction. *The American Statisitician*, 48(1):37–39, 1994.

[47] R. Oberuc. *Dynamic Portfolio Theory and Management: Using Asset Allocation to Improve Profits and Reduce Risk*. McGraw-Hill, 2003.

[48] W. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley-Interscience, 2007.

[49] J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1):122–136, 1964.

[50] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

[51] A. D. Roy. Safety-first and the holding of assets. *Econometrica*, 20(3):431–449, 1952.

[52] P. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economic and Statistics*, 51(3):239–246, 1969.

[53] B. Scholkopf and A. Smola. *Learning with Kernels*. The MIT Press, Cambridge, MA, 2002.

[54] A. Scowcroft and J. Sefton. *Advances in Portfolio Construction and Implementation*, chapter Enhanced Indexation. Butterworth-Heinemann, 2003.

[55] W. F. Sharpe. The sharpe ratio. *Journal of Portfolio Management*, 21(1):49–58, 1994.

[56] J. Siegel. *Stocks for the Long Run, Fourth Edition*. McGraw-Hill Companies, 2007.

[57] J. Stock and M. Watson. Testing for common trends. *Journal of the American Statistical Association*, 83(404):1097–1107, 1988.

[58] J. Tobin. Liquidity preference as behavior towards risk. *The Review of Economic Studies*, 67:65–86, 1958.

[59] J. Tobin. *Theory of Interest Rates*, chapter The Theory of Portfolio Selection. MacMillan and Co Ltd, 1962.

[60] R. Tsay. *Analysis of Financial Time Series*. Wiley, 2005.

[61] M. Villani. Bayesian point estimation of the cointegration space. *Journal of Econometrics*, 134:645–664, 2006.

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

[63] M. Wagner. A comparison of johansen's, bierens', and the subspace algorithm method for cointegration analysis. *Oxford Bulletin of Economics and Statistics*, 66(3):399–424, 2004.

[64] W. H. Wagner. Best execution. *Financial Analysts Journal*, 49(1):65–71, 1993.

[65] M. Wallace and J. T. Warner. The fisher effect and the tern structure of interest rates: Tests of cointegration. *The Review of Economics and Statistics*, 1993.

[66] J. Yong and X. Y. Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer, 1999.

[67] G. Zelniker and F. Taylor. *Advanced Digital Signal Processing: Theory and Applications*. CRC Press, 1993.

[68] H. Zhang. Treasury yield curves and cointegration. *Applied Economics*, 25:361–367, 1993.