ESSAYS ON PORTFOLIO SELECTION

by

Thiago de Oliveira Souza

A Thesis submitted to the faculty of
University of London
Queen Mary
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D.) in Finance

Department of School of Economics and Finance

University of London

Queen Mary

December 2011

AUTHOR'S DECLARATION

I wish to declare this thesis, titled as "Essays on Portfolio Selection", submitted to the University of London in pursuance of the degree of Doctor of Philosophy (Ph.D.) in Finance is my own work.

Signed

Thiago de Oliveira Souza

ABSTRACT

This thesis began with an introduction and literature review in Chapter 1. In Chapter 2, I propose a new intertemporal asset-pricing model based on heterogeneous beliefs to bring together the concurrent theories that could generate value and momentum effects. In this model, I assume that such behaviour occurs simply due to an agnostic view of forecasting returns considering the dominant strategy in the market. Given the endogenous price determination in the model, individuals were expected to adjust their own strategies to match the dominant strategy to obtain higher profits (from more accurate forecasts). The idea was to bridge the literature on intertemporal asset allocation with the one on heterogeneous beliefs.

In Chapters 3 and 4, I consider the empirical problem of implementing Markowitz (1952) mean-variance optimisation on a portfolio of stocks. In particular, I focus on the out-of-sample performance of the minimum-variance portfolio obtained from the use of asset group information and regularisation methods to obtain more stable estimates of the parameters in the model.

Specifically, in Chapter 3, I introduce the use of regularisation methods to the portfolio selection problem and a literature review on

the subject. In Chapter 4, I propose two alternative approaches for the use of the group structure information and to obtain more stable and regularised minimum-variance portfolios. I show that these procedures produce significantly better results in the portfolios compared with the unconstrained minimum-variance portfolios estimated from the whole data set in terms of portfolio variance and the Sharpe ratio.

ACKNOWLEDGEMENT

As with everything I achieve in life, I would like to thank my family: My mother, my father, my grandparents, my three brothers and my sister. I am grateful for the support that has allowed me pursue this project. This thesis is not solely the result of my own work, but was possible because of the combination of my family support and academic mentorship.

I would also like to thank two of my advisors who are also my friends: Prof. Marcelo Fernandes and Prof. Marco Bonomo. Marcelo advised me during my Ph.D. training, and Marco advised me during my MSc. degree. In addition to the outstanding help, patience, support and guidance during my research, Marcelo also believed in this project and really "bought" the idea, frequently exceeding what one would expect from an advisor. Marco was the person from whom I first learned what being a researcher in Finance was about. He is the one who introduced me to the topic and also the first to believe in my project. During the Ph.D. he has also performed the time consuming role of my tutor along with CAPES and I am also really grateful for that.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

Chapter 1

Introduction

1.1 Outline

The subject of portfolio selection is considerably broad. In this thesis, I concisely cover both theoretical and empirical issues due to space concerns. In Chapter 2, which is the more theoretically oriented chapter, I propose a new intertemporal asset-pricing model based on heterogeneous beliefs to bring together the concurrent theories that could generate value and momentum effects.

In this model, I assume that such behaviour occurs simply due to an agnostic view of forecasting returns considering the dominant strategy in the market. Given the endogenous price determination in the model, individuals are expected to adjust his or her own strategies to match the dominant strategy to obtain higher profits (from more accurate forecasts). The idea was to bridge the literature on intertemporal (strategic) asset allocation with the literature on heterogeneous beliefs.

In Chapters 3 and 4[1] , I consider the empirical problem of implementing Markowitz (1952) mean-variance optimisation on a portfolio of stocks. In particular, I focused on the out-of-sample performance of the minimum-variance portfolio obtained from the use of its asset group (class) information and regularisation methods to obtain more stable estimates of the parameters in the model. Chapter 3 introduces the discussion of regularisation methods applied to portfolio selection problems and the motivation for the ideas in Chapter 4 is also introduced.

In Chapter 4, I propose two regularization methods: first, I use the group structure information from the data without explicitly regularising the solution. I apply a simple 2-step procedure: in the first step, the assets are split into classes, and the within-class minimum-variance portfolios are found. In the second step, I use the portfolios obtained in the first step as assets for a second optimisation across classes. I show that this procedure produces significantly better results in the portfolios than the unconstrained minimum-variance portfolios that are estimated from the whole data set in terms of portfolio variance and the Sharpe ratios. Later, I show that the 2-step procedure could be interpreted as a regularisation (or shrinkage) operation of the

[1] Both Chapters 3 and 4 are based on collaborative work with Marcelo Fernandes and Guilherme Rocha.

covariance matrix of the returns.

For the second approach, I use a more sophisticated econometric procedure to include asset-grouping information in an explicit regularisation framework. With this approach, assets are assigned to a certain group (economy sectors), and the obtained portfolios are regularised towards portfolios with a reasonable record of out-of-sample performance in the literature, such as the equally weighted portfolio, or to the portfolio with constrained short sales. This approach entailed a single step using the results that were originally applied to express the group structure among Ordinary Least Square OLS regressors.

1.2 Motivation

Asset pricing and portfolio selection problems face a special challenge in that data are not generated by experiments; instead, they are obtained naturally. Therefore, researchers cannot control the amount, shocks, or other features of the data. As stated by Campbell et al. (1997), what distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation. Therefore, random fluctuations that require the use of statistical theory to estimate and test financial models are intimately related to the uncertainty upon which those models are based.

We start our discussion from the well-established paradigm based on no-arbitrage arguments. From a theoretical perspective, the paradigm can be summarised by a stochastic discount factor (SDF) that prices all assets in the economy and all its consequences for asset-pricing models. Among these consequences, the challenges in obtaining equilibrium asset-pricing models that conciliate the momentum and the value effects.

From an empirical perspective, the problem is related to implementing the results from the theory to real data. In this field, however, problems start to appear at a much earlier stage, at the estimation step. Implementing the mean-variance optimisation of Markowitz (1952), which simply captures the relationship between risk-return and the effects of diversification, is already problematic.

As previously mentioned, the two features are intrinsically related. In the subsequent sections, we show that the frontier between econometric and theoretical modelling could be very difficult to define, especially for implementing the optimisation method in Markowitz (1952). We note, for instance, that restricting the weights of the Markowitz (1952) portfolio improved its out-of-sample performance. We examine this operation as an econometric procedure and attempted to obtain better finite sample properties for the estimations. However, interpreting this as a new empirical model for portfolio selection would also be possible.

Therefore, the idea that portfolio optimisation is divided into two steps (i.e., in the first step, the investor learns about the generating process of the returns, and in the second step, he or she uses this information in choosing his or her portfolio) and that his or her research is only "concerned with the second stage" in Markowitz (1952) may not be completely accurate. This appears to be the case at least when the same uncertainty in the data used to validate the model is also present in the data used to generate the predictions, as is normally the case.

After understanding the fact that theory and practice are so intimately related, especially in this field, we address both empirical and theoretical questions in this work.

1.3 The general asset pricing theory

As previously mentioned, most of the recent research on asset pricing can be cast into an SDF framework, as in Campbell (2000). The most basic equation of asset pricing is written as:

$$P_{it} = E_t[M_{t+1}X_{t+1}], \tag{1.3.1}$$

where P_{it} the price of asset i in time t, $E_t[.]$ is the expectation operator, conditioned on the information available at time t, X_{t+1} is the realisation of the payoff of asset i in time $t+1$, and M_{t+1} is the "stochastic discount factor" (SDF), a random variable that prices all future payoffs in current terms. Realisations of the SDF are always positive given that a set of positive state prices exist, which is a condition fulfilled by no arbitrage. The SDF will be unique if markets are complete.

The SDF is intrinsically related to uncertainty; in the absence of uncertainty, the SDF is simply a constant that discounts values from the future to the present. Because prices in time t are fixed, the SDF also measures risk.

It is easier to understand the nature of the SDF if we consider the optimisation problem of an agent with the time-separable utility. In this case, the first-order condition of his or her utility maximisation problem will give

$$U'(C_t)P_{it} = \delta E_t[U'(C_{t+1})X_{i,t+1}], \tag{1.3.2}$$

where the marginal cost of an extra unit of asset i is equal do the expected marginal utility of the extra payoff in time $t+1$. We can rearrange the equation as:

$$P_{it} = E_t[\delta \frac{U'(C_{t+1})}{U'(C_t)} X_{i,t+1}], \tag{1.3.3}$$

such that the SDF is given by

$$M_{t+1} = \delta \frac{U'(C_{t+1})}{U'(C_t)}. \tag{1.3.4}$$

This equation illustrates the fact that the SDF is a random variable because C_{t+1} is unknown in time t. It may be interesting to note that the realisation of C_{t+1} in this case is related to the return of the whole portfolio of the individual. Because this return is uncertain, C_{t+1} and M_{t+1} are also uncertain. The volatility of the SDF in this case is related to that of $U\prime(C_{t+1})$.

Risk enters the equation because marginal utility is a decreasing function. Therefore, assets that pay relatively high values when consumption is high (i.e., assets that have a positive covariance with the SDF) will have a lower price than assets with lower covariance with the SDF (i.e., future consumption) for the same expected payoff.

The SDF can also be represented in terms of returns. We can define (gross) returns as

$$(1 + R_{i,t+1}) = \frac{X_{i,t+1}}{P_{it}}, \tag{1.3.5}$$

in this case, we can write

$$1 = E_t[M_{t+1}(1 + R_{i,t+1})]. \tag{1.3.6}$$

Campbell (2000) lists some of the possible uses for the equation above. For instance, it is possible to examine the implications of SDF from the data on the mean, variance and predictability of asset returns. Examining the properties of the returns results in the well known "equity premium puzzle" of Mehra and Prescott (1985), which basically states that the variance of the SDF should be much larger than what can be reasonably assumed for the model to correctly precify risky assets.

This simple case illustrates the situation with a representative agent. Because all agents are equal, the SDF of each agent is also an SDF for the entire (aggregate) economy. We relaxed this assumption in Chapter 2, in which agents can differ in terms of beliefs (the expectation in equation 1.3.3).

1.3.1 Equity premium puzzle

First, note that the implied mean of the SDF can be obtained from equation (1.3.1) because it applies for all assets (including the risk-free asset). Assuming that this risk-free asset pays one unit tomorrow, its mean is

$$P_{it} = E_t[M_{t+1}] = \frac{1}{1 + R_{f,t+1}}. \tag{1.3.7}$$

Given that no truly risk-free asset exists in the economy (in real terms because of inflation risks), we can still use short-term treasury bills as a proxy for a risk-free asset. In this case, the conditional expectation of the SDF is implied to be slightly lower than 1% pa (approximately 0.8% pa for the U.S. as reported in Campbell and Viceira (1999)) and not very volatile (1.76% standard deviation in the same paper).

A second important piece of information from the relationship above is that the risk premium restricts the volatility of the SDF. This can be shown below, where equation (1.3.6) is applied to the risky and risk-free assets to obtain

$$
\begin{aligned}
0 &= E_t[M_{t+1}(R_{i,t+1} - R_{f,t+1})] = \\
&E_t M_{t+1} E_t(R_{i,t+1} - R_{f,t+1}) + Cov_t(M_{t+1}, (R_{i,t+1} - R_{f,t+1})),
\end{aligned} \tag{1.3.8}
$$

which can be rearranged to:

$$
E_t(R_{i,t+1} - R_{f,t+1}) = \frac{-Cov_t(M_{t+1}, (R_{i,t+1} - R_{f,t+1}))}{E_t M_{t+1}}. \tag{1.3.9}
$$

Because the coefficient of correlation must be larger than -1, the negative covariance in the above equation must be smaller than the product of the standard deviations of the excess return and the SDF. This produces

$$
\frac{\sigma_t(M_{t+1})}{E_t M_{t+1}} \geq \frac{E_t(R_{i,t+1} - R_{f,t+1})}{\sigma_t(R_{i,t+1} - R_{f,t+1})}, \tag{1.3.10}
$$

where the Sharpe ratio for asset on the right-hand side of the equation puts a lower bound on the volatility of the SDF. The largest lower bound is found by the asset with the highest Sharpe ratio; therefore, this bound is the most restrictive.

The equity premium puzzle arises when we consider assets that imply a very large number for the left-hand side in equation (1.3.10). We may consider the aggregate U.S. stock market as an example. Campbell and Viceira (1999) showed that the annualised Sharpe ratio for a value-weighted stock index is approximately 0.5, implying a minimum of 50% annualised standard deviation on the SDF. This is a random variable that is always positive with a mean that is slightly lower than 1.

These extreme values only worsen when we use equilibrium models to derive the SDF while looking at marginal utilities as in equation (1.3.3).). In a representative agent model with power utility, the coefficient of relative risk aversion must be on the order of 50 to match the volatility of the SDF, whereas typical values are less than 5. The key problem in this case is that variations in the SDF are driven by changes in aggregate consumption, which is rather stable over time. To conciliate these two facts, investors would need to be extremely risk-averse.

Several other tentative explanations for this phenomenon also exist. For instance, one of them is the uncertainty in the moments that enter equation (1.3.10).Another one is the existence of "Peso problems", i.e., some catastrophic event that is rationally reflected in stock prices or even the use of U.S. stock market, given the selection bias of a case that "worked". In Chapter 2, we propose that these changes were actually driven by consumption and changes in beliefs. This added another source of variation to the SDF and helps to conciliate high returns, low consumption volatility and the typical levels of risk aversion.

The long-documented predictability of returns increases the equity premium puzzle. This happens because this predictability it allows the creation of a managed portfolio with a Sharpe ratio higher than that of the market and increases the lower bound of the SDF volatility. For instance, Campbell (2000) noted that an extensive amount of the literature documents the predictability of aggregate stock returns from past information. This information includes the following: lagged returns (Fama and French (1988b), Poterba Lawrence and James (1988)), the dividend-to-price ratio (Campbell and Shiller (1988a), Fama and French (1988a)), the earnings-to-price ratio (Campbell and Shiller (1988b)), the book-to-market ratio (Lewellen (1999)), the dividend payout ratio (Lamont (1998)), the share of equity in new finance (Baker and Wurgler (2000)), yield spreads between long-term and short-term interest rates and between low- and high-quality bond yields (Campbell (1987), Fama and Kenneth (1989), Keim and Stambaugh (1986)), recent changes in short-term interest rates (Campbell (1987), Hodrick (1992)), and the level of consumption relative to income and wealth (Ludvigson and Lettau (1999)). Many of these variables are related to the stage of the business cycle and predict countercyclical variation in stock returns (Fama and Kenneth (1989),Ludvigson and Lettau (1999)).

1.3.2 Factor models and the cross-sectional of returns: Value, size and momentum effects

Multi-factor models, such as the ones used in Chapter 2, can also be cast into the SDF framework. The general idea was to model the SDF as a linear function of K factors; then, the risk premium will be the sum of the asset's loads on those factors multiplied by the risk prices of the factors.

Considering a mean-variance optimisation, for instance, the single period optimisation consumption equals wealth; with quadratic utility, the marginal utility is linear. In this case, the SDF must be a linear function of future wealth, i.e., it should be linear in the market portfolio return. If there are K common shocks and there is completely diversifiable risk, then the SDF can depend only on the K common shocks.

Factor models can also be used to describe the behaviour of asset returns atheoretically while choosing factors to fit the empirical evidence. The three-factor model of Fama and French (1993) is an example of such approach, while Carhart (1997) extends it to account for a momentum factor.The following three effects are the most commonly documented: the size effect, by which firms with small market value tend to have higher returns than what is predicted by the CAPM; the value effect, by which fundamentalist ratios are capable of forecasting future returns (also associated with the mean reversion of De Bondt and Thaler (1985), where stocks with previous bad performances in the last three to five years tend to outperform in the future)); and the momentum effect of Jegadeesh and Titman (1993), where stocks with high returns over the last three to twelve months tend to outperform in the future.

Assuming that these anomalies are not the result of mispecified tests, there are some tentative explanations for the above effects . However, none of these explanations is conclusive. Fama and French (1993) and Fama and French (1996), for instance, interpret their factor model as evidence of a "distress factor" without explaining why this occurs in the first place. In fact, models with time-varying discount rates will be successful in generating the value effect by a simple fact: stocks with high discount rates will have lower prices and higher future returns regardless of the reason why these discount rates are high in the first place. The momentum effect, on the other hand, is much harder to generate. Fama and French (1996), for instance, claimed that it might be the result of data mining and did not attempt to model it. Even behavioural models had difficulties explaining this effect because momentum is consistent with the slow reaction to news but is difficult to reconciliate with the subsequent over reaction that leads the value effect.

In Chapter 2, we assume that investors use factor models to predict returns. Fundamentalist strategies are then based on fundamentalist factors[2] while chartist (i.e., momentum) strategies are based on past returns.

---

[2] In the empirical section, we used the dividend-price ratio as a return forecaster, but any other fundamentalist ratio could be used in practice.

1.3.3 Strategic asset allocation: Consumption and portfolio choice for long-term investors

Before addressing the intertemporal case, we consider the one-period optimisation. Assume that an investor lives off financial wealth alone and only for one period. If the investor derives increasing utility from consumption, the investor's future consumption will be the payoff of his portfolio: $C_{t+1} = X_{t+1}$. From equation (1.3.4), setting $\theta_t = \delta \frac{1}{U'(C_t)}$, which is known at time t, then

$$M_{t+1} = \theta_t U'(C_{t+1}) = \theta_t U'(X_{t+1}). \tag{1.3.11}$$

which implies

$$X_{t+1} = {U'}^{-1}(M_{t+1}/\theta_t). \tag{1.3.12}$$

The Markowitz (1952) mean variance portfolio can be obtained, allowing $U(.)$ to be quadratic. In this case $U'(.)$ is linear, and the result is a linear trade-off between mean and variance of returns. In the particular case where M_{t+1} is linear in the returns of a market portfolio (as is the case in a CAPM framework), this investor holds a portfolio that consists of the risk-free asset and the market portfolio. It is also possible to examine the equity premium puzzle from this perspective: if M_{t+1} is highly volatile, then X_{t+1} also needs to be highly volatile unless the investor is very risk-averse, meaning that $U(.)$ is very concave, with $U''(.)$ being very negative and $U'(.)$ declining very rapidly.

The problem is that this "myopic" and intuitive solution does not necessarily correspond to optimality for the intertemporal case. In fact, only a few very restrictive situations exist in which the solution for both problems will be the same. As Merton (1969) and Merton (1971) show, investors do not only care not only about wealth but also about shocks to these investment opportunities (i.e., the productivity of wealth) in a long horizon framework with time-varying investment opportunities. The term "strategic asset allocation" from Brennan et al. (1997) describes the long-term investors' hedging against these shocks.

However, intertemporal models are very difficult to solve analytically. Very few special cases have closed-form solutions, and the remaining ones rely on either numerical methods or approximate solutions from perturbations of known exact solutions as the ones used in Chapter 2.

Another important decision involves which utility function should be used to describe the investor's preferences. Many utility functions, such as the quadratic function, have implications that go against empirically stylised facts. Conciliating constant risk premia and interest rates with the upward trend in consumption observed in the last century in the U.S. is one of these problems. One example of a utility with good empirical features is the generalisation of the power utility proposed by Epstein and Zin (1989), Epstein and Zin (1991) and Weil (1989) which are discussed in detail in Chapter 2. The main advantage of this utility function over the power utility is that it separates the coefficient of relative risk aversion, which is meaningful even in an atemporal framework, from the elasticity of intertemporal substitution of consumption, which is meaningful even in the absence of risk.

Investment opportunities can vary in time because real interest rates vary, in which case the present value of the portfolio also varies, and because risk premia vary. There is evidence of both in the literature , increasing the difference between the myopic and intertemporal solutions.

For instance, it is often argued that stocks are safer for long-term investors. However, this cannot be true if asset returns are IID because it means that the means and variances of all assets increase with time. This situation can only be supported if stock returns are predictable and that the variance of stock returns increases less than the variance would proportionally increase with time. This is normally called mean reversion, and it implies that investment opportunities vary in time.

One important problem to reconcile time-varying opportunities and a representative agent framework is the following: investors are supposed to time the market, altering their allocations of stocks as conditions change. However, this cannot happen in a representative agent framework that is in general equilibrium because not all investors can buy or sell stocks at the same time given a fixed supply of stocks. As we see in Chapter 2, this is not a problem in a heterogeneous agent framework.

1.3.4 Equilibrium models with a representative agent

In an equilibrium model with a representative agent, it is possible to obtain the corresponding SDF from the specified utility function and stream of consumption. In Lucas (1978) it is assumed that the economy could be described by a representative agent with a standard utility function that consumes aggregate consumption. In this case, the SDF could be obtained from the first-order condition in the utility maximisation problem. However, this approach resulted in three puzzles.

The first and most important one is the equity premium puzzle discussed above, which arose because consumption growth is very smooth over time. Thus, the covariance could never be large regardless of how highly correlated it is with the assets' returns.

The second puzzle is that the volatility of stock returns was too great to be explained by traditional models. This was because stock returns are driven by shocks on consumption growth via the SDF, which affects the expected future dividends and discount rates. However, unexpected consumption growth is, once again, too small to justify the volatility of stock returns (i.e., the volatility of the discount rate/SDF).

The third puzzle is the risk-free rate puzzle that occurs if a power utility function is used. This puzzle occurrs because the increased risk-aversion coefficient required to solve the equity premium puzzle made the elasticity of intertemporal substitution (its inverse) very small. This implied a preference for a constant (or close to constant) stream of consumption. The only way to conciliate this preference for constant consumption with the empirical evidence of historical upward consumption would be to use a very low or even negative rate of time preference or a very high real interest rate, which does not occur in reality.

The last puzzle could be solved with the help of Epstein-Zin utility functions. However, the other two puzzles could not be easily solved. For the equity premium, the best that could be done was to assume that risk aversion is actually much higher than what is normally accepted. The volatility puzzle is driven by an actual change in the equity premium over time because real interest rates (the other possible source for this variation) are too stable to explain those swings.

One tentative method that could be used to generate time-varying risk premia is to model the utility itself and explain that changes in the equity premium occur because of certain features of the utility function. For instance, the utility in some habit-formation models is dependent on time or, more precisely, on consumption history. The present increase in consumption makes agents more willing to consume in the future (for habit formation), increasing the marginal utility of future consumption. This change is enough to generate the fluctuations in the real interest rate when it is applied to a representative agent, which solves the volatility puzzle. However, it does not solve the equity premium puzzle because large risk aversion is still required to explain the excess returns of stocks.

1.3.5 Equilibrium models with heterogeneous agents

In an attempt to explain the puzzles listed in the previous section, one alternative is to use models with heterogeneous agents. A type of heterogeneity is on the constraints to which agents are subject. The idea is that not all investors participate in the stock market; therefore, the relevant consumption is not the aggregate consumption but only part of it. Consumption by constrained agents (that do not participate in the stock market) is irrelevant. Evidence (e.g., Brav et al. (2002)) also shows that consumption by stockholders is more volatile than that by non-stockholders. The relevant consumption becomes unobservable if one wants to employ a representative agent formulation.

A second alternative is to model heterogeneous income constraints. The idea is that, in an incomplete market, individuals may have very different consumption paths. Any individual consumption growth would generate a valid SDF, but the same may not be true regarding aggregate consumption. In addition, these models also have problems in solving the asset-pricing puzzles. For instance, in the Constantinides and Duffie (1996) model, heterogeneity should be very large to have significant effects on the SDF.

It is also possible to model heterogeneity in preferences. Different degrees of risk aversion or time preferences among investors may lead to time-varying risk price. For instance, risk-tolerant investors hold more risky assets and control a larger share of wealth in good states than in bad states. This makes aggregate risk aversion increase during bad times, just as in habit-formation models.

A final source of heterogeneity is the heterogeneity of beliefs. In these models, agents model future returns based on different strategies. One alternative is to assume that the agents choose these strategies based on their previous performance (fitness). Each agent type has a different SDF implied by the strategy used. The equilibrium asset price (and aggregate SDF) is given endogenously. Heterogeneity contributes to the volatility of the aggregate SDF because it is affected by the (changing) proportion of agents in the economy. Even if the higher Sharpe ratio obtained using this managed portfolio increases the lower bound of the SDF volatility, it would not be a problem because this volatility and the change in beliefs are linked. This differs from the representative agent approach, in which the only source of variation in the SDF comes from changes in aggregate consumption. We presented and discussed a model of this type in Chapter 2.

1.4 The mean-variance optimization

As discussed earlier, the mean-variance portfolio can be cast into an SDF framework by assuming that an investor with quadratic utility function lives off financial wealth alone and only for one period. These strong and restrictive assumptions give rise to the extensions described before. However, the mean-variance optimisation in Markowitz (1952) is still by far the best-known formulation of portfolio choice problems. Its closed-form solution and intuitive results have features attractive to both academics and practitioners. The most important feature of this model is that the model captures the effect of diversification and the positive association between risk and expected returns. The model intrinsically describes a short-term condition because it relies on a single-period optimisation.

1.4.1 Clarification on notation

From one period to the next, investors need to choose how to allocate their wealth among the N risky assets in the economy. We stacked these N assets' returns between time t and $t+1$ in the $N \times 1$ vector R_{t+1}. Expected returns that are conditional on the information in time t, are given by

$$E_t\left[R_{t+1}\right] = \mu_t, \tag{1.4.1}$$

while the conditional covariance matrix is given by

$$E_t\left[(R_{t+1}-\mu_t)(R_{t+1}-\mu_t)\prime\right]=\Sigma_t. \tag{1.4.2}$$

The goal was to find a $N \times 1$ vector of portfolio weights x, where each element in x was the proportion of wealth allocated to each of the N available assets. If the investor's wealth must be fully invested, then

$$x\prime\iota = 1, \tag{1.4.3}$$

where ι is a $N \times 1$ vector of ones. The return of the portfolio is then given by

$$R_{p,t+1} = x\prime R_{t+1}$$

and we obtain the conditional expected return and covariance matrix of the portfolio respectively:

$$E_t\left[R_{p,t+1}\right] = x\prime\mu_t, \tag{1.4.4}$$

and

$$Var_t[R_{p,t+1}] = x\prime\Sigma_t x. \tag{1.4.5}$$

1.4.2 The optimization problem

In the formulation developed by Markowitz (1952), investors associate a positive utility to expected returns and a negative one to expected variance. Thus, the optimization problem could be solved in two ways: either choosing a portfolio that minimises the variance for a given level of expected return or maximises the expected utility incorporating the trade-off between expected returns and variance given by the investor's attitude towards risk. The first formulation is very useful when we do not want to make any assumption on how investors trade variance and expected returns because we can map any combination of risk and return that is available to the investor.

In this framework, we set the minimal expected return to, for instance, $\bar{\mu}$, and find the portfolio that produces this return: $R_{p,t+1} = x' R_{t+1}$, with the smallest variance. As different values for $\bar{\mu}$ are selected, all of the optimal combinations of risk and return (the so-called efficient frontier of available securities) are recovered.

The problem, in this case, becomes

$$
\begin{aligned}
x_M^*(\bar{\mu}, \Sigma_t, \mu_t) \quad = \quad & \arg\min_x \quad x^T \Sigma_t x \\
& \text{s.t.} \qquad x^T \mu_t = \bar{\mu} \\
& \qquad\qquad x^T \iota = 1.
\end{aligned}
\tag{1.4.6}
$$

When R_{t+1} represents a vector of excess returns of the form $R_{t+1} = R_{f,t+1} + \mu$, the first order conditions obtained using the Lagrangian produces the optimal portfolio weights:

$$x^* = \Lambda_1 + \Lambda_2 \bar{\mu}, \tag{1.4.7}$$

with

$$\begin{aligned} \Lambda_1 &= \tfrac{1}{D}[B\left(\Sigma_t^{-1}\iota\right) - A(\Sigma_t^{-1}\mu)], \\ \Lambda_2 &= \tfrac{1}{D}[C\left(\Sigma_t^{-1}\mu\right) - A(\Sigma_t^{-1}\iota)], \end{aligned} \tag{1.4.8}$$

and

$$\begin{aligned} A &= \iota^T \Sigma_t^{-1} \mu, \\ B &= \mu^T \Sigma_t^{-1} \mu, \\ C &= \iota^T \Sigma_t^{-1} \iota, \text{ and} \\ D &= BC - A^2. \end{aligned}$$

The minimized variance is $Var[R^*_{p,t+1}] = x^{*\prime}\Sigma_t x^*$.

In this step, we see the two key features of the model. Diversification plays a central role because it is possible to obtain lower portfolio variances that combine assets with less-than-perfectly-correlated returns for a given level of expected return. The other feature, the risk-return trade-off notes that x^* is linear in $\bar{\mu}$. Therefore, higher expected returns could only be obtained through riskier (more extreme) allocations.

1.4.3 The minimum-variance portfolio

The problem of finding the minimum-variance portfolio is equivalent to the previous problem but without the restriction of expected returns.

When an investor is looking for the minimum-variance portfolio in the absence of any constraints, the problem is to find

$$\begin{aligned} x^*_{MINU}(\Sigma_t) &= \arg\min_x \quad x^T \Sigma_t x \\ & \quad \text{s.t.} \qquad x^T \iota = 1. \end{aligned} \tag{1.4.9}$$

where $x \in \mathbb{R}^N$ is the vector of portfolio weights, $\Sigma_t \in \mathbb{R}^{N \times N}$ is the estimated covariance matrix and $\iota \in \mathbb{R}^N$ is a vector of ones. The constraint assured that the portfolio weights added up to one.

Apart from the issues of optimality under broader conditions, as described earlier, the theory up to this point has not often been challenged . However, there are still problems in connecting the theoretical solution to the data even assuming that all of the conditions for the optimality of this solution would hold (e.g., one-period investment horizon, quadratic utility and only financial wealth). The tentatives of obtaining mean variance efficient portfolios often end up generating highly concentrated portfolios that perform poorly out of sample. This problem is due to the finite sample error in estimating the parameters of the model that is associated with the sensitivity of the solution to these parameters. The optimization problem is often referred to as ill-posed (or ill-conditioned) given that the obtained parameters depend extremely heavily on the sample used to estimate them.

In Chapter 3, we discuss some of the econometric procedures used to realistically implement the model and work our way around the issues mentioned before. We pay close attention to the finite sample properties of these estimators and propose regularisation methods to improve the estimates. The general idea of these methods is that, in finite samples, it is possible to improve the performance of estimators (e.g., in terms of mean squared errors) by averaging them with a constant. This is obtained by introducing some bias into the result (given by the constant) but reducing the final variance of the estimate.

Chapter 2

Strategic asset allocation with heterogeneous beliefs

In this chapter, I show how the presence of agents with heterogeneous beliefs generates the price trends observed in the financial markets. I develop an asset pricing model in which agents have long horizon objectives, based on a stream of consumption. Each agent chooses a forecasting model and maximises a recursive utility function. The choice of the forecasting model in each period determines the agent type. However their types change over time according to the relative performance of the forecasting models. This happens because agents have an incentive to adopt the forecasting model with the best performance in the previous period to coordinate with the market. I estimate the asset pricing model using data on the international stock markets.

The exercise shows that especially for very risk averse individuals, the accounting for the intertemporal hedging demand is crucial.

2.1 Introduction

This paper bridges the literatures on intertemporal asset allocation and on heterogeneous beliefs. From the intertemporal asset allocation framework, the asset pricing model inherits the ability to reproduce the behaviour of consumption-based utility maximizing investors with long horizon objective functions. I solve the intertemporal asset allocation problem introduced by Merton (1969) and Samuelson (1969) using the approximate solution of Campbell et al. (2003). I use the class of preferences in Epstein and Zin (1989, 1991) and Weil (1989) in order to individuate the agent's risk aversion and elasticity of intertemporal substitution. This framework is convenient because it allows to solve the portfolio selection problem in the presence of multiple risky assets. This is in contrast with the usual myopic mean-variance framework with a single risky asset in the literature on heterogeneous beliefs (e.g., Brock and Hommes (1997, 1998) or Boswijk et al. (2007)), providing an alternative to the multiple risky assets formulation of Wenzelburger (2004).

By assuming heterogeneous beliefs I am able to better describe the individual and market behaviours and, as such, reproduce the stylized facts of asset returns. There are many attempts in the literature to reproduce these effects. Nevertheless, none of them is able to fully solve all the puzzles and explain the momentum and value effects at the same time. The biggest challenges in most cases are solving the equity premium puzzle and generating the momentum effect. Models with adaptive heterogeneous beliefs are able to conciliate momentum and value effects as well as to generate higher volatility in returns. These are achieved because the agents change beliefs (i.e., forecasting models) over time.

Heterogeneous beliefs models alleviate the equity premium puzzle. Abel (1989) notes that heterogeneity *per se* does not necessarily invalidate the representative agent approach, but heterogeneity in *beliefs* does. This happens because the cross-sectional distribution of expectations cannot be summarized by a single sufficient statistics. Abel (1989) also shows that introducing heterogeneity in beliefs can substantially increase the equity premium (see also Basak (2005) and Kurz and Beltratti (1996))

The formulation matches several theoretical and empirical evidence of heterogeneity of expectations. It also matches the evidence of variability over time in the choice of forecasting models as in Frankel and Froot (1987) for instance. In addition, the assumption of heterogeneous agents avoids a no-trade equilibrium that arises as a consequence of theorems such as those in Milgrom and Stokey (1982).

Kandel and Pearson (1995) and Bamber et al. (1999) provide evidence of heterogeneity in analyst expectations for stocks regarding earnings around announcements. Analysing bubbles, Shiller (2002) provides evidence of heterogeneity in the expectations of the future performance of the market. Finally, Patton and Timmermann (2010) study the sources of disagreement about forecasts of macroeconomic variables and find that they are persistent and indicate that they stem from heterogeneity in priors or models, not different information sets.

Frankel and Froot (1987) and Taylor and Allen (1992) report survey evidence of heterogeneity in expectations. In particular, Frankel and Froot (1987) find that forecasting companies use different models to project returns, and that the number of companies using different classes of models changes over time. Surveying exchange rate expectations of financial specialists, Menkhoff (1997) shows that investors tend to use different trading strategies. Their strategy choice depends on the investment horizon they are trying to forecast. They basically use chartist strategies in the short run, and keep fundamentalist strategies for long horizons.

In this paper, I use the approximate solution of Campbell et al. (2003) to calculate the demands for assets of each agent type, and apply the framework of Brock and Hommes (1997, 1998) and Boswijk et al. (2007) to model the evolution of types. Modelling the evolution of types corresponds to describing how the proportions of agents using a given return forecasting model evolve over time. Therefore, I extend the models of Brock and Hommes (1997, 1998) and Boswijk et al. (2007) to consider also long term investors.

The agents adjust their forecasts trying to match what they believe to be the dominant forecasting strategy in the market. I assume that the market is populated by many agents choosing among different forecasting models. These agents are aware that other agents are also choosing their models in the same way. Therefore, they know that the most accurate forecast is the one given by the strategy chosen by the majority of them. This happens regardless of the theoretical support that a given model may enjoy. They make this choice in each period and this determines their types. However, they do not receive perfect information regarding the performance of the strategies.

In the empirical section, I assume the perspective of an investor in the U.S.A. who would like to diversify his/her portfolio using the international stock markets. For simplicity, I assume that there are only two agent types in each market: fundamentalists and chartists. Fundamentalists use value strategies and chartists use momentum strategies. I estimate the model using stock market data from the U.S., the U.K., Japan, and Hong Kong. I start by estimating a simple dividend-price factor model and a simple momentum factor model for each of these four markets. Next, I use these factor models as the fundamentalist and chartist strategies and examine the resulting dynamics. Given these forecasting models, I obtain the demand for assets of each agent type. I use these demands to compute the relative performances of their strategies and this determines the fraction of agents using each strategy.

I show that myopic and long-term investors have different demands for assets and, therefore, different performances. I also show that the investment horizon has different effects on the demand for assets of fundamentalists and chartists. The component of the demand for assets that is ignored in a myopic framework can be significantly large and impact the estimation of the proportions of agents. This is especially true when agents are very risk averse. In fact, in this case the omitted term in the myopic framework can be the dominant one in certain markets. Therefore, the agent's decision of using a fundamentalist or chartist forecast in these markets will often depend on whether he/she believes that agents are myopic or not. In addition, I show that the level of noise in the observed performances also has different impacts on the model results whether we consider the complete intertemporal demand for assets or only its myopic component.

The paper is organised in three sections following this introduction. In Section 1, I derive the asset-pricing model and discuss its theoretical results. In Section 2, I estimate the model and analyse the results focusing on the differences between the intertemporal solution and the myopic solution previously obtained in the literature. Section 3 concludes.

2.2 The model

There is an infinite number of long-term investors of H different types. The trading strategy used to forecast returns determines the agent type h. In most of the paper, I restrict the analysis to $H = 2$ (i.e. fundamentalist or chartist types), but I develop the model for the general case with a given number of types H. Agents extract information from prices: they switch between trading strategies (change their types) as they respond to the previous performance of the strategies. However, they do not receive perfect information regarding the performance of the strategies. They all have access to the same information set but use different return forecasting models. Therefore, I model differences in opinions (i.e., forecasting models), and not differences in information sets.

2.2.1 The investor's maximization problem

Time is discrete, and investors that live infinitely maximise the recursive preferences defined over a stream of consumption, as described by Epstein and Zin (1989, 1991) and Weil (1989).[1] There are n risky assets

[1] The power utility is a special case of the Epstein-Zin function. We can obtain it by letting $\gamma = \psi^{-1}$ (and hence $\theta = 1$). In addition, the log utility is a special case of the power utility, it can be easily obtained by adding the restriction $\gamma = 1 = \psi^{-1}$.

in the economy, and investors allocate their wealth among these assets and consumption. The investor's problem is to choose the portfolio allocation, $\alpha^*_{h,t}$, and consumption, $C^*_{h,t}$, that maximises his/her utility at every time t given his/her type. Each investor is, however, restricted by a budget constraint.[2] So, their problem is given by:

$$
\begin{aligned}
(\alpha^*_{h,t}, C^*_{h,t}) \quad & = \\
\arg \max_{\alpha_{h,t} \in \mathbb{R}^n, C_{h,t} \in \mathbb{R}} \quad & U_t(C_t, E_t[U_{t+1}]) = \left[(1-\delta) C_t^{\frac{1-\gamma}{\theta}} + \delta (E_t(U_{t+1}^{1-\gamma}))^{\frac{1}{\theta}}\right]^{\frac{\theta}{1-\delta}} \\
\text{s.t.} \quad & W_{t+1} = (W_t - C_t)(1 + R_{p,t+1}), \\
& R_{p,t+1} = \textstyle\sum_{i=2}^{n} \alpha_{h,i,t}(R_{i,t+1} - R_{1,t+1}) + R_{1,t+1}.
\end{aligned}
\tag{2.2.1}
$$

where C_t is the agent's consumption and $E_t(.)$ is the agent's conditional expectation operator at time t. The agent's relative risk aversion coefficient is $\gamma > 0$, $\psi > 0$ is the agent's elasticity of intertemporal substitution coefficient, $0 < \delta < 1$ is the agent's time discount factor and $\theta \equiv (1-\gamma)/(1-\psi^{-1})$. In the consumption-based budget constraint, W_t is wealth at time t, and $R_{p,t+1}$ is the portfolio return on the next period. Finally, $\alpha_{h,i,t}$ is the portfolio weight on asset i at time t and $R_{i,t+1}$ is the return on the next period. The first asset ($i = 1$) is proxy for a risk free asset with a real return of $R_{1,t+1}$.

With time varying investment opportunities, this condition generates the myopic portfolio allocation. However, as Giovannini and Weil (1989) showed, $\gamma = 1$ or $\psi^{-1} = 1$ alone are not sufficient for this result.

[2] Because the maximisation problem is the same for every agent type, I do not write the subscripts here to simplify the notation.

Epstein-Zin preferences and agent-based models

The class of utility functions in Epstein and Zin (1989, 1991) and Weil (1989) represent intertemporal preferences. The utility at time t, $U_t(C_t, E_t[U_{t+1}])$, depends on consumption at time t, C_t, and also on the expected utility in time $t+1$, given by $E_t[U_{t+1}]$. Recursively substituting future expected utilities highlights the intertemporal characteristics of this class of preferences.

In order to gain intuition, we can look at the following equations:

$$U_t(C_t, E_t[U_{t+1}]) = \left[(1-\delta)C_t^{\frac{1-\gamma}{\theta}} + \delta(E_t(U_{t+1}^{1-\gamma}))^{\frac{1}{\theta}}\right]^{\frac{\theta}{1-\delta}}, \tag{2.2.2}$$

$$U_{t+1}(C_{t+1}, E_{t+1}[U_{t+2}]) = \left[(1-\delta)C_{t+1}^{\frac{1-\gamma}{\theta}} + \delta(E_{t+1}(U_{t+2}^{1-\gamma}))^{\frac{1}{\theta}}\right]^{\frac{\theta}{1-\delta}}. \tag{2.2.3}$$

The extended form of equation (2.2.2) is obtained by simply substituting equation (2.2.3) into equation (2.2.2):

$$U_t(C_t, E_t[U_{t+1}]) = \left[\begin{array}{c} (1-\delta)C_t^{\frac{1-\gamma}{\theta}} + \\ \delta(E_t\left\{\left[(1-\delta)C_{t+1}^{\frac{1-\gamma}{\theta}} + \delta(E_{t+1}(U_{t+2}^{1-\gamma}))^{\frac{1}{\theta}}\right]^{\frac{\theta}{1-\delta}}\right\}^{1-\gamma})^{\frac{1}{\theta}} \end{array}\right]^{\frac{\theta}{1-\delta}}, \tag{2.2.4}$$

where the part inside the curly brackets is the utility at time $t+1$, $U_{t+1}(C_{t+1}, E_{t+1}[U_{t+2}])$. The equation shows that the utility in time t depends on consumption in time t and also in time $t+1$. Continuing with the recursive substitution, it is possible to show that the utility being maximised in fact depends on an infinity stream of consumption. This is what is responsible for the intertemporal maximization characteristics of the model.

The use of Epstein and Zin (1989, 1991) and Weil (1989) preferences, together with the usual constraints, brings new insights to the literature on heterogeneous beliefs models because it addresses a multi-period optimization with intermediate consumption. This is in line with the models of Merton (1969) and Samuelson (1969). The most important characteristics of this utility function is that it disentangles the elasticity of intertemporal substitution, that is relevant even in the absence of risk, from the coefficient of relative risk aversion, that is meaningful even in an atemporal formulation.

As given in the formulation of the problem, each agent solves an infinity intertemporal optimization. Therefore, all agents in the economy have long investment horizons, and maximize a stream of consumption. This is particularly new with respect to heterogeneous beliefs models with evolutionary selection of expectations such as those introduced in Brock and Hommes (1997, 1998). In these models, agents are either assumed to be myopic or to maximise their utility at a given point in the future.

The intertemporal optimization contrasts, therefore, with the myopic one (in one or multiple periods). Examples are the heterogeneous

CAPM of Chiarella et al. (2006), the overlapping generations (OLG) model with heterogeneous beliefs in Böhm and Chiarella (2005) or the mean-variance investors in Horst and Wenzelburger (2008) or Wenzelburger (2004), that can also be seen as an extension of Brock and Hommes (1997, 1998) with multiple types of agents and risky assets. The intertemporal optimization with intermediate consumption using the Epstein and Zin (1989, 1991) and Weil (1989) preferences also contrasts with multi-period final wealth optimizations obtained, for instance, in the extension of Hillebrand and Wenzelburger (2006). This happens because the two formulations generate different demands for assets at every point in time.

One of the most important characteristics of the intertemporal formulation is the possibility of comparing myopic and long-term investors. This happens because the demand for assets is separable into a myopic and an intertemporal hedging demand terms. In the empirical section I extensively compare the estimated results considering one or the other investment horizon.

First order conditions and approximate solution (Chan et al. (2003))

Epstein and Zin (1989, 1991), as noted by Chan et al. (2003), find that solving the problem in (2.2.1) for a single agent results in the Euler equation:

$$E_t\left[\left\{\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\psi}}\right\}^{\theta}(1+R_{p,t+1})^{-(1-\theta)}(1+R_{i,t+1})\right]=1 \qquad (2.2.5)$$

that must hold for any asset i, (including the portfolio p) along the optimum consumption path. The equation shows the relationship between portfolio allocation, consumption and expectations (or *beliefs*). It highlights the importance of the forecasting model used by the agent. The forecasting model impacts both on the planned growth in consumption and also on the asset allocation through the expectation operator.

In general there is no closed form solution to this problem. I, thus, apply the same approximate solution of Campbell et al. (2003) to obtain an expression for the asset's demand of each agent type. I describe the procedure in details in the following subsections. It begins by postulating that agents describe the dynamics of the relevant state variables as a first-order vector auto-regressive process $VAR(1)$. Campbell and Viceira (1999) shows that the approximate solution exists if the elasticity of intertemporal substitution is close to 1. We then log-linearize the portfolio return and budget constraints in (2.2.1) as well as the Euler equation in (2.2.5) close to this value. This step produces an expression for the expected excess return of each asset. Next, we write everything in terms of the state variables in the VAR. Solving for the consumption and portfolio rules, we finally obtain the optimal asset demand for each investor type. See Appendix A for a basic derivation of excess returns using a stochastic discount factor framework.

Dynamics of returns

Formally, I define

$$x_{t+1} = \begin{bmatrix} r_{2,t+1} - r_{1,t+1} \\ r_{3,t+1} - r_{1,t+1} \\ \dots \\ r_{n,t+1} - r_{1,t+1} \end{bmatrix}, \qquad (2.2.6)$$

where $r_{i,t+1} = \ln(1 + R_{i,t+1}) \ \forall \ i$, and x_{t+1} is a vector of excess returns. I also include other state variables s_{t+1}, such as the price-earnings ratio, realised returns or other return forecasters, stacking $r_{1,t+1}$, x_{t+1} and s_{t+1} into an $m \times 1$ vector z_{t+1}:

$$z_{t+1} = \begin{bmatrix} r_{1,t+1} \\ x_{t+1} \\ s_{t+1} \end{bmatrix}. \qquad (2.2.7)$$

A fundamentalist agent will model the market dynamics by considering fundamentalist predictors. Chartists will decide based exclusively on past returns. The difference between them is in the coefficients of the VAR:

$$z_{h,t+1} = \phi_{h,0} + \phi_{h,1} z_t + v_{h,t+1}. \qquad (2.2.8)$$

The trading strategy that agent h is actually using determines the coefficients $\phi_{h,0}$, the $m \times 1$ vector of intercepts, and $\phi_{h,1}$, the $m \times m$ matrix of slope, with shocks $v_{h,t+1}$ that satisfy

$$v_{h,t+1} \sim i.i.d.\ N(0, \Sigma_{h,v}), \tag{2.2.9}$$

$$\Sigma_{h,v} \equiv Var_t(v_{h,t+1}) = \begin{bmatrix} \sigma^2_{h,1} & \sigma'_{h,1x} & \sigma'_{h,1s} \\ \sigma_{h,1x} & \Sigma_{h,xx} & \Sigma'_{h,xs} \\ \sigma_{h,1s} & \Sigma_{h,xs} & \Sigma_{h,ss} \end{bmatrix}. \tag{2.2.10}$$

These distributional assumptions allow for a cross-sectional correlation between the shocks, which are otherwise iid over time.[3] Given the homoskedastic $VAR(1)$ formulation, it is easy to derive the unconditional distribution of z_{t+1} because it inherits the normality of the shocks. Note that unlike Brock and Hommes (1998), I assume that agents may also disagree on how to estimate these variances and covariances.

Approximate solution

Epstein and Zin (1989, 1991) show that it is possible to write the value

[3] The homoskedasticity assumption is rather restrictive because it rules out the possibility that state variables predict changes in risk. This means that they can only affect the portfolio choice by predicting changes in expected returns. However, many previous studies show that the effect of those risk changes over portfolio choice is limited. Campbell (1987), Harvey (1991) and Glosten et al. (1993) found only modest effects that are dominated by the effects of the state variables on expected returns. Also, Chacko and Viceira (2005) show that changes in risk are not persistent enough to have large effects on the intertemporal hedging demand.

function obtained from the maximization in (2.2.1) per unit of wealth as a power function of the optimal consumption-wealth ratio:

$$V_t \equiv \frac{U_t}{W_t} = (1-\delta)^{-\frac{\psi}{1-\psi}} \left(\frac{C_t}{W_t}\right)^{\frac{1}{1-\psi}}. \qquad (2.2.11)$$

Campbell and Viceira (1999) note that, under the assumptions made here,

$$\lim_{\psi \to 1} \frac{C_t}{W_t} = (1-\delta), \qquad (2.2.12)$$

which guarantees that the value function (2.2.11) has a finite limit as ψ tends to 1. This result is important because it allows for an approximation close to this limit where an analytical solution to the model exists.

Following Campbell and Viceira (2001) and Campbell et al. (2003), it is possible to approximate the return on the portfolio in (2.2.1). The approximation is exact in continuous time and very close to the true value at short time intervals. It is given by:

$$r_{p,t+1} = r_{1,t+1} + \alpha_t' x_{t+1} + \frac{1}{2}\alpha_t'(\sigma_x^2 - \Sigma_{xx}\alpha_t), \qquad (2.2.13)$$

where lower cases indicate variables in log and $\sigma_x^2 \equiv diag(\Sigma_{xx})$ is a vector with the diagonal elements of Σ_{xx}, i.e., the variances of the excess returns.

Similar to Campbell (1993, 1996), we can also log-linearise the budget constraint in the same problem. We do this around the unconditional mean of the log consumption-wealth ratio. This results in

$$\Delta w_{t+1} \approx r_{p,t+1} + \left(1 - \frac{1}{\rho}\right)(c_t - w_t) + k, \tag{2.2.14}$$

where Δ is the difference operator; $\rho \equiv 1 - \exp(E[c_t - w_t])$; and $k = \ln(\rho) + (1-\rho)\ln(1-\rho)/\rho$ is endogenous because it depends on the optimal level of c_t relative to w_t. When $\psi = 1$, $c_t - w_t$ is constant and $\rho = \delta$. In this case, the budget constraint approximation is exact.

Applying a second-order Taylor expansion to the Euler equation in (2.2.5) around the conditional means of Δc_{t+1}, $r_{p,t+1}$, $r_{i,t+1}$ gives way to

$$\begin{aligned} 0 &= \theta \ln \delta - \frac{\theta}{\psi} E_t \Delta c_{t+1} - (1-\theta) E_t r_{p,t+1} + E_t r_{i,t+1} \\ &\quad + \frac{1}{2} Var_t \left[-\frac{\theta}{\psi} \Delta c_{t+1} - (1-\theta) r_{p,t+1} + r_{i,t+1} \right]. \end{aligned} \tag{2.2.15}$$

This log-linearised Euler equation is exact if consumption and asset returns are jointly lognormally distributed. This is the case when the elasticity of intertemporal substitution equals one ($\psi = 1$).

Now, we subtract (2.2.15) evaluated in $i = 1$ from (2.2.15) evaluated in i. Further noting that $\Delta c_{t+1} = \Delta(c_{t+1} - w_{t+1}) + \Delta w_{t+1}$ yields

$$\begin{aligned} E_t(r_{i,t+1} - r_{1,t+1}) + \frac{1}{2} Var_t(r_{i,t+1} - r_{1,t+1}) &= \frac{\theta}{\psi}(\sigma_{i,c-w,t} - \sigma_{1,c-w,t}) \\ &\quad + \gamma(\sigma_{i,p,t} - \sigma_{1,p,t}) \\ &\quad - (\sigma_{i,1,t} - \sigma_{1,1,t}), \end{aligned} \tag{2.2.16}$$

where

$$
\begin{aligned}
\sigma_{i,c-w,t} &= Cov_t(r_{i,t+1}, c_{t+1} - w_{t+1}), \\
\sigma_{1,c-w,t} &= Cov_t(r_{1,t+1}, c_{t+1} - w_{t+1}), \\
\sigma_{i,p,t} &= Cov_t(r_{i,t+1}, r_{p,t+1}), \\
\sigma_{1,p,t} &= Cov_t(r_{1,t+1}, r_{p,t+1}), \\
\sigma_{i,1,t} &= Cov_t(r_{i,t+1}, r_{1,t+1}), \\
\sigma_{1,1,t} &= Var_t(r_{1,t+1}).
\end{aligned}
$$

On the left hand side of (2.2.16), we have the average excess return of asset i over asset 1 that each agent requires. We add one-half of the variance of the excess return because we consider log returns.[4]

The factors that determine the required excess return on each asset are shown on the right-hand side. Factors that contribute to raise the risk premium are the excess covariance with consumption growth and excess covariance with the portfolio return. The last term cancels out when the asset is risk free. It relates the covariance of the asset's excess return with the benchmark return to the required risk premium. Because consumption growth and portfolio return are endogenous, this is only a first-order condition describing the optimal solution. Thus, to solve the model, it is necessary to determine both those values.

[4] The left-hand side of equation (2.2.16) is determined by the dynamics of z_t, which also determines the variances and covariances on the right-hand side. However, the second term ($\gamma(\sigma_{i,p,t} - \sigma_{1,p,t})$) is a function of portfolio choice, α_t. This is calculated to make both sides equal for a given consumption policy. See Apendix A

Assuming that the optimal portfolio rule is linear in the VAR state vector but with a quadratic optimal consumption rule produces (2.2.17) and (2.2.18):

$$\alpha_t = A_0 + A_1 z_{t,} \tag{2.2.17}$$

$$c_t - w_t = b_0 + B_1' z_t + z_t' B_2 z_t. \tag{2.2.18}$$

Here A_0, A_1, b_0, B_1, and B_2 are constant coefficient matrices with dimensions $(n-1)\times 1$, $(n-1)\times m$, 1×1, $m\times 1$, and $m\times m$,, respectively, that we need to determine.

Now, we simply write the conditional moments that appeared in (2.2.16) as functions of the VAR and the unknown parameters in (2.2.17) and (2.2.18). Finally we solve for the parameters that satisfy (2.2.16).

For agent type h, we write the conditional expectation on the left-hand side of (2.2.16) as

$$E_{h,t}(x_{t+1}) + \frac{1}{2} Var_{h,t}(x_{t+1}) = H_x \phi_{h,0} + H_x \phi_{h,1} z_t + \frac{1}{2}\sigma^2_{h,x}, \tag{2.2.19}$$

where H_x is matrix that selects the vector of excess returns from the full state vector, and $Var_{h,t}$ is the conditional volatility estimated by agent h at time t.

Campbell and Viceira (2001) and Campbell et al. (2003) also show that it is possible to write the right-hand side of (2.2.16) as linear functions of the state variables:

$$\sigma_{h,c-w,t} - \sigma_{h,1,c-w,t}\iota \equiv \left[\sigma_{h,i,c-w,t} - \sigma_{h,1,c-w,t}\right]_{i=2,3,\ldots,n} = \Lambda_{h,0} + \Lambda_{h,1} z_t, \tag{2.2.20}$$

$$\sigma_{h,p,t} - \sigma_{h,1,p,t}\iota \equiv \left[\sigma_{h,i,p,t} - \sigma_{h,1,p,t}\right]_{i=2,3,\ldots,n} = \Sigma_{h,xx}\alpha_{h,t} + \sigma_{h,1x}, \tag{2.2.21}$$

$$\sigma_{h,1,t} - \sigma_{h,1,1,t}\iota \equiv \left[\sigma_{h,i,1,t} - \sigma_{h,1,1,t}\right]_{i=2,3,\ldots,n} = \sigma_{h,1x}, \tag{2.2.22}$$

where ι is a vector of ones.

The approximate demand for assets from agent h

By plugging (2.2.19) to (2.2.22) into the Euler equation (2.2.16) and solving for the portfolio rule, we finally obtain the optimal asset demand for each investor type h:

$$\begin{aligned} \alpha^*_{h,t} \;=\; & \overbrace{\frac{1}{\gamma}\Sigma^{-1}_{h,xx}\left[E_{h,t}(x_{t+1}) + \frac{1}{2}Var_{h,t}(x_{t+1}) + (1-\gamma)\sigma_{h,1x}\right]}^{\text{Myopic Demand}} \\ & \underbrace{+\frac{1}{\gamma}\Sigma^{-1}_{h,xx}\left[-\frac{\theta}{\psi}\left(\sigma_{h,c-w,t} - \sigma_{h,1,c-w,t}\iota\right)\right]}_{\text{Intertemporal hedging demand}}. \end{aligned} \tag{2.2.23}$$

Equation (2.2.23) is the generalised multiple-asset demand of Restoy (1992) and Campbell and Viceira (1999) for agent type h. It characterises the optimal portfolio choice as the sum of two components. The first one is exactly the myopic demand with many risky assets and lognormal returns. It does not depend on the elasticity of intertemporal substitution because this is a myopic component. The second is the intertemporal hedging demand term. With time-varying investment opportunities, the prediction in Merton (1969, 1971) is that an investor more risk-averse than a logarithmic investor would want to hedge against those shocks.[5] We verify this by noting that the second term indeed depends on the excess covariance between the shocks on the return on the risky asset and the shocks on consumption growth. The investor demands more assets with returns that are negatively correlated with the consumption growth because he is willing to smooth consumption. This makes the intertemporal hedging demand term usually positive for such assets.

Equation (2.2.23) highlights the difference between the myopic and the intertemporal frameworks. In this equation, we see that the myopic term is only a fraction of the complete demand for assets. The intertemporal hedging demand term is the part that is ignored when we cast investment problems within a myopic framework.

[5] A logarithmic investor has coeficient of risk aversion $\gamma = 1$; hence $\theta = 0$. Therefore, the portfolio rule of the investor is simply myopic, as we would expect. $\theta = 0$ sets the intertemporal hedging demand term to zero, and the only term left (that does not depend on θ) is the myopic one.

Define the Intertemporal Hedging Demand at time t for agent type h as

$$IHD_{h,t} \equiv \frac{1}{\gamma}\Sigma_{h,xx}^{-1}\left[-\frac{\theta}{\psi}\left(\sigma_{h,c-w,t} - \sigma_{h,1,c-w,t}\iota\right)\right]. \tag{2.2.24}$$

Note that the hedging demand depends on h and can also vary over time. We can now rewrite (2.2.23) as

$$\alpha_{ht}^{*} = \frac{1}{\gamma}\Sigma_{h,xx}^{-1}\left[E_{h,t}(x_{t+1}) + \frac{1}{2}Var_{h,t}(x_{t+1}) + (1-\gamma)\sigma_{h,1x}\right] + IHD_{h,t}. \tag{2.2.25}$$

2.2.2 Evolution of trader types

Thus far, we derived the demand for assets of a given agent type but with no discussion on how the agents initially choose their types. In this section, we model the evolution of η_{ht}, the fraction of agent type h at time t. This is the so-called evolutionary part of the model and describes how beliefs about the best strategy are updated over time.

Following Brock and Hommes (1997, 1998), agents observe the past performance of each strategy and then decide between them. Agents have access to fitness measures that are subjected to noise due to measurement errors or non-observable characteristics. The observed fitness of strategy h, $\tilde{U}_{h,t}$, is given by

$$\tilde{U}_{h,t} = U_{h,t} + \varepsilon_{h,t}, \tag{2.2.26}$$

where $U_{h,t}$ is the deterministic part of the measure, and $\varepsilon_{h,t}$ an *iid* noise across types, drawn from a double exponential distribution. In this case, the probability that a given agent chooses strategy h is given by the multinomial logit probabilities of a discrete choice when the number of agents tends to infinity. So, we describe the fractions n_{ht} of trader types as follows:

$$\eta_{ht} = \frac{\exp(\beta U_{h,t-1})}{\sum_{h=1}^{H} \exp(\beta U_{h,t-1})}, \tag{2.2.27}$$

where $U_{h,t-1}$ is the fitness measure of strategy h evaluated in period $t-1$, and β is a parameter regulating the intensity of choice. The later is inversely proportional to the variance of the noise $\varepsilon_{h,t}$.[6]

The measure of evolutionary fitness of strategy h is the realised profits over a certain period, which is given by

$$U_{h,t} = (x_t) \bullet \alpha_{h,t} + \omega U_{h,t-1}, \tag{2.2.28}$$

where ω is a memory parameter that reflects how slowly agents discount the success of past strategies when selecting their trading rules and $\bullet$ is the direct product operator. We then consider the simplest case: no memory, i.e., $\omega = 0$. In this case, (2.2.28) becomes

$$U_{h,t} = (x_t) \bullet \alpha_{h,t}. \tag{2.2.29}$$

[6] This ensures that $\beta = 0$ when the variance of noise is infinity. In this case, agents cannot observe differences in fitness and are not sensitive to differences in the performance of strategies. The other extreme situation is when the performances of the strategies could be perfectly observed, or $\beta = \infty$. In this case, all agents switch strategies when they see any difference in relative performances.

2.2.3 A tentative equilibrium

One of the drawbacks of using intertemporal models is the difficulty in determining uniqueness or even the existence of an equilibrium. One possibility to obtain it is to assume that if the market is in equilibrium, total demand, α_t^d, equals total supply, α_t^s, for each asset. In this case, the following equation holds true:

$$\sum_{h=1}^{H} \eta_{ht} \bullet \alpha_{ht} = \alpha_t^d = \alpha_t^s \tag{2.2.30}$$

where the vector η_{ht} denotes the (possibly different) fraction of trader type h at date t in each of the asset markets while considering H different trader types.

Now, combining (2.2.25) and (2.2.30) for the case of zero outside supply shares (i.e., $\alpha_t^s = 0$) yields the the market-clearing condition:

$$\sum_{h=1}^{H} \eta_{ht} \bullet \left\{ \begin{array}{c} \frac{1}{\gamma}\Sigma_{h,xx}^{-1}\left[E_{h,t}(x_{t+1}) + \frac{1}{2}Var_{h,t}(x_{t+1}) + (1-\gamma)\sigma_{h,1x}\right] \\ +IHD_{h,t} \end{array} \right\} = 0. \tag{2.2.31}$$

However, there is nothing that guarantees that the equation above has a unique or even multiple solutions and therefore the model is not guaranteed to be in general equilibrium. In this paper, I fit decision models of different types of investors to empirical data, but not an equilibrium model.

2.3 Empirical application

In this section, I assume the perspective of investors in the U.S.A. who would like to diversify their portfolio using the international stock markets. These investors classify strategies into fundamentalists and chartists. In each period, they need to decide whether to use the forecast given by one or the other model (i.e., choose their types). I estimate the model assuming that investors can allocate funds between four major stock markets: U.S. (Dow Jones Industrials), UK (FTSE all share), Japan (Nikkei 500) and Hong Kong (Hang Seng).

The main objective of the exercise is to evaluate the impact of considering only the myopic component of the demand for assets on the results of the model. I find that if investors believe that the market participants are very risk averse, then their assumptions about their investment horizons are extremely important. This happens because using the intertemporal demand for assets or only its myopic component often results in different conclusions in this case. I show that the intertemporal hedging demand term is not only significant, but it dominates for very risk averse agents. I also show that the inclusion of the intertemporal hedging term in the demand for assets has different effects for fundamentalist and chartist agents. These effects also depend on their risk aversion. In addition, I show that the level of noise in the observed performances also has different impacts on the model results whether we consider the complete intertemporal demand for assets or only its myopic component.

Finally, I show that the proportion of trader types fluctuates according to the market conditions. These fluctuations are relatively more prominent for the Nikkei and Hang Seng markets. One explanation is that these two markets show clearer regime switches during the observed period.

2.3.1 Data Description

I use quarterly data from the U.S., UK, Japan and Hong Kong stock markets. Table 2.3.1 reports the main descriptive statistics with all returns in U.S. dollars. I estimate the fundamentalist and chartist models using the complete data set for each individual market. These data sets go until the first quarter of 2007, but they start at different dates. The Dow Jones starts at the second quarter of 1978; the FTSE starts at the first quarter of 1965; the Hang Seng starts at the third quarter of 1973; and the Nikkei starts at the first quarter of 1992.

For the estimation of the intertemporal asset allocation problem, however, I restrict attention to the common sample ranging from the first quarter of 1993 until the first quarter of 2007. This is the period when forecasts of these models exist.

Datastream is the source for the index values and dividend-price ratios. Quarterly data regarding the American consumption-wealth

Returns	Dow	FTSE	Nikkei	Hang Seng
Mean	0.0085	0.0070	−0.0004	0.0178
Median	0.0172	0.0150	−0.0001	0.0437
Maximum	0.1966	0.5763	0.2427	0.5158
Minimum	−0.3211	−0.3493	−0.3737	−0.7288
Std.Dev.	0.08	0.11	0.13	0.19
Skewness	−0.84	0.34	−0.46	−0.83
Kurtosis	5.1	7.2	3.2	5.6
Observations	226	180	85	170

Table 2.3.1: Descriptive statistics for the series of Dow Jones Industrials, FTSE all shares, Nikkei 500 and Hang Seng real quarterly returns in US dolars

ratio comes from the Martin Lettau's website[7] and corresponds to the updated data set in Ludvigson and Lettau (2004). Finally, the CPI series comes from the U.S. Department of Labor Statistics.

2.3.2 Estimation

I construct the real stock return using the difference between the return on the stock index of each country and the U.S. inflation in

[7] `http://faculty.haas.berkeley.edu/lettau/`

the same period using the CPI. I report the results for $\Psi = 0.98$, $\beta = 10$ and $\gamma = 5$ or $\gamma = 50$. However, the model estimates for $\beta = \{0.25, 0.75, 0.5, 1, 5, 10, 20\}$ and $\gamma = \{1, 2, 5, 20, 50\}$ have the same qualitative results.

I restrict attention to a simple version of the model with two agents and four assets. I find the proportions of fundamentalists and chartists in two steps. First, I determine their demand for assets as in (2.2.23). Next, I use this as an input to determine the corresponding proportion of types given by (2.2.27). I use the constant conditional correlation GARCH specification proposed by Bollerslev (1990) to estimate the conditional variances and covariances in (2.2.23).

Estimated agents' models

Fundamentalist agents predict the real return on every asset using the past dividend-price ratio:

$$x_{t+1} = \mu + \rho_0 x_t + \rho_1 DP_t + \rho_2 DP_{t-l_2} + \rho_3 DP_{t-l_3} + e_t. \qquad (2.3.1)$$

Past real return (x_t) is included to eliminate serial correlation in the equation; e_t is an error term; l_2 and l_3 are lags that vary according to the asset that agents are forecasting. I choose the lags empirically to match the data.

Chartist traders use only past returns to forecast future returns for each asset. This model is given by

$$x_{t+1} = \mu + \rho_0 x_{t-l_0} + \rho_1 x_{t-l_1} + e_t. \tag{2.3.2}$$

I choose the lags empirically but the aim is to keep these lags small, given that momentum is mostly a short term effect.

I estimate these models for each one of the $n = 4$ assets. They provide the inputs for the (restricted) VAR that agents use to describe the market. Agents estimate the parameters in (2.3.1) and (2.3.2) recursively, based on the information available on each date. For example, agents use the information available up to third quarter of 1999 to estimate ρ in fourth quarter of 1999.

Table 2.3.2 displays the results of these estimations for the two agent types in each market (using the whole data set in the estimation). The fundamentalist models fit the data much better than the chartist ones. The positive coefficients of the lagged returns in the chartist models, however, are in accordance with the previous findings of momentum effect. The overall positive coefficients of the dividend-price ratios in the fundamentalist models are also in accordance with the literature. The biggest difference among the markets is that we can find a relationship between future return and the dividend-price ratio at much shorter horizons for the Nikkei index. The shorter estimation sample (first quarter of 1991 until the first quarter of 2007) for the Nikkei does not allow to test if there is a stronger relationship at longer horizons.

Forecasting models - Estimation results

Chartist				Fundamentalist		
Dow quarterly return				**Dow quarterly return**		
Variable	Coefficient	t-Statistic		Variable	Coefficient	t-Statistic
Dow Return (-1)	0.065	0.97		Dow Return (-1)	-0.322	-2.40
Dow Return (-3)	0.025	0.37		DP	0.084	1.74
C	0.007	1.40		DP (-40)	-0.107	-1.99
Adjusted R-squared		-0.004		DP (-60)	0.155	2.91
				C	-0.142	-2.35
				Adjusted R-squared		0.16
FTSE quarterly return				**FTSE quarterly return**		
Variable	Coefficient	t-Statistic		Variable	Coefficient	t-Statistic
FTSE Return (-1)	0.075	1.00		FTSE Return (-1)	-0.060	-0.59
FTSE Return (-3)	0.105	1.39		DP	0.136	3.09
C	0.006	0.81		DP (-40)	-0.074	-1.54
Adjusted R-squared		0.005		DP (-80)	0.040	1.04
				C	-0.099	-0.95
				Adjusted R-squared		0.06
Hang Seng quarterly return				**Hang Seng quarterly return**		
Variable	Coefficient	t-Statistic		Variable	Coefficient	t-Statistic
Hang Seng Return (-3)	0.139	1.80		Hang Seng Return (	-0.022	-0.19
C	0.017	1.19		DP	0.294	3.51
Adjusted R-squared		0.013		DP (-40)	0.087	1.47
				DP (-70)	-0.006	-0.10
				C	-0.419	-2.34
				Adjusted R-squared		0.16
Nikkei quarterly return				**Nikkei quarterly return**		
Variable	Coefficient	t-Statistic		Variable	Coefficient	t-Statistic
Nikkei Return (-3)	0.162	1.48		Nikkei Return (-1)	-0.083	-0.47
C	-0.003	-0.22		DP	0.077	0.44
Adjusted R-squared		0.015		DP (-1)	0.258	1.38
				DP (-3)	0.273	2.65
				C	0.082	3.66
				Adjusted R-squared		0.27

Table 2.3.2: The table shows the estimated models of the two agent types in each market. Fundamentalist agents estimate for each market a model of the form: $x_{t+1} = \mu + \rho_0 x_t + \rho_1 DP_t + \rho_2 DP_{t-l_2} + \rho_3 DP_{t-l_3} + e_t$, where x_t is the real return in time t; DP_t is the dividend-price ratio in time t; e_t is an error term and μ and the ρ are the coefficients to be estimated. Chartist agents estimate for each market a model of the form: $x_{t+1} = \mu + \rho_0 x_{t-l_0} + \rho_1 x_{t-l_1} + e_t$. In the table, the number in parenthesis gives the number of lags of the corresponding variable.

GARCH(2,1) estimation results

Chartist			Fundamentalist		
Dow Jones			Dow Jones		
	Alpha-1	0.14		Alpha-1	0.18
	Alpha-2	0.00		Alpha-2	0.37
	Beta	0.79		Beta	0.20
	Sum	0.92		Sum	0.75
FTSE			FTSE		
	Alpha-1	0.00		Alpha-1	0.00
	Alpha-2	0.25		Alpha-2	0.12
	Beta	0.61		Beta	0.59
	Sum	0.86		Sum	0.70
Hang Seng			Hang Seng		
	Alpha-1	0.12		Alpha-1	0.06
	Alpha-2	0.78		Alpha-2	0.32
	Beta	0.10		Beta	0.27
	Sum	1.00		Sum	0.65
Nikkei			Nikkei		
	Alpha-1	0.29		Alpha-1	0.04
	Alpha-2	0.03		Alpha-2	0.00
	Beta	0.36		Beta	0.33
	Sum	0.69		Sum	0.37

Table 2.3.3: Considering the $GARCH$ $(2,1)$ given by $\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \beta\sigma_{t-1}^2$, the table shows the estimated coefficients α_1, α_2 and β in each market. It also shows the sum of these coefficients.

Table 2.3.3 reports the coefficient estimates of the $GARCH(2,1)$ specification, given in:

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \beta\sigma_{t-1}^2 \tag{2.3.3}$$

These univariate GARCH are inputs for the constant conditional correlation GARCH of Bollerslev (1990). The sum of the coefficients in the GARCH models are at most one. This indicates that they are all weakly stationary, though not necessarily with finite unconditional variance.

2.3.3 Model results

The components of the demand for assets

Figures 2.3.1 and 2.3.2 show the relative importance of each component in the demand for assets of fundamentalist and chartist agents with two different levels of risk aversion. Equation (2.2.23) describes how to obtain the myopic and the intertemporal hedging demand terms in each graph.

Comparing the two columns in Figures 2.3.1 and 2.3.2, we see that as agents become more risk averse, the importance of the intertemporal hedging demand in relation to the myopic demand for assets increases. For very risk averse individuals and depending on the asset, the intertemporal hedging term in fact becomes the dominant component in the demand for assets. Figure 2.3.1 reveals that this happens with the Dow Jones and the FTSE for the fundamentalists whereas Figure 2.3.2 shows that this applies for every stock market index, but the Hang Seng, in the case of the chartist agents.

There are two main reasons that explain why the increase in risk aversion leads to an increase in the relative importance of the intertemporal hedging demand. The first is that it decreases the overall demand for risky assets, reducing the myopic demand term. The second is that the agents become more willing to hedge against changes in the investment opportunity set. Therefore, they demand more of assets with such

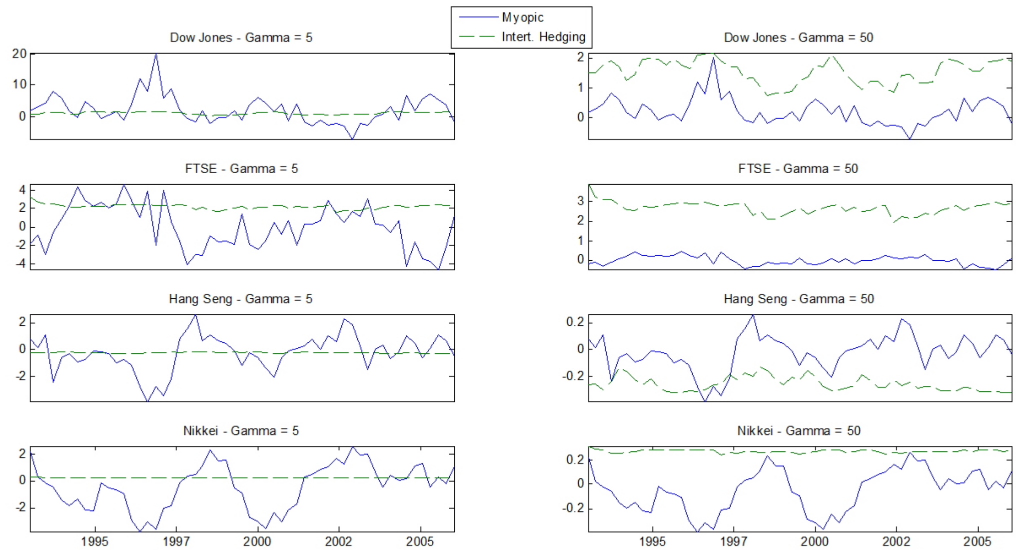

Figure 2.3.1: The figure shows the relative importance of the myopic and the intertemporal hedging terms in the demand for assets of fundamentalist agents. The results correspond to the four markets: U.S. (Dow Jones), Japan (Nikkei), U.K. (FTSE) and Hong Kong (Hang Seng) and to a coefficient of relative risk aversion $\gamma = 5$ or $\gamma = 50$.

properties (via the intertemporal hedging demand term).

Examining the pairs of Figures 2.3.2 and 2.3.1, we see that the intertemporal hedging demand term has different effects on the total demand for assets of fundamentalist and chartist agents. For instance, the intertemporal hedging demand for the Hang Seng is positive for the chartist agents, and negative for the fundamentalists.

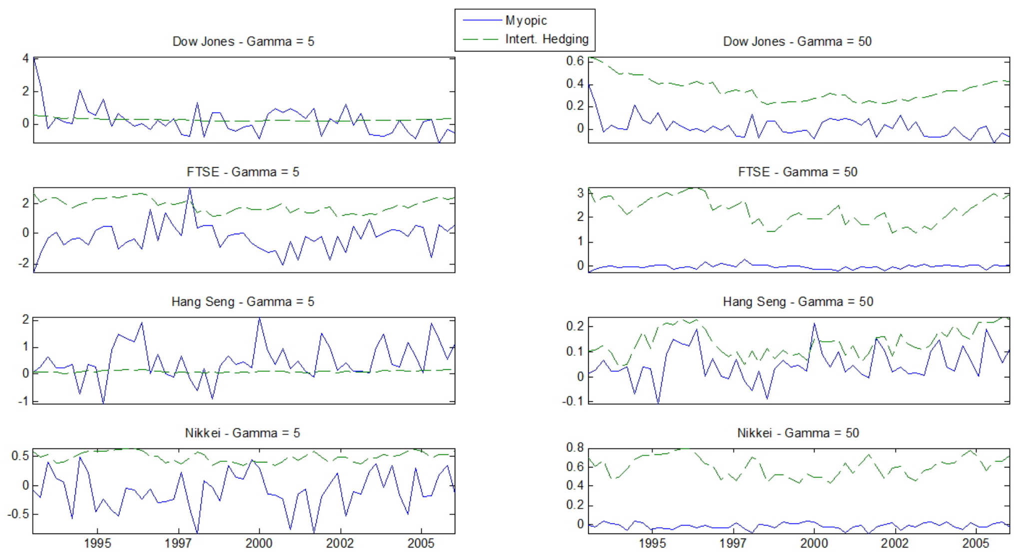

Figure 2.3.2: The figure shows the relative importance of the myopic and the intertemporal hedging components in the demand for assets of chartist agents. The results correspond to the four markets: U.S. (Dow Jones), Japan (Nikkei), U.K. (FTSE) and Hong Kong (Hang Seng) and to a coefficient of relative risk aversion $\gamma = 5$ or $\gamma = 50$.

The intertemporal hedging demand term with multiple assets

The four assets present desirable intertemporal hedging characteristics given the negative covariance between shocks on their returns and shocks on the consumption-wealth ratio.[8] Therefore, in a single asset framework, the intertemporal hedging demand would be positive for all of them. In a multiple asset framework, however, the results can be different.

Fundamentalists have a negative intertemporal hedging demand for the Hang Seng because from a portfolio perspective this asset is very risky. Table 2.3.4 shows that the shocks on the return on the Hang Seng have the largest variance and covariance with the shocks on the returns on the other assets. This happens for both fundamentalist and chartist agents. However, fundamentalists have an overall larger intertemporal hedging demand. We can find the intuition for this result in Campbell et al. (2003). They note that the predictability of returns increases the demand for intertemporal hedge. As mentioned before, fundamentalist agents use models that predict returns more accurately than chartists. Therefore, keeping everything else constant, fundamentalists should have higher intertemporal hedging demands. As a con-

[8] Although not reported, the estimated excess covariance between shocks on the asset's return and shocks on the consumption-wealth ratio, as given in equation (2.2.24), is negative.

For $\Psi = 0.98 < 1$ and for an agent that is more risk averse than a logarithmic one (i.e., $\gamma > 1$), it would be possible to obtain a negative value for the intertemporal hedging demand term if the excess covariance between shocks on the asset's return and shocks on the consumption-wealth ratio was positive.

sequence, they short the Hang Seng index to reduce the risk of their *overall* portfolio.

Chartist agents

Covariances ($\times 10^{-3}$)	Dow Jones	FTSE	Hang Seng	Nikkei
Dow Jones	6.4			
FTSE	5.5	7.3		
Hang Seng	8.1	8.4	21.1	
Nikkei	4.2	5.0	6.3	16.2

Fundamentalist agents

Covariances ($\times 10^{-3}$)	Dow Jones	FTSE	Hang Seng	Nikkei
Dow Jones	4.4			
FTSE	3.6	5.0		
Hang Seng	4.7	5.6	15.8	
Nikkei	2.6	2.6	4.0	10.3

Table 2.3.4: Variance-Covariance matrix of the shocks on the expected returns estimated by chartist and by fundamentalist agents.

Estimated proportions of types

Figure 2.3.3 plots the estimated proportions of fundamentalists given two different levels of risk aversion, $\gamma = 5$ and $\gamma = 50$. It compares the proportions obtained from the complete intertemporal demand for assets (i.e. including also the intertemporal hedging demand term) with the ones obtained from its myopic component alone.

As expected, considering only the myopic component or the complete demand for assets results in significantly different estimations when the agents are very risk averse, i.e., $\gamma = 50$. When agents are not extremely risk averse, i.e., $\gamma = 5$, the estimated proportions do not change much from one formulation to another in the data set used here. This happens regardless of the fact that the intertemporal hedging demand term, shown earlier, is significantly large for agents with both levels of risk aversion.

Changing the intensity of choice β

Figure 2.3.4 shows how the estimated proportions of agents change with the noise in the observed performances (captured by the values of β) given the myopic or intertemporal framework used. The plot shows that changing the value of β affects the variation in the proportions of agents. The intensity of choice, β, is negatively correlated with the magnitude of the noise in the observed performance of the strategy. In other words, a high value of β corresponds to a situation in which traders observe relative differences in performance more clearly. It increases the likelihood of the traders changing their types. This in turn results in a higher variation over time in the proportions of fundamentalists,

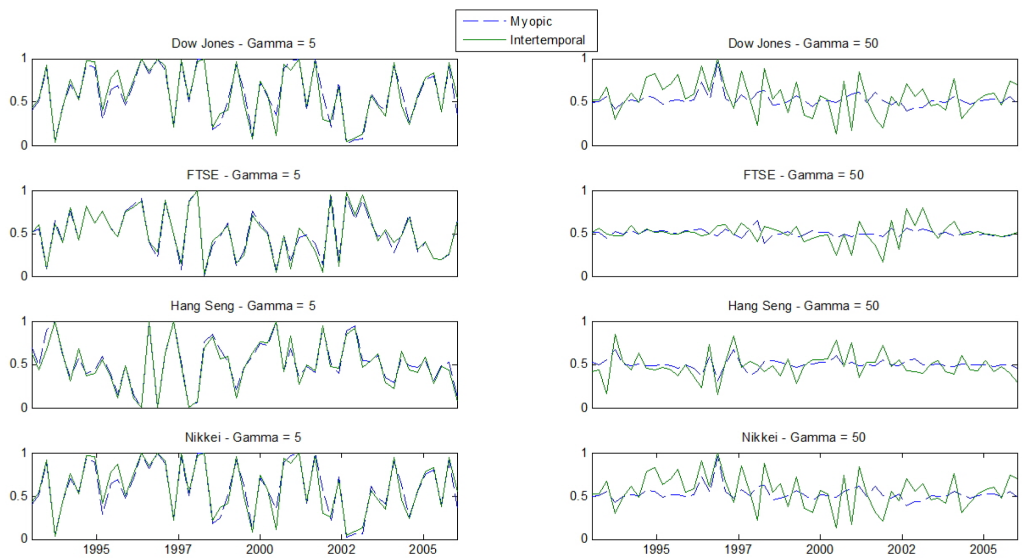

Figure 2.3.3: The figure shows the proportions of fundamentalists in each market estimated from the complete intertemporal demand for assets and also from its myopic component alone. I assume a coefficient of relative risk aversion $\gamma = 5$ or $\gamma = 50$.

as we see in the graph.

The picture also shows that this effect is stronger when we consider the complete demand for assets, as opposed to the myopic component alone. Considering the complete demand increases the differences in the performances of the strategies. This happens because the additional intertemporal hedging demand term is different across types and these differences do not cancel the differences in the myopic components.

The proportions of fundamentalists and the markets

Figure 2.3.5 displays the variation in the proportions of fundamentalist traders according to the market conditions. It shows these variations in the four different markets plotting each index level (in US$) with the corresponding fundamentalist proportion. In common, the plots show a pattern of a decrease in the fraction of fundamentalists being followed by a reversal in prices and a subsequent increase in the fraction of fundamentalists. This pattern is clearer in the Hang Seng index, or during the period between the last quarter of 1998 and the last quarter of 2002 in the Nikkei and also, to a lesser extent, in the FTSE.

The decrease in the fraction of fundamentalists occurs because fundamentalist strategies are not successful in forecasting returns when prices do not follow the fundamentals. This is what happens between 1998 and 1999 especially in the Hang Seng and Nikkei indices.

When prices start to revert to the fundamentals, the traders begin

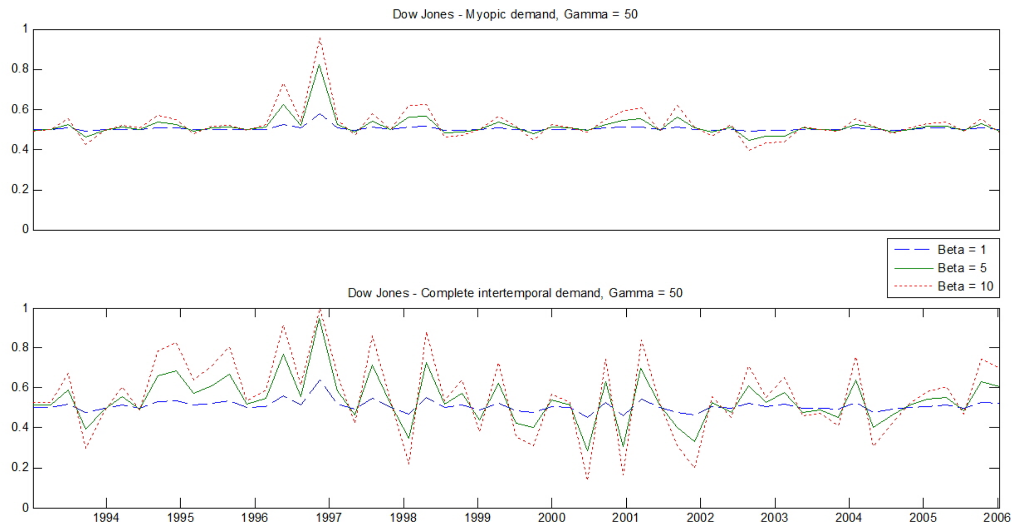

Figure 2.3.4: The graph shows the estimated proportions in the Dow Jones for $\beta = \{1, 5, 10\}$ considering the complete demand for assets or only its myopic component for $\gamma = 50$. It shows the relationship between the estimated proportions, the demand for assets and the different levels of noise in the observed performances (captured by the values of β).

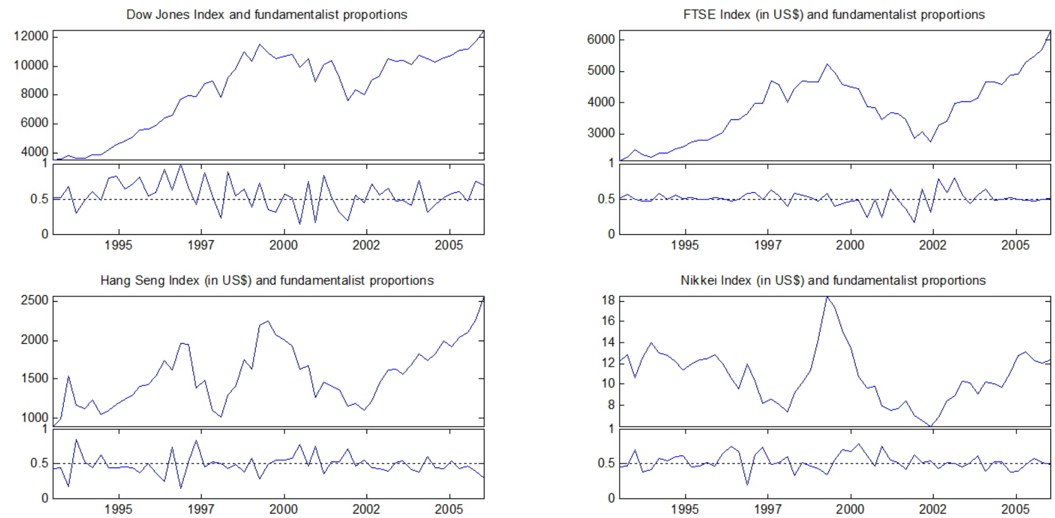

Figure 2.3.5: The figure shows the proportions of fundamentalists and the corresponding stock market index value when agents have a coefficient of relative risk aversion $\gamma = 50$.

to believe that fundamentalist strategies are correct. Subsequently, the proportion of fundamentalists increases until the first half of 2000, when the market prices are back to the level that they were in 1998. In the Hang Seng (and to a lesser extent in the FTSE also), we do not observe a reversal, but we see that the last increase in prices starting in the second half of 2003 is not consistent with fundamentals, as provided by our model. Finally, the participation of fundamentalists in the Dow Jones does not oscillate much.

2.4 Concluding remarks

In this paper I develop a new asset-pricing model in which agents with long investment horizons maximise a recursive utility function and choose the strategy used to forecast returns based on its previous profitability. The model keeps many characteristics of earlier asset pricing models with heterogeneous beliefs. For instance, it has the ability to generate changes in prices that are not driven by fundamentals without requiring restrictive assumptions about the agent's preferences or rationality.

The paper extends the literature on heterogeneous beliefs into two different directions. First by considering agents with long-term investment horizons as opposed to myopic investors. The empirical exercise shows that the component on the demand for assets that is ignored in a myopic framework can be significantly large. This is especially true when agents are very risk averse. In addition, the impact of changes in the parameters of the model is also different whether we consider the complete intertemporal asset demand or only its myopic component. These parameters are for instance the noise in observed performances, captured by β, or the level of risk aversion, γ.

The paper also extends the literature on heterogeneous beliefs by considering an arbitrary large number of assets, n. The negative intertemporal hedging demand for the Hang Seng by fundamentalist agents, for instance, would be positive in a single risky asset formulation given its desirable hedging properties.

2.5 Chapter Appendix

2.5.1 Excess returns and the stochastic discount factor

The basic equation of asset pricing (in terms of returns) can be written as follows:

$$1 = E_t \left[M_{t+1}(1 + R_{i,t+1}) \right], \tag{2.5.1}$$

where M_{t+1} is the "stochastic discount factor" (SDF) that prices any asset in the economy. Equation (2.5.1) can be developed into:

$$\begin{aligned}
1 &= E_t \left[M_{t+1}(1 + R_{i,t+1}) \right] \\
&= E_t \left[M_{t+1} \right] \cdot E_t \left[(1 + R_{i,t+1}) \right] + Cov_t \left(M_{t+1}, (1 + R_{i,t+1}) \right) \\
&= E_t \left[M_{t+1} \right] \cdot E_t \left[(1 + R_{i,t+1}) \right] + Cov_t \left(M_{t+1}, R_{i,t+1} \right).
\end{aligned}$$

Using the fact that $E_t \left[M_{t+1} \right]^{-1} = (1+R_{f,t+1})$, obtained from (2.5.1) for the risk-free asset, we have:

$$\begin{aligned}
E_t \left[(1 + R_{i,t+1}) \right] &= \frac{1 - Cov_t \left(M_{t+1}, R_{i,t+1} \right)}{E_t \left[M_{t+1} \right]} \\
&= (1 + R_{f,t+1}) - \frac{Cov_t \left(M_{t+1}, R_{i,t+1} \right)}{E_t \left[M_{t+1} \right]},
\end{aligned}$$

and finally, the expression for the excess returns rearranging the equation once again is given by:

$$\begin{aligned} E_t\left[(1+R_{i,t+1})-(1+R_{f,t+1})\right] &= -\frac{Cov_t\left(M_{t+1},R_{i,t+1}\right)}{E_t\left[M_{t+1}\right]} \\ E_t\left[R_{i,t+1}-R_{f,t+1}\right] &= -\frac{Cov_t\left(M_{t+1},R_{i,t+1}\right)}{E_t\left[M_{t+1}\right]}. \end{aligned} \quad (2.5.2)$$

During the derivation of the approximate solution in the text, we obtain (2.2.16), that is similar to (2.5.2). The left-hand side of (2.5.2) is the expected excess return for asset i. This expectation is given by the beliefs of the agents, modeled by the VAR described in the text. Agents with heterogeneous beliefs have different expectations of returns and therefore different demands for assets.

Note that, in the Euler equation (2.2.5), we obtain:

$$M_{t+1}=\left\{\delta\left(\frac{C_{h,t+1}}{C_{h,t}}\right)^{-\frac{1}{\psi}}\right\}^{\theta}\left(1+R_{h,p,t+1}\right)^{-(1-\theta)}. \quad (2.5.3)$$

In this case, the right hand side of (2.5.2) depends on the agent's choices, i.e., the covariance term depends on the portfolio composition, $\alpha^*_{h,t}$ (via $R_{h,p,t+1}$) and consumption, $C^*_{h,t}$, the two variables that the agent chooses and also the only two sources of variability in the SDF.

So, given the expectation on the left-hand side of (2.5.2) and a consumption policy, we are able to determine the portfolio choice (i.e., asset demands).

Chapter 3

Regularization and portfolio selection

3.1 Introduction

The biggest issue regarding the implementation of mean-variance optimisation is estimation error. Small differences between asset returns or covariances are exploited by the optimiser. When these differences are due to estimation error instead of the real differences in the data generating process of returns for each asset, the problem becomes more significant.

In theory, having highly concentrated portfolios is not always problematic, as pointed out by Green and Hollifield (1992). When the concentration of the portfolios is due to estimation error, however, a suboptimal allocation would result that tends to worsen as the estimation error grows and more extreme allocations are selected. Portfolios in this case tend to be underdiversified and have extreme allocations. The result is that they tend to perform poorly out of sample even when they are compared to naive portfolios, such as the $1/N$ equal allocation shown in DeMiguel et al. (2009). Estimation error becomes an even bigger issue as the number of available assets grows.

Even when the problem is to find the minimum-variance portfolio (which ignores expected returns and reduces estimation error), the result is still an underdiversified portfolio with poor out-of-sample performance. In an attempt to improve the out-of-sample performance of these estimates, several econometric procedures are proposed. However, no unique solution has been presented so far.

This chapter begins with a brief review of plug-in estimation, discussing its asymptotic and finite sample properties. Later, we present the regularisation methods and how they can be applied to the portfolio choice problem.

3.1.1 Plug-in estimation

Plug-in estimation is the most widely used econometric approach in the portfolio choice literature. In this approach, the parameters of a given model are estimated and plugged into the analytical solution obtained from the theoretical model. Naturally, the estimation error in the parameters obtained with this approach will be passed on to portfolio weights, and the resulting allocation is different from the optimal allocation in almost every case.

Single period problem - Asymptotic properties

Consider the mean variance problem with a risk-free asset as an example. With iid excess returns, the optimal portfolio weights are given by

$$x^* = \frac{1}{\gamma}\Sigma^{-1}\mu, \tag{3.1.1}$$

where μ is the (constant) risk premia and Σ is the variance-covariance matrix of returns. Given the excess return data $\{r_{t+1}\}_{t=1}^{T}$, the moments μ and Σ can be estimated using the sample counterparts:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_{t+1} \tag{3.1.2}$$

and

$$\hat{\Sigma} = \frac{1}{T-N-2}\sum_{t=1}^{T}(r_{t+1}-\hat{\mu})(r_{t+1}-\hat{\mu})'. \tag{3.1.3}$$

Plugging these values into (3.1.1) results in the estimated weights $\hat{x}^* = (1/\gamma)\hat{\Sigma}^{-1}\hat{\mu}$. Under normality, this estimator is unbiased:

$$E[\hat{x}^*] = \frac{1}{\gamma} E[\hat{\Sigma}^{-1}] E[\hat{\mu}] = \frac{1}{\gamma} \Sigma^{-1} \mu = x^*. \tag{3.1.4}$$

The first equality comes from independence between $\hat{\Sigma}$ and $\hat{\mu}$, and the second is due to the unbiasedness of $\hat{\Sigma}^{-1}$ and $\hat{\mu}$.[1]

The second moments of the plug-in estimator can be derived by expanding the estimation around the true risk premia and return covariance matrix. To illustrate the technique, we consider only one risky asset. In this case, it can be shown that the variance of the estimator is given by

$$var[\hat{x}^*] = \frac{1}{\gamma^2} \left(\frac{\mu}{\sigma^2} \right)^2 \left(\frac{var[\hat{\mu}]}{\mu^2} + \frac{var[\hat{\sigma}^2]}{\sigma^4} \right). \tag{3.1.5}$$

This illustrates that the imprecision of the plug-in estimator is proportional to the magnitude of the true optimum portfolio weights

$$x^* = \frac{1}{\gamma} \left(\frac{\mu}{\sigma^2} \right), \tag{3.1.6}$$

and depends on the imprecision in the estimation of the volatility and risk premia, each scaled by their respective true magnitudes. For real applications, portfolio weights tend to be estimated very imprecisely because inputs to the estimator are difficult to pin down.

The second moments become more difficult to estimate with fat tails because outliers have a strong impact on the estimates. Therefore, stylised facts of financial series, such as the conditional heteroskedasticity,

[1] Without normality, the plug in estimator is still consistent with $plim\ \hat{x}^* = x^*$.

inflate the variance of the unconditional sample variance considerably. This illustrates the point that both return moments can contribute to asymptotic imprecision of plug-in portfolio weight estimates.

Plug-in estimation in finite samples

The asymptotic results derived previously are useful in characterising statistical uncertainty when the sample size is **large enough** with respect to the number of parameters to be estimated. On the other hand, it is easy to find real-life applications of portfolio selection where the number of assets is on the order of thousands while the length of observations is still on the order of decades. This characterises this solution as an ill-posed or ill-conditioned problem. The prevailing issue with plug-in estimates in portfolio selection in these cases becomes finite sample performance.

A substantial amount of the literature describes the shortcomings of plug-in estimates. In general, findings show that plug-in estimates could be very unreliable even with a relatively large sample size. This is especially true when the number of assets in the portfolio increases.[2]

Much of the recent literature on portfolio selection focuses on finding econometric methods with better finite sample properties, and shrinkage estimation is one of the most prominent methods.

[2] Note that the number of unique elements of the return covariance matrix increases at a quadratic rate with the number of assets.

3.2 Regularisation to improve estimates: constraining, penalising and shrinking

It is well known that the plug in estimator's performance can be improved in finite samples by constraining/shrinking the estimates. An early example of this is given by James and Stein (1961) with respect to estimates of the mean of a multivariate normal. These estimates are shrunk towards a common mean and were shown to outperform the sample mean in terms of mean squared error for dimensions as small as 3.

In the context of portfolio selection, regularisation can be introduced in at least two alternative ways. We can obtain an empirical minimum-variance portfolio by plugging in a regularised return covariance matrix to the unconstrained problem in:

$$\begin{aligned} x^*_{MINU}(\Sigma_t) \quad = \quad & \arg\min_x \quad x^T \Sigma_t x \\ & \text{s.t.} \qquad x^T \iota = 1. \end{aligned} \tag{3.2.1}$$

Alternatively, we can interpret the weights of the portfolio as the coefficients to be estimated and apply regularisation techniques to these coefficients while plugging in the sample covariance matrix to same problem.

In this section, we first provide an overview of regularisation for statistical estimates. Then, we discuss the two alternative paths to regularising empirically optimised portfolios. We conclude with a brief discussion on the equivalence of the two alternative approaches.

3.2.1 Statistical regularisation

The basic idea in shrinkage estimation is that it is possible to reduce an estimator's variance by averaging it with a given constant that, by definition, has no variance. This can be done at the expense of including some bias in the estimation, and the goal to correctly apply these estimators is to find the optimum balance between bias and variance.

Penalised estimates shrink the maximum likelihood estimators (MLE) towards a deterministic minimiser of a deterministic function. A good penalty is one that introduces the least bias in the estimates but reduces a large part of their variance.

Regularization by explicit shrinking

James and Stein (1961) introduced what is perhaps the earliest example of a regularised statistical estimate. In their set-up, the N-dimensional vector of means of a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$ are to be estimated. The classical non-regularised maximum likelihood estimate for μ_j was the sample average $\bar{y}_j$. Letting the $\hat{\mu}_0$ be the "grand mean" (i.e., the average of the sample averages $\bar{y}_j$), James and Stein (1961) introduce a family of estimators indexed by $\delta \in (0, 1)$ as

$$\hat{\mu}_j(\delta) = \delta\hat{\mu}_0 + (1 - \delta)\bar{y}_j, \text{ with } \delta \in (0, 1). \tag{3.2.2}$$

In other words, each estimator in this family shrinks each sample mean towards a common mean with the amount of shrinkage given by δ. James and Stein (1961) show that the best performance in terms of mean squared error is achieved by setting δ to be

$$\delta^* = \min\left[1, \frac{(N-2)/T}{(\mu - \mu_0\iota)^T\Sigma^{-1}(\mu - \mu_0\iota)}\right], \text{ where } \mu_0 = \frac{1}{N}\sum_{j=1}^{N}\mu_j. \tag{3.2.3}$$

Although the optimal weighting δ^* was derived for estimating the means, the equation above provides an intuitive guidance on how to tune the regularisation parameter. More aggressive shrinkage is advised in three cases: first, when the bias introduced by the shrinkage is small (as measured by $(\mu-\mu_0\iota)\Sigma^{-1}(\mu-\mu_0\iota)$); second, when the number of parameters being estimated (N) is large; and third, when the noise level is high (small $\det(\Sigma)$). As the sample size increases, however, less shrinkage seems to be necessary.

The general lesson seems to be that the less stable the non-regularised estimate ($\bar{y}_j$) is, whether the instability is due to noisier measurements or the smaller sample size or the smaller bias introduced by shrinkage, the more can be gained by borrowing strength from the more restricted estimate ($\hat{\mu}_0$). Although enlightening, the expression in (3.2.3) is not feasible because it involves the very parameters we are trying to estimate. Practical methods for tuning δ are required, and some are discussed later.

A natural extension of the James and Stein (1961) estimate for a vector of means is to consider the family of estimates:

$$\hat{\theta}(\delta) = \delta\hat{\theta}_0 + (1-\delta)\hat{\theta}_{MLE}, \tag{3.2.4}$$

where $\hat{\theta}_{MLE}$ is the MLE estimate for parameter θ in a broader model (low bias, high variance) and $\hat{\theta}_0$ is an MLE estimate for the same parameter under more restrictive assumptions (higher bias, lower variance). Again, finding an appropriate way to tune δ is an integral part of any such method.

Regularisation by penalisation

Penalised estimates are another way to obtain regularised stable statistical estimates. For a given loss function $L(Z, t)$ and

$$\hat{\theta}(\lambda) = \arg\min_{t\in\mathbb{R}^p} \left[\sum_{i=1}^{n} L(Z_i, t) + \lambda \cdot T(t) \right], \tag{3.2.5}$$

where L is a loss function, T is a penalty function, Z_i are the observed data points, and $\lambda \geq 0$ is a tuning parameter that trades off between the loss and penalty functions. The estimate $\hat{\theta}(\lambda)$ in (3.2.5) and (3.2.10) can be interpreted as a compromise between the unpenalised M-estimator $\hat{\theta}(0)$ and the deterministic a priori estimate $\hat{\theta}(\infty) := \arg\min_t T(t)$ with λ controlling the emphasis put on the prior/penalty T.

The ℓ_1- and ℓ_2-penalised portfolios Considering a constant λ_1, a vector of portfolio weights x, the sample covariance matrix $\hat{\Sigma}_t$, and a vector of ones ι, we may write

$$\sum_{i=1}^{n} L(Z_i, x) = x^T \hat{\Sigma}_t x + \lambda_1 \cdot x^T \iota$$

in equation (3.2.5). In this case we obtain, respectively, the ℓ_1- or ℓ_2-penalised portfolios assuming

$$T(x) = \sum_{i=1}^{n} \|x_i\|_1 \tag{3.2.6}$$

or

$$T(x) = \sum_{i=1}^{n} \|x_i\|_2, \tag{3.2.7}$$

where the right hand side of equation (3.2.6) represents the ℓ_1-norm of the portfolio weights x, and the right hand side of equation (3.2.7) represents its ℓ_2-norm.

Therefore, the ℓ_1- or ℓ_2-penalised estimates are given, respectivelly, by equation (3.2.8) or equation (3.2.9):

$$\hat{\theta}(\lambda) = \arg\min_{x \in \mathbb{R}^p} \left[x^T \hat{\Sigma}_t x + \lambda_1 \cdot x^T \iota + \lambda \cdot \sum_{i=1}^{n} \|x_i\|_1 \right], \tag{3.2.8}$$

$$\hat{\theta}(\lambda) = \arg\min_{x \in \mathbb{R}^p} \left[x^T \hat{\Sigma}_t x + \lambda_1 \cdot x^T \iota + \lambda \cdot \sum_{i=1}^{n} \|x_i\|_2 \right]. \tag{3.2.9}$$

In equations (3.4.4) and (3.4.14) in the next subsection, we show how to obtain the same portfolios using constraints.

Often, the loss L and penalty T can be interpreted as neg-log likelihood functions. In such cases, $p(\theta) \propto [\exp(-T(\theta))]^{\lambda}$ plays the role of a prior distribution on θ and $\hat{\theta}(\lambda)$can be interpreted as the maximum a posteriori likelihood estimate (MAPLE) for θ, that is, $\hat{\theta}(\lambda)$ maximizes the a posteriori likelihood for θ:

$$\hat{\theta}(\lambda) = \arg\max_{t \in \mathbb{R}^p} \left\{ \left[\prod_{i=1}^{n} \exp\left(-L(Z_i, t)\right) \right] \cdot \left[\exp(-T(\theta)) \right]^{\lambda} \right\}. \quad (3.2.10)$$

Penalised estimates in the form $\hat{\theta}(\lambda)$ have garnered increasing attention in the statistical literature as a means of fitting increasingly complex models (large p) with limited amounts of data. The ridge regression of Hoerl and Kennard (1970) provides an early example of a penalised estimate for linear regression models using the squared error loss ℓ_2-norm of errors and a penalty on the Euclidean (ℓ_2) norm of the regression parameters. In such cases, the MAPLE estimate can be interpreted as a random effects model with a normal $\mathcal{N}(0, \frac{1}{\lambda} \cdot \mathbf{I}_p)$ prior to the parameters. More recently, the penalised approach was extended to a myriad of loss and penalty functions.

Penalised estimate as constrained estimates

Penalised estimates as defined in (3.2.5) are often defined in terms of a constrained M-estimator. Consider estimates defined as:

$$\begin{aligned} \tilde{\theta}(\mathcal{T}) \quad := \quad & \arg\min_{t \in \mathbb{R}^p} \quad \textstyle\sum_{i=1}^{n} L(Z_i, t) \\ & \text{s.t.} \qquad T(t) \leq \mathcal{T}, \end{aligned} \quad (3.2.11)$$

for the same functions L and T as in (3.2.5) above. As long as $\mathcal{T} \leq \hat{\theta}(0)$, writing this problem in Lagrangian form shows that a corresponding λ exist in (3.2.5) for each $\mathcal{T}$ in (3.2.11). This is certainly true for the loss functions and penalty functions considered in this paper; therefore, we will use the terms “constraints” and “penalties” interchangeably.

As a final remark, notice that the parameter λ in this framework played a similar role as the δ parameter in shrinkage estimates, as defined in (3.2.4): that is, more emphasis is placed on the shrinking target values with larger values of the parameter λ. As before, a method for tuning λ is needed to select one estimate $\hat{\theta}(\lambda)$ from the family of estimates defined by (3.2.5), often called the regularisation path. The same principles in choosing δ apply when selecting λ: more constrained estimates are preferred when there is more noise in the data, when the sample size is smaller, and when the introduction of a constraint/penalty produces less distortion.

Tuning the regularisation/shrinkage parameter

In empirical applications, we may use cross validation to select the optimal amount of regularisation to be applied to the problem. The selected value in both problems gives us information regarding the optimal strategy to be used.

When we solve the 2-norm restricted optimisation and find the optimal amount of regularisation to be equal to $1/N$,for instance, the noise in the data for that sample is so large that it is preferable to use the naive $1/N$ allocation as in DeMiguel et al. (2009). If the optimal δ is different from $1/N$, the naive allocation is expected to be sub-optimal.

The same applies to the 1-norm restriction formulation: if cross validation gives us $\delta \neq 1$, we have evidence that using a no-short-sale portfolio is not optimal; nevertheless, this portfolio may be better than the portfolio with no restrictions at all.

These results also hold in the reverse case. If we find an optimal non-binding δ, the regularised solution is not needed; investors would obtain better portfolios by using the unrestricted formulation of the mean variance optimisation.

3.3 Plug-in portfolios with regularised empirical covariance matrices

It is possible to apply shrinkage estimation to portfolio selection problems not only to estimate the risk premia but also to estimate the covariance matrix.[3]. In this last case, a covariance matrix $\hat{\Sigma}_s$ is usually proposed; the matrix is the convex combination of the sample estima-

[3] See Jobson et al. (1979), Jobson and Korkie (1980), Frost and Savarino (1986) and Jorion (1986) for risk premia shrinkage estimation, and Frost and Savarino (1986) and Ledoit and Wolf (2004) for shrinking the covariance matrix of returns.

tion $\hat{\Sigma}$ and a shrinkage target $\hat{S}$, as in:

$$\hat{\Sigma}_s = \delta \hat{S} + (1 - \delta)\hat{\Sigma}. \tag{3.3.1}$$

The usual candidates include the identity matrix, the equal correlation covariance matrix or the one-factor matrix.

3.3.1 No-short-sales constraint: Jagannathan and Ma (2003)

Imposing no-short-sales constraints on the minimum-variance portfolio changes the problem to

$$\begin{aligned} x^*_{NSS}(\hat{\Sigma}_t) \quad = \quad & \arg\min_x \quad x^T \hat{\Sigma}_t x \\ & \text{s.t.} \quad x^T \iota = 1, \\ & \qquad\quad x \geqslant 0, \end{aligned}$$

where the last restriction ensures that all weights are positive (meaning that short sale is not permitted). Jagannathan and Ma (2003) shows that the solution to this problem is equivalent to the unconstrained problem if the sample covariance matrix is replaced by

$$\hat{\Sigma}_{JM} = \hat{\Sigma}_t - \lambda \iota' - \iota \lambda', \tag{3.3.2}$$

where $\lambda \in \mathbb{R}^N$ is the vector of Lagrangian multipliers from the restricted optimisation.

3.3.2 Shrinking the covariance matrix towards a deterministic one: Ledoit and Wolf (2003, 2004)

Ledoit and Wolf (2004) consider several shrinking targets such as the identity matrix, the constant correlation matrix, and the covariance matrix obtained from estimating a 1-factor model with the market as the factor. They use an alternative covariance matrix that is a convex combination between the sample covariance and the selected target matrix, as given by

$$\hat{\Sigma}_{LW} = \frac{1}{1+\nu}\hat{\Sigma} + \frac{\nu}{1+\nu}\hat{\Sigma}_{TARGET}, \tag{3.3.3}$$

where $v \in \mathbb{R}^{+}$ is a constant and $\hat{\Sigma}_{TARGET} \in \mathbb{R}^{N \times N}$ is the target matrix.

Ledoit and Wolf also show how to find the asymptotically optimum value for v that minimises the expected Frobenius norm of the difference between the matrix $\hat{\Sigma}_{LW}$ and the true covariance matrix.

3.4 Constraining/shrinking portfolio weights

An alternative to obtaining stable, data-driven minimum-variance portfolios relies on interpreting the weights of the optimal portfolio themselves as the parameter of interest and applying regularisation techniques directly to the portfolio weights, as in DeMiguel et al. (2009).

This framework is particularly appealing because it is often easier to set sensible shrinking targets to portfolio weights (e.g., equal weighting, market capitalisation) than it is to model the structure of the assets correlations. The use of theoretically supported targets derived from asset-pricing models is also easier within this framework, which also better captures the link between first and second moments because this link doesn't need to be modelled.

DeMiguel et al. (2009) also argues that constraining portfolio weights give extra flexibility and interpretation to the constraints that are not easily obtained when constraints on the moments of returns are imposed.

3.4.1 Shrunk portfolios

Given the unstable empirical minimum-variance portfolio $x^*_{MINU}(\hat{\Sigma})$ and a more stable target portfolio $\hat{x}_0$ (which may or may not depend on the data), a shrunken portfolio can be defined in the spirit of James and Stein (1961) as

$$\hat{x}^*_{JS}(\delta) = \delta \cdot \hat{x}_0 + (1-\delta)\hat{x}^*_{MINU}(\hat{\Sigma}), \qquad (3.4.1)$$

where $\delta$is a shrinkage parameter that must be determined empirically. Because there is no self-evident counterpart to the grand mean in the portfolio selection problem, different alternatives can be used as the shrinkage target $\hat{x}_0$Two possible targets with good empirical performance are: the completely balanced portfolio with equal weights on all assets, and the no-short-sales empirical minimum-variance portfolio $x^*_{NSS}(\hat{\Sigma})$.

Norm-constrained minimum-variance portfolio

An alternative way to reduce statistical error in portfolio selection problems is through the use of portfolio constraints. Frost and Savarino (1988) shows that portfolio constraints truncate the extreme portfolio weights. While the theory does not rule out the optimality of portfolio with extreme weights (e.g., Green and Hollifield (1992)), the empirical evidence is that constrained mean-variance optimisers do lead to better out-of-sample performance, as in DeMiguel et al. (2009), Michaud (1989)and others. Such empirical results suggest that the extreme weights observed in plug-in empirical minimum-variance portfolios $x^*_{MINU}\left(\hat{\Sigma}\right)$ are actually artefacts of estimation errors rather than a reflection of the true correlation among the assets.

Several alternative constraints can be proposed as a means to obtain more stable portfolios. Given that both the minimum-variance portfolio selection and the estimation of linear regression parameters are defined as minimisers of quadratic functions, many portfolio constraints can be traced back to the literature on penalised linear regression. The ℓ_p-penalised minimum-variance portfolio introduced by DeMiguel et al.

(2009) is defined as

$$\begin{aligned} x^*_{\ell_p}(\hat{\Sigma}_t, \delta) &= \arg\min_x \quad x^T \hat{\Sigma}_t x \\ &\quad \text{s.t.} \quad x^T \iota = 1, \\ &\qquad\qquad \|x\|_p \leq \delta, \end{aligned}$$

where $\|x\|_p$ denotes the ℓ_p-norm of the portfolio weights vector, i.e.,

$$\|x\|_p := \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p}. \tag{3.4.2}$$

These portfolios find their counterpart in the penalised linear regression literature in the bridge estimates of Frank and Friedman (1993).]. The CAP portfolios to be defined can also be traced back to the linear regression literature. These portfolios are based on penalties that translate grouping information into penalties for the estimation problem (see Zhao et al. (2009)).

Penalised portfolios can also be obtained without explicitly constraining the norm of the minimum-variance portfolio. This is the case of the partial minimum-variance portfolios of DeMiguel et al. (2009)which is related to the 2-norm-constrained minimum-variance portfolios obtained without constraining any norms[4]. These portfolios can also be traced back to the linear regression literature, and their counterpart can be found in the partial least squares of Wold (1975). The 2-step unconstrained minimum-variance portfolios that we would

[4] DeMiguel et al. (2009) view these portfolios as a discrete first-order approximation to the 2-norm-constrained portfolios.

describe later constitute another example that can be interpreted as a penalised portfolio without explicit norm constraints. Before we introduce our grouped portfolios, we discuss two important particular cases of norm-penalised portfolios here: the ℓ_1 and the ℓ_2 penalised minimum-variance portfolios.

The ℓ_1-constrained portfolio and short-sales constraint

Setting $p = 1$ in equation (3.4.2) gives

$$||x||_1 = \sum_{i=1}^{N} |x_i|, \tag{3.4.3}$$

which is the sum of the absolute values of the portfolio weights. The optimisation problem in this particular case becomes

$$\begin{aligned} x^*_{\ell_p}(\hat{\Sigma}_t, \delta) \quad = \quad & \min_x \quad x^T \hat{\Sigma}_t x \\ & \text{s.t.} \quad x^T \iota = 1, \\ & \qquad \textstyle\sum_{i=1}^{n} \|x_i\| \leq \delta. \end{aligned} \tag{3.4.4}$$

It is easy to see that when $\delta = 1$ the budget restriction along with the ℓ_1 restriction are equivalent to the usual short-sales constraint.[5] DeMiguel et al. (2009) also consider less restrictive alternatives where $\delta > 1$. They argue that the short-sales-constrained portfolios in this case are generalised to allow limits for short sales (negative weights) in

[5] To verify, just note that it is impossible to satisfy both restrictions if there is any negative value in the vector of weights x_i.

the portfolio. By constructing the modulus function, we could separate the positive and negative terms in the sum as

$$||x||_1 = \sum_{i=1}^{N} |x_i| = \sum_{x_i \in \mathbb{R}^+} x_i - \sum_{x_i \in \mathbb{R}^-} x_i. \tag{3.4.5}$$

Considering the fact that portfolio weights need to add up to one, we would then have

$$x'\iota = 1 = \sum_{x_i \in \mathbb{R}^+} x_i + \sum_{x_i \in \mathbb{R}^-} x_i, \tag{3.4.6}$$

and therefore

$$\sum_{x_i \in \mathbb{R}^+} x_i = 1 - \sum_{x_i \in \mathbb{R}^-} x_i. \tag{3.4.7}$$

Using both conditions and substituting equation (3.4.7) in (3.4.5) we obtain

$$||x||_1 = 1 - 2 \sum_{x_i \in \mathbb{R}^-} x_i. \tag{3.4.8}$$

For any $||x||_1 < \delta$, we rearrange the previous equation as follows:

$$- \sum_{x_i \in \mathbb{R}^-} x_i < \frac{\delta - 1}{2}. \tag{3.4.9}$$

In this expression, the left-hand side is the total short-selling weight of the portfolio. On the right-hand side, this position is restricted by $(\delta - 1)/2$. In this formulation, the investor may choose which assets for short sale as long as its total weights are kept under this limit. As we increase the amount of regularisation and decrease δ, the solution approaches the no-short-sales portfolio that could be interpreted as our target portfolio. Jagannathan and Ma (2003) prefer to interpret this portfolio as the one resulting from shrinking some of the covariance matrix elements, but the results are equivalent.[6]

The ℓ_2-constrained portfolio and naive diversification

It is also possible to consider the 2-norm-constrained portfolio by setting $p = 2$ in equation (3.4.2). In this case, our constraint would limit the portfolio weight's Euclidian norm in $\mathbb{R}^N$.

$$||x||_2 = \left(\sum_{i=1}^{N} |x_i|^2 \right)^{1/2} \tag{3.4.13}$$

[6] DeMiguel et al. (2009) show that it is possible to obtain the 1-norm-constrained portfolio (x_{NC1}) by solving the unconstrained problem using

$$\hat{\Sigma}_{NC1} = \hat{\Sigma} - \nu n \iota' - \nu \iota n'. \tag{3.4.10}$$

where $\hat{\Sigma}_{NC1}$ is the updated covariance matrix, ν is the Lagrangian multiplier for the 1-norm constraint in the constrained optimization, and $n \in \mathbb{R}^N$ is an indicator vector that tells which covariances should be shrunk: Its i^{th} element assumes a value of 1 if this asset is sold short in the 1-norm constrained optimization and 0 otherwise:

$$x_{NC1,i} < 0 \Rightarrow n_i = 1, \tag{3.4.11}$$

$$x_{NC1,i} > 0 \Rightarrow n_i = 0. \tag{3.4.12}$$

Equation (3.4.10) gives the 1-norm-constrained portfolio a moment shrinkage interpretation.

The problem then becomes

$$
\begin{aligned}
x^*_{\ell_p}(\hat{\Sigma}_t, \delta) \quad = \quad & \min_x \quad x^T \hat{\Sigma}_t x \\
& \text{s.t.} \quad x^T \iota = 1, \\
& \qquad \textstyle\sum_{i=1}^N |x_i|^2 \leq \delta,
\end{aligned}
\qquad (3.4.14)
$$

where we substitute $\delta = \acute{\delta}^2$ for analytical tractability.

Lastly, as shown in DeMiguel et al. (2009), we note that the restriction in (3.4.13) can be rewritten as

$$
\sum_{i=1}^{N} \left(x_i - \frac{1}{N} \right)^2 \leq \left(\delta - \frac{1}{N} \right). \qquad (3.4.15)
$$

As we decrease the value of δ,the selected portfolio tends to be closer to the equal-weighted portfolio because the 2-norm difference between the selected portfolio weights and the equal-weighted weights are constrained to be smaller than the difference on the right-hand side of equation (3.4.15). In the limit where $\delta = 1/N$, the resulting regularised portfolio is exactly the equal weighted one.

Chapter 4

Group information and asset allocation

Intuitively, securities that belong to a given class share some class-risk factors; therefore, further diversification can be achieved by including assets from different classes in a portfolio. Thus, information on how the assets are grouped is potentially useful in both the mean-variance and the inter-temporal asset allocation frameworks. In the latter, grouping information is especially important because different classes of securities may respond differently to shocks on the productivity of wealth. Despite the importance of dynamics and inter-temporal hedging in asset allocation between classes, as shown in Campbell and Viceira (2002), we focus here on the mean-variance framework.

Within this framework, we consider two different ways in which grouping information can be incorporated into the asset allocation problem. In both cases, the classification of the n assets into K groups of assets is assumed and denoted by the non-overlapping subsets $\mathcal{G}_k \subset \{1, \ldots, n\}$ containing each of the n_k indices of the assets belonging to group k, for $k = 1, \ldots, K$.

4.1 Two-step portfolio selection

A common practice to select portfolios in multi-class asset allocation problems is to divide the problems into two steps. In the first step, a within-class portfolio containing only assets from the each of the K classes is constructed using only estimates for the intra-class covariance matrix. In the second step, the across-class portfolio is selected containing only these synthetic class assets constructed in the first step.

In the first step, the investor solves an asset allocation problem for each of the K classes with the form

$$\begin{aligned} x^*_{\mathcal{G}_k}(\Sigma, \mathcal{G}_k, \gamma_k, \delta_k) \;=\; \arg\min_{x \in \mathbb{R}^{n_k}} \quad & x^T \Sigma_{\mathcal{G}_k, \mathcal{G}_k} x \\ \text{s.t.} \quad & x^T \iota = 1, \\ & \ell_{\gamma_k}(x) \leq \delta_k. \end{aligned} \tag{4.1.1}$$

For sufficiently large δ_k, $x^*_{\mathcal{G}_k}(\Sigma, \mathcal{G}_k, \delta_k)$ corresponds to the within-class minimum-variance portfolio for class k. Smaller values of δ_k result in within-class minimum-variance portfolios similar to the ones in DeMiguel et al. (2009) or Frost and Savarino (1988). The K portfolios arising from the K separate optimisation problems can also be written as

$$
\begin{bmatrix} x^*_{\mathcal{G}_1}(\Sigma, \mathcal{G}_1, \gamma_1, \delta_1) \\ x^*_{\mathcal{G}_2}(\Sigma, \mathcal{G}_2, \gamma_2, \delta_2) \\ \vdots \\ x^*_{\mathcal{G}_K}(\Sigma, \mathcal{G}_K, \gamma_K, \delta_K) \end{bmatrix} =
$$

$$
\begin{aligned}
&\arg\min_{x \in \mathbb{R}^N} && \begin{bmatrix} x_{\mathcal{G}_1} \\ x_{\mathcal{G}_2} \\ \vdots \\ x_{\mathcal{G}_K} \end{bmatrix}^T \begin{bmatrix} \Sigma_{\mathcal{G}_1,\mathcal{G}_1} & 0 & \cdots & 0 \\ 0 & \Sigma_{\mathcal{G}_1,\mathcal{G}_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{\mathcal{G}_m,\mathcal{G}_m} \end{bmatrix} \begin{bmatrix} x_{\mathcal{G}_1} \\ x_{\mathcal{G}_2} \\ \vdots \\ x_{\mathcal{G}_K} \end{bmatrix}, \\
&\text{s.t.} && \left\{ \begin{array}{rcl} x^T_{\mathcal{G}_j}\iota & = & 1, \\ \|x_{\mathcal{G}_k}\|_{\gamma_k} & \leq & \delta_k, \end{array} \right. \quad \text{for } j = 1, \ldots, m.
\end{aligned} \tag{4.1.2}
$$

Apart from a difference in the budget constraints, the asset allocation problem in (4.1.2)is similar to the problem of finding the minimum-variance portfolio under within-group constraints while completely ignoring the across-group covariances.

While the cross-class correlation information is intuitively important, ignoring it in this first step allows the investor to obtain more stable within-group minimum-variance portfolios using empirical data. This is because the same sample size is now used to estimate fewer elements in the covariance matrix, restricting the impact of estimation error to those elements only and bringing stability to the matrix estimate.

We denote the within class minimum-variance portfolios by $x^*_{\mathcal{G}_k}(\hat{\Sigma}, \mathcal{G}_k, \delta_k)$,which are obtained by plugging in the sample covariance matrix into the asset allocation problems in (4.1.1).

Across-group correlations are taken into account in the second step of the procedure. In this step, the investor decides how to allocate his wealth over the K portfolios built during the first step. Formally, the variance minimisation in the second step is given by

$$\begin{aligned} z^*(\Sigma, \mathbf{G}, \Gamma, \Delta) \quad = \quad & \arg\min_{z \in \mathbb{R}^K} \quad z^T \cdot \Sigma_{AG}\left(\Sigma, \mathbf{G}, \delta_1, \ldots, \delta_k\right) \cdot z \\ & \text{s.t.} \qquad z^T \iota = 1, \\ & \qquad\qquad \ell_\gamma(z) \leq \delta_0, \end{aligned} \tag{4.1.3}$$

where $\mathbf{G}$ denotes the set of subsets defining the grouping of assets, δ_0 is the across-group norm-constraint, Δ is a vector containing the norm-constraints $\delta_1, \ldots, \delta_K$ and $\Sigma_{AG}\left(\Sigma, \mathbf{G}, \delta_1, \ldots, \delta_k\right)$ is the covariance matrix of the returns of the assets built in step 1. Formally, Σ_{AG} is defined by

$$\begin{aligned} \left[\Sigma_{AG}\left(\Sigma, \mathbf{G}, \delta_1, \ldots, \delta_k\right)\right]_{k_1,k_2} \quad &: \quad = \\ x^*_{\mathcal{G}_{k_1}}(\Sigma, \mathcal{G}_{k_1}, \delta_{k_1})^T \cdot \Sigma_{\mathcal{G}_{k_1},\mathcal{G}_{k_2}} \cdot x^*_{\mathcal{G}_{k_2}}(\Sigma, \mathcal{G}_{k_2}, \delta_{k_2}), \text{ for } k_1, k_2 \quad &= \quad 1, \ldots K. \end{aligned}$$

Even though the investor is selecting the minimum-variance portfolio in this second step, he or she is constrained to choosing from the K

synthetic assets built during the first step. Here, this constraint may once again lead to more stable selection of portfolios given the fact that the smaller $K \times K$ across-group covariance matrix Σ_{AG} can be better estimated than the larger $n \times n$ asset covariance matrix Σ.

The advantages of this procedure include the low computational cost and the possibility of having unique solutions even when the number of assets is large relative to the sample size once the assets are split into smaller groups. Another advantage is the flexibility afforded by many "fudge" parameters. For instance, it is possible to use a given penalisation norm and value for each optimisation problem and to be flexible in the group structure.

4.2 Grouping constraints

The two-step procedure above incorporates group information into the selection of portfolios by completely ignoring the across-group correlation at first; then, the group structure in the second step is imposed when the across-group covariance is taken into account.

We now introduce a single-step procedure that incorporates the group information in the portfolio selection problem by means of group-inducing constraints. Such constraints were initially proposed in the statistical model selection literature in the form of penalties for performing structured variable selection in linear regression problems; this is seen in Zhao et al. (2009), Kim et al. (2006) or Yuan and Lin (2006). In such penalties/constraints, the different properties of ℓ_γ-norm penalised optimisers are exploited to induce different group behaviours on the penalised optimisers.

On one hand, the penalised optimisers for $0 < \gamma \leq 1$, tend to have many zero components (sparsity). On the other hand, the penalised optimisers for $2 \leq \gamma \leq \infty$ tend to concentrate on the diagonals (similarity). The CAP penalties in Zhao et al. (2009) combine these different behaviours of ℓ_γ-penalised optimisers to incorporate group information into linear regressors.

Taking the groups of assets $\mathcal{G}_1, \ldots, \mathcal{G}_K$ as given, the CAP penalty for portfolio selection is defined by first setting a vector of norm parameters $\Gamma = (\gamma_0, \gamma_1, \ldots, \gamma_K)$. The γ_0 norm-parameter is the across-group norm, and for $k = 1, \ldots, K$, the γ_k parameter is the within-group-k norm. Once the grouping and the Γ parameter are given, the CAP penalty could be computed as

$$\ell_\Gamma(x) := \ell_{\gamma_0}(N_\Gamma(x)) = \left[\sum_{k=1}^{K} \ell_{\gamma_k}(x_{\mathcal{G}_k})^{\gamma_0}\right]^{\frac{1}{\gamma_0}} .$$

The rationale is that γ_k determines how the weights within a group are related to one another. $\gamma_{k=1}$, for instance, promotes sparsity within group k. The γ_0 parameter then determines how the groups are related to one another: setting $\gamma_0 = 2$, for instance, would promote weights with balanced group-norms. For a more detailed discussion on the properties of CAP-penalised optimisers, we refer the reader to Zhao et al. (2009).

With a CAP penalty at hand, the group-constrained portfolio could then be defined as

$$\begin{aligned} x^*_{\ell_\Gamma}(\Sigma,\delta) \quad = \quad & \arg\min_{x\in\mathbb{R}^n} \quad x^T\Sigma x \\ & \text{s.t.} \qquad x^T\iota = 1, \\ & \qquad\qquad \ell_\Gamma(x) \leq \delta. \end{aligned} \tag{4.2.1}$$

In the follow section, we will focus on the CAP penalties having $\gamma_0 \in \{1,2\}$ and $\gamma_k = \gamma_1 \in \{1,2\}$ for all $k = 1,\ldots,K$. Intuitively, we have:

- $\gamma_0 = 1, \gamma_1 = \cdots = \gamma_k = 2$ leads to portfolios with balanced weights within a few selected groups;
- $\gamma_0 = 2, \gamma_1 = \cdots = \gamma_k = 1$ leads to portfolios with few assets from within a group and balanced weights across groups;
- $\gamma_0 = \gamma_1 = \cdots = \gamma_k = 2$ recovers the penalised portfolios introduced in DeMiguel et al. (2009);
- $\gamma_0 = \gamma_1 = \cdots = \gamma_k = 1$ recovers the penalised portfolios introduced in Jagannathan and Ma (2003);

Before moving on, we emphasise that the CAP penalty reduces to the ℓ_γ penalty discussed in DeMiguel et al. (2009) when $\gamma_0 = \gamma_k = \gamma$ for all $k = 0,\ldots,K$. Of particular interest are the cases where $\gamma_j \geq 1$ for all $j = 0,\ldots,m$. In those cases, the optimisation problem in (4.2.1) is convex; thus, efficient computational tools are available for computing large-scale constrained minimum-variance portfolios.

4.3 Empirical exercise

In this section, we apply the two-step and group-constraint approaches to construct a constrained minimum-variance portfolio from observed data. Next, we describe the data set that we use, the methodology and the results.

4.3.1 Fitted portfolios

Four penalised versions of the minimum variance portfolios were fitted to empirical data, namely:

1. The ℓ_1-penalised portfolio $x^*_{\ell_1}$

$$\begin{aligned} x^*_{\ell_1} \quad = \quad \arg\min_x \quad & x^T \hat{\Sigma} x \\ \text{s.t.} \quad & x^T \iota = 1, \text{ and} \\ & \|x\|_1 \leq \delta. \end{aligned} \tag{4.3.1}$$

2. The ℓ_2-penalised portfolio $x^*_{\ell_2}$

$$\begin{aligned} x^*_{\ell_2} \quad = \quad \arg\min_x \quad & x^T \hat{\Sigma} x \\ \text{s.t.} \quad & x^T \iota = 1, \text{ and} \\ & \|x\|_2 \leq \delta. \end{aligned} \tag{4.3.2}$$

3. The grouped $\ell_{1,2}$-penalised portfolio $x^*_{\ell_{1,2}}$

$$\begin{aligned} x^*_{\ell_{1,2}} &= \arg\min_x \quad x^T\hat{\Sigma}x \\ & \qquad \text{s.t.} \qquad x^T\iota = 1, \ \text{and} \\ & \qquad\qquad\qquad \left\|N_2^{\mathcal{G}}(x)\right\|_1 \leq \delta, \end{aligned} \tag{4.3.3}$$

   where

   - a given $\mathcal{G}$ group structure is given (sectors of the economy) involving K groups,
   - $N_2^{\mathcal{G}}(x)$ is a K dimensional-vector containing the ℓ_2 norm of the within group portfolios

$$N_2^{\mathcal{G}}(x) = \left[\begin{array}{cccc} \|x_{\mathcal{G}_1}\|_2 & \|x_{\mathcal{G}_2}\|_2 & \cdots & \|x_{\mathcal{G}_K}\|_2 \end{array} \right], \tag{4.3.4}$$

     with $x_{\mathcal{G}_k}$ denoting the weights of assets in group k.

4. The grouped ℓ_2, ℓ_1-penalised portfolio $x^*_{\ell_{2,1}}$

$$\begin{aligned} x^*_{\ell_{2,1}} &= \arg\min_x \quad x^T\hat{\Sigma}x \\ & \qquad \text{s.t.} \qquad x^T\iota = 1, \ \text{and} \\ & \qquad\qquad\qquad \left\|N_1^{\mathcal{G}}(x)\right\|_2 \leq \delta, \end{aligned} \tag{4.3.5}$$

   where

   - a given $\mathcal{G}$ group structure is given (sectors of the economy) involving K groups,

- $N_1^{\mathcal{G}}(x)$ is a K dimensional-vector containing the ℓ_1 norm of the within group portfolios

$$N_1^{\mathcal{G}}(x) = \left[\begin{array}{cccc} \|x_{\mathcal{G}_1}\|_1 & \|x_{\mathcal{G}_2}\|_1 & \cdots & \|x_{\mathcal{G}_K}\|_1 \end{array} \right], \tag{4.3.6}$$

with $x_{\mathcal{G}_k}$ denoting the weights of assets in group k.

Portfolio paths and selection criteria

For a fixed penalty function and observed data, a portfolio path is the set of solutions for all different values of the regularisation parameter δ. Given a portfolio path, a portfolio is selected according to the selection criteria that will be described in detail.

4.3.2 The data set

To obtain and compare the performance of the different portfolios, we use the monthly return data covering the period between January/1973 and April/2009. Our data set contained the 237 stocks that are part of the S&P500 index during this period. We group the assets into 9 sectors according to the ICB (Industry Classification Benchmark): basic materials, conglomerates, consumer goods, financial, health care, industrial goods, services, technology, and utilities.

Within this period, the data was divided into rolling windows 120 Because the data covered 433 months, portfolio estimates has 314 windows. For each method (penalty+selection criterion), the constrained portfolio fitted using the sample covariance matrix for returns r_{t-119} through r_t is used in period $t+1$.

4.3.3 Compared portfolios

We compare the performance of the following four penalised portfolios:

- ungrouped ℓ_1,
- ungrouped ℓ_2,
- sector grouped ℓ_1, ℓ_2 portfolio, and
- sector grouped ℓ_2, ℓ_1 portfolio.

For each constrained method, the regularisation parameter δ is selected according to three different criteria:

- Most constrained (MC) portfolio: this is obtained by setting $\delta$to be the minimum value for which a feasible portfolio satisfying $x^T\iota = 1$ exists. For the ℓ_2-constrained portfolio, MC portfolio completely disregards the data and is reduced to the naive diversification $\frac{1}{N}$-portfolio. For ℓ_1-constrained portfolios, the MC portfolio corresponds to the no-short-sale minimum variance constrained (MINC) portfolio.
- K-fold cross-validated (CV) portfolio: this portfolio is obtained by first splitting the fitting data (January/1973 to January/2001)

into K subsets (folds). In the k-th fold, the path of portfolios $x(\delta)$ is computed on a grid of values of δ using data points not in the k-th subsets of observations. The k-th subset of data points is used to estimate the out-of-sample variance for each portfolio $x(\delta)$ on the grid.on the grid, and the process is repeated K times. The K estimates of the out-of-sample variance of the portfolio $x(\delta)$ are averaged for each value of δ on the grid. The estimate then is chosen to be $x(\hat{\Sigma}, \delta^*)$, where δ^* is the one delta on the grid for which the mean out-of-sample variance is minimal.

If K equals the number of observations, this corresponds to using the jackknife/leave-one-out method to select a single portfolio from the portfolio path. In the exercise that follows, we set $K = 5$ as the number of cross-validation folds.

- The maximum return (MR) portfolio: this portfolio is obtained by selecting the portfolio on the regularisation path with the maximum return at the last observed data point (see DeMiguel et al. (2009)).

At time t, each method is applied to the data observed between times $t-w$ and t, and the fitted portfolio is used in time $t+1$, where w is a window size. For a given method (penalisation + selection criterion), the portfolio used at time $t+1$ is constructed by using the sample covariance matrix $\hat{\Sigma}_t$ computed with the data stretching from $t-w$ to

t. In the comparisons below, the window size was set to $w = 120$.

4.3.4 Evaluation criteria

Applying each method to the data produces a portfolio trajectory. Each trajectory is evaluated according to the following out-of-sample results calculated over the 120-month rolling window:

1. Variance:

$$\begin{aligned} \sigma_r^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(x_t^T R_t - \bar{r}\right)^2, \\ \bar{r} &= \frac{1}{T}\sum_{t=1}^{T} x_t^T R_t, \end{aligned} \tag{4.3.7}$$

2. Sharpe Ratio:

$$S = \frac{\bar{r}}{\sigma_r} \tag{4.3.8}$$

4.3.5 Significance of results

The statistical significance of the variance and Sharpe ratios differences are obtained by bootstrap when portfolio returns are not independently and identically distributed as a multivariate normal. Following DeMiguel et al. (2009), we compute the p-values for the Sharpe ratios using the bootstrapping methodology of Ledoit and Wolf (2008).This is recommended for financial time series that are generally serially correlated and exhibit volatility clustering.

We test the hypothesis that the Sharpe ratio of the return of portfolio i is equal to that of portfolio j:

$$H_0 : \mu_i/\sigma_i - \mu_j/\sigma_j = 0. \tag{4.3.9}$$

We report a two-sided p-value using the studentised circular block bootstrap of Ledoit and Wolf (2008), with $B = 1000$ bootstrap resamples and block size $b = 5$.

We also test the hypothesis that the variance of the returns of two portfolios are equal:

$$H_0 : \sigma_i^2 - \sigma_j^2 = 0. \tag{4.3.10}$$

For this test, we use the (nonstudentised) stationary bootstrap of Politis and Romano (1994) to construct a two-sided confidence interval for the difference using the same $B = 1000$ bootstrap resamples and block size $b = 5$. We then construct the p-values using the methodology in Ledoit and Wolf (2008).

4.3.6 Methodology

Because we only consider constraints constructed using the ℓ_1 and ℓ_2 norms, all portfolios we studied are defined as the solution to a convex optimisation problem. To fit the path of portfolios for a given constraint (and input data), we first compute the minimal value the constraint could assume over the $x^T\iota = 1$ hyperplane. Starting from this point (and corresponding value of δ) we progressively increase δ until the Lagrange multiplier of the constraint is numerically zero. To compute the initial value of δ and to compute each portfolio along the path, we used the `CVX` disciplined convex optimisation suite for Matlab (Grant and Boyd (2008) and Grant and Boyd (2011)). Although we are able to use this generic tool for this particular problem, larger problems involving thousands of assets may require exact-path following algorithms similar to those used in Osborne et al. (2000), Efron et al. (2004) or Brodie et al. (2009) (in a portfolio selection framework), or approximate algorithms (as seen in Zhao and Yu (2004) or Rosset (2004)) to be efficiently computed.

4.3.7 Empirical Results

Table 4.3.1 shows the out-of-sample variances of each portfolio: 1/N is the equal weighted portfolio, MINU is the minimum-variance unconstrained portfolio, and MINC is the minimum-variance short-sales constrained portfolio. Portfolios that ignore the group information are specific cases of portfolios that did not ignore the group information: ℓ_1 is obtained by penalising the ℓ_1 norm, and ℓ_2 is obtained by penalising the ℓ_2 norm. Both of these portfolios ignore the group structure and are equivalent to the 1-step $\ell_1 - \ell_1$ and $\ell_2 - \ell_2$ portfolios, respectively. Portfolios that used the group information are obtained in 1 or 2 steps. In general, $\ell_i - \ell_j$ is a portfolio that penalises the ℓ_i norm across groups and the $\ell_j$norm intra groups. The next entries correspond to the amount of penalisation needed to obtain the portfolios. In the table, CV corresponds to cross-validation; MR corresponds to "maximised return from the last period", MC stands for "most constrained" and MRL for "most relaxed".

Among the benchmarks, MINC is the portfolio that shows the lower out-of-sample variance; hence, this portfolio is chosen to be compared with the others . Therefore, the p-values in Table 4.3.1 refer to the test in equation 4.3.10 between the respective portfolio and MINC.

Table 4.3.1shows that we could not reject the hypothesis of equal variances for any of the portfolios constructed in 2 steps; the hypothesis also could not be rejected for most of the portfolios constructed in 1 step at the usual significance levels. In fact, all group regularised portfolios that use the information in the data to choose the amount of penalisation (CV and MR) has variances that are statistically equivalent to the benchmark.

Although the null hypothesis of equal variances could not be rejected because the series is too noisy , we could still analyse the estimated values. First, we note that CV portfolios that does not use the group information (i.e., the ones in DeMiguel et al. (2009))already have very low variances. These 1-step portfolios are the $\ell_2 - \ell_2$ CV (shrinking towards the 1/N, which is equivalent to a single penalisation on the

$\ell_2$norm chosen by cross-validation) and the $\ell_1 - \ell_1$ CV (which is equivalent to a single penalisation in the $\ell_1$norm chosen by cross-validation, shrinking the portfolio towards the MINC portfolio). The variances for $\ell_2 - \ell_2$ CV, and $\ell_1 - \ell_1$ CV were 1.068 and 1.096, respectively.

The other cross-validation portfolios obtained in 1 step had variances around these values; again, the portfolio with the lowest variance is the $\ell_1 - \ell_2$ CV, which tends to select equal weighted portfolios intra class and several economic sectors. The variance of this portfolio is marginally smaller than that of the previous portfolios: 1.067.We see that the variances of the 2-step portfolios are typically higher than those of the 1-step portfolios.

Table 4.3.1 suggests the following: if computational costs are not an issue, investors may use the 1-step solution penalising the $\ell_1 - \ell_2$, or the $\ell_2 - \ell_2$ norms as a very close second best solution. Only the first solution actually actively uses the grouping information: it tends to select a few sectors with equally weighted portfolios in each sector. The second one tends to allocate assets in an equal fashion, penalising the ℓ_2 portfolio norm. In both cases, there are marginal improvements over the MINC portfolio when using cross validation to select the amount of penalisation. The MINC portfolio could be used when computational costs are an issue given its simple implementation and good results.

Table 4.3.2 shows the Sharpe ratios of the portfolios that are created. Among the benchmarks, the 1/N portfolio corresponds to the

Benchmark portfolios				
1/N	1.978			
MINU	1.866			
MINC	**1.137**			
No group information: De Miguel et al. (2009)				
	Variance	p-values		
	L1			
CV	1.096	0.99		
MR	1.224	0.99		
	L2			
CV	1.068	1.00		
MR	1.483	0.80		
Group information				
	2-step		1-step	
	Variance	p-values	Variance	p-values
L1-L1				
CV	1.182	0.99	1.096	0.99
MR	1.318	0.96	1.224	0.99
MC	1.139	0.99	1.137	1.00
MRL	1.347	0.93	1.797	0.26
L1-L2				
CV	1.265	0.99	1.067	0.99
MR	1.254	0.97	1.402	0.91
MC	1.374	0.88	2.069	0.16
MRL	1.347	0.93	1.619	0.53
L2-L1				
CV	1.168	1.00	1.179	1.00
MR	1.262	0.99	1.345	0.95
MC	1.419	0.86	1.352	0.93
MRL	1.347	0.93	1.968	0.02
L2-L2				
CV	1.245	1.00	1.068	1.00
MR	1.520	0.77	1.483	0.80
MC	2.077	0.22	1.978	0.10
MRL	1.347	0.92	1.654	0.47

Table 4.3.1: Monthly out-of-sample variances of selected portfolios (x1000). P-values refer to the MINC portfolio

highest Sharpe ratio; therefore, it is chosen as the reference for the hypothesis tests given by equation 4.3.9. Similar to the variances, we cannot reject the null hypothesis of equality of Sharpe ratios at the usual significance levels in almost any case because the p-values are too high . In fact, we could only reject the null hypothesis at 10% for the MR 1-step $\ell_1 - \ell_2$ portfolio that has the same interpretation as before: selecting a few sectors of the economy but equally allocating the portfolio among the assets in each of the portfolios.

It is interesting to note that the use of group information results in portfolios with Sharpe ratios that are not worse than the ones in the literature; furthermore, the performance is significantly improved (at least in one particular norm penalization).

As we examine the estimated Sharpe ratios of the 1-step portfolios, we see that choosing the penalisation by maximising the last period return yields the highest values. However, cross validation yields better results with the 2-step portfolios. Comparing 2-step and 1-step portfolios, we see that 1-step portfolios tend to have higher Sharpe ratios, except in the $\ell_2 - \ell_1$ case.

Finally, we compare the use of the group structure in the estimated portfolios. We find that the original portfolios in DeMiguel et al. (2009), that ignore the group structure have Sharpe ratios that are higher than all 2-step portfolios but are statistically similar to the Sharpe ratios obtained in the 1-step portfolios. The MR $\ell_1 - \ell_2$ is the only portfolio that has a Sharpe ratio statistically higher than the benchmark; it is also higher than that of all portfolios that ignore the group structure. The best version of the $\ell_2 - \ell_1$ on the other hand[1], still has a Sharpe ratio that is lower than the ones obtained by ignoring the group structure.

An alternative that should increase the performance of portfolios using group information is to select better characteristics that are used to group assets in the first place. For stocks, characteristics such as the firm size, fundamentalist ratios, momentum and the proximity to major announcements could be used. It is also possible to use assets from different classes, such as commodities, stocks, bonds, and real estate and group them accordingly . The main idea is to have assets with similar behaviour grouped together. Considering only stocks and using their economic sector to group them is just one of many possibilities to improve portfolio performance.

4.3.8 Conclusion

In this section, we presented an alternative approach for the portfolio selection problem in the presence of estimation error in finite samples; this approach used the group structure of the asset. Like DeMiguel et al. (2009) and Brodie et al. (2009), we also chose to shrink the portfolio weights rather than the covariance matrix to improve the performance

[1] Choosing the amount of penalization that maximizes the returns of the last period.

Benchmark portfolios				
1/N	**7.67**			
MINU	5.45			
MINC	7.04			
No group information: De Miguel et al. (2009)				
	Variance	p-values		
	L1			
CV	5.88	0.65		
MR	9.75	0.67		
	L2			
CV	7.44	0.96		
MR	9.74	0.57		
Group information				
	2-step		1-step	
	Variance	p-values	Variance	p-values
L1-L1				
CV	9.14	0.72	5.88	0.65
MR	6.35	0.77	9.75	0.67
MC	7.79	0.96	7.04	0.89
MRL	7.90	0.96	8.89	0.86
L1-L2				
CV	8.02	0.93	6.58	0.80
MR	3.93	0.44	14.13	0.10
MC	1.43	0.11	8.20	0.81
MRL	7.90	0.95	6.82	0.88
L2-L1				
CV	9.64	0.54	3.49	0.34
MR	8.86	0.74	5.83	0.70
MC	8.84	0.41	6.93	0.75
MRL	7.89	0.96	7.45	0.97
L2-L2				
CV	8.87	0.72	7.44	0.96
MR	7.15	0.87	9.74	0.57
MC	7.88	0.65	7.59	0.27
MRL	7.90	0.95	5.50	0.72

Table 4.3.2: Monthly out-of-sample Sharpe ratios of selected portfolios (x100)

of the portfolio. We showed that our framework nests the portfolios in DeMiguel et al. (2009)as a special case, and these portfolios nest several others.

Empirically, we found that the use of the group structure could be beneficial and could provide an improvement over the existing benchmark portfolios. Specifically, we saw that the 1-step portfolios tend to have equal or better performance than portfolios that ignore the group structure. We also presented the 2-step portfolios with much lower computational costs, which generate more stable estimates with better out-of-sample performance when compared to some of the benchmarks used in the literature. The 2-step portfolios could be used as an alternative when computational costs are an issue.

Finally, we found that these results were true to our objective function, both in terms of the minimised variance and in terms of maximising the Sharpe ratio. This was achieved indirectly by reducing the variance of the portfolios and by choosing appropriate penalisation parameters.

Bibliography

Abel, A. (1989). Asset prices under heterogeneous beliefs: Implications for the equity premium. Rodney L. White Center for Financial Research Working Papers 9.

Baker, M. and J. Wurgler (2000). The equity share in new issues and aggregate stock returns. Journal of Finance 55(5), 2219–2257.

Bamber, L., O. Barron, and T. Stober (1999). Differential interpretations and trading volume. Journal of Financial and Quantitative Analysis 34(03), 369–386.

Basak, S. (2005). Asset pricing with heterogeneous beliefs. Journal of Banking and Finance 29(11), 2849–2881.

Böhm, V. and C. Chiarella (2005). Mean variance preferences, expectations formation, and the dynamics of random asset prices. Mathematical Finance 15(1), 61–97.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. Review of Economics and Statistics 72(3), 498–505.

Boswijk, H., C. Hommes, and S. Manzan (2007). Behavioral heterogeneity in stock prices. Journal of Economic Dynamics and Control 31(6), 1938–1970.

Brav, A., G. Constantinides, and C. Geczy (2002). Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. Journal of Political Economy 110(4), 793–823.

Brennan, M., E. Schwartz, and R. Lagnado (1997). Strategic asset allocation. Journal of Economic Dynamics and Control 21(8-9), 1377–1403.

Brock, W. and C. Hommes (1997). A rational route to randomness. Econometrica 65(5), 1059–1095.

Brock, W. and C. Hommes (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control 22(8-9), 1235–1274.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009). Sparse and stable markowitz portfolios. Proceedings of the National Academy of Sciences 106(30), 12267.

Campbell, J. (1987). Stock returns and the term structure. Journal of Financial Economics 18(2), 373–399.

Campbell, J. (1993, June). Intertemporal asset pricing without consumption data. American Economic Review 83(3), 487–512.

Campbell, J. (1996, April). Understanding risk and return. Journal of Political Economy 104(2), 298–345.

Campbell, J. (2000). Asset pricing at the millennium. Journal of Finance 55, 1515–1567.

Campbell, J., Y. Chan, and L. Viceira (2003). A multivariate model of strategic asset allocation. Journal of Financial Economics 67, 41–80.

Campbell, J., A. Lo, and A. MacKinlay (1997). The econometrics of financial markets, Volume 1. princeton University press Princeton, NJ.

Campbell, J. and R. Shiller (1988a). The dividend-price ratio and expectations of future dividends and discount factors. Review of financial studies 1(3), 195.

Campbell, J. and R. Shiller (1988b, July). Stock prices, earnings, and expected dividends. Journal of Finance 43(3), 661–76.

Campbell, J. and L. Viceira (1999). Consumption and portfolio decisions when expected returns are time varying. Quarterly Journal of Economics 114(2), 433–495.

Campbell, J. and L. Viceira (2001). Who should buy long-term bonds? American Economic Review 91(1), 99–127.

Campbell, J. and L. Viceira (2002). Strategic asset allocation: portfolio choice for long-term investors. Oxford University Press, USA.

Carhart, M. (1997). On persistence in mutual fund performance. Journal of finance 52(1), 57–82.

Chacko, G. and L. Viceira (2005). Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. Review of Financial Studies 18(4), 1369.

Chan, Y., L. Viceira, and J. Campbell (2003). A multivariate model of strategic asset allocation. Journal of Financial Economics 67, 41–80.

Chiarella, C., X.-Z. He, R. Dieci, and U. of Technology Sydney (2006, July). A dynamic heterogeneous beliefs capm. (181).

Constantinides, G. and D. Duffie (1996). Asset pricing with heterogeneous consumers. Journal of Political Economy 104(2), 219–240.

De Bondt, W. and R. Thaler (1985). Does the stock market overreact? Journal of Finance 40(3), 793–808.

DeMiguel, V., L. Garlappi, F. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science 55(5), 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies 22(5), 1915.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. The Annals of statistics 32(2), 407–499.

Epstein, L. and S. Zin (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. Econometrica, 937–969.

Epstein, L. and S. Zin (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. Journal of Political Economy 99(2), 263–286.

Fama, E. and K. French (1988a). Dividend yields and expected stock returns. Journal of Financial Economics 22(1), 3–25.

Fama, E. and K. French (1988b). Permanent and temporary components of stock prices. Journal of Political Economy 96(2), 246–273.

Fama, E. and K. French (1993). Common risk factors in the returns on stocks and bonds. Journal of financial economics 33, 3–56.

Fama, E. and K. French (1996). Multifactor explanations of asset pricing anomalies. Journal of Finance 51(1), 55–84.

Fama, E. and R. Kenneth (1989). French, 1989, business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.

Frank, I. and J. Friedman (1993). A statistical view of some chemometrics regression tools. Technometrics 35(2), 109–135.

Frankel, J. and K. Froot (1987, March). Using survey data to test standard propositions regarding exchange rate expectations. American Economic Review 77(1), 133–53.

Frost, P. and J. Savarino (1986). An empirical bayes approach to efficient portfolio selection. Journal of Financial and Quantitative Analysis 21(03), 293–305.

Frost, P. and J. Savarino (1988). For better performance: Constrain portfolio weights. Journal of Portfolio Management 15(1), 29–34.

Giovannini, A. and P. Weil (1989, 1989). Risk aversion and intertemporal substitution in the capital asset pricing model. Working Paper 2824, National Bureau of Economic Research.

Glosten, L., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48(5), 1779–1801.

Grant, M. and S. Boyd (2008). Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura (Eds.), Recent Advances in Learning and Control, Volume 371 of Lecture Notes in Control and Information Sciences, pp. 95–110. Springer Berlin / Heidelberg.

Grant, M. and S. Boyd (2011). Cvx: Matlab software for disciplined convex programming, version 1.21.

Green, R. and B. Hollifield (1992). When will mean-variance efficient portfolios be well diversified? Journal of Finance 47(5), 1785–1809.

Harvey, C. (1991). The world price of covariance risk. Journal of Finance 46(1), 111–157.

Hillebrand, M. and J. Wenzelburger (2006, August). The impact of multiperiod planning horizons on portfolios and asset prices in a dynamic capm. Journal of Mathematical Economics 42(4-5), 565–593.

Hodrick, R. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. Review of Financial studies 5(3), 357.

Hoerl, A. and R. Kennard (1970). Ridge regression: applications to nonorthogonal problems. Technometrics 12(1), 69–82.

Horst, U. and J. Wenzelburger (2008). On non-ergodic asset prices. Economic Theory 34(2), 207–234.

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. Journal of Finance 58(4), 1651–1684.

James, W. and C. Stein (1961). Estimation with quadratic loss. In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability: held at the Statistical Laboratory, University of California, June 20-July 30, 1960, pp. 361. Univ of California Press.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance 48(1), 65–91.

Jobson, J. and B. Korkie (1980). Estimation for markowitz efficient portfolios. Journal of the American Statistical Association 75(371), 544–554.

Jobson, J., B. Korkie, and V. Ratti (1979). Improved estimation for markowitz portfolios using james-stein type estimators. In Proceedings of the American Statistical Association, Business and Economics Statistics Section, Volume 41, pp. 279–284.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. Journal of Financial and Quantitative Analysis 21(03), 279–292.

Kandel, E. and N. Pearson (1995). Differential interpretation of public signals and trade in speculative markets. Journal of Political Economy 103(4), 831–872.

Keim, D. and R. Stambaugh (1986). Predicting returns in the bond and stock markets. Journal of Financial Economics 17, 357–390.

Kim, Y., J. Kim, and Y. Kim (2006). Blockwise sparse regression. Statistica Sinica 16(2), 375.

Kurz, M. and A. Beltratti (1996). The equity premium is no puzzle. Stanford University Dept. of Economics WP# 96-004.

Lamont, O. (1998, October). Earnings and expected returns. Journal of Finance 53(5), 1563–1587.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis 88(2), 365–411.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the sharpe ratio. Journal of Empirical Finance 15(5), 850–859.

Lewellen, J. (1999). The time-series relations among expected return, risk, and book-to-market. Journal of Financial Economics 54(1), 5–43.

Lucas, R. J. (1978). Asset prices in an exchange economy. Econometrica, 1429–1445.

Ludvigson, S. and M. Lettau (1999). Consumption, aggregate wealth and expected stock returns. Staff Reports.

Ludvigson, S. and M. Lettau (2004). Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption. American Economic Review 94(1), 279–299.

Markowitz, H. (1952). Portfolio selection. Journal of Finance 7, 77–91.

Mehra, R. and E. Prescott (1985). The equity premium: A puzzle. Journal of monetary Economics 15(2), 145–161.

Menkhoff, L. (1997). Examining the use of technical currency analysis. International Journal of Finance & Economics 2(4), 307–318.

Merton, R. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. Review of Economics and Statistics 51(3), 247–257.

Merton, R. (1971, December). Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory 3(4), 373–413.

Michaud, R. (1989). The markowitz optimization enigma: is' optimized'optimal? Financial Analysts Journal 45(1), 31–42.

Milgrom, P. and N. Stokey (1982). Information, trade and common knowledge. Journal of Economic Theory 26(1), 17–27.

Osborne, M., B. Presnell, and B. Turlach (2000). A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis 20(3), 389.

Patton, A. and A. Timmermann (2010, October). Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion. Journal of Monetary Economics 57(7), 803–820.

Politis, D. and J. Romano (1994). The stationary bootstrap. Journal of the American Statistical Association 89(428), 1303–1313.

Poterba Lawrence, H. and M. James (1988). Mean reversion in stock prices: Evidence and implications. Journal of Financial Economics 22(1), 27–59.

Restoy, F. (1992). Optimal portfolio policies under time-dependent returns. Working Paper, Banco de España.

Rosset, S. (2004). Tracking curved regularized optimization solution paths. In in Advances in Neural Information Processing Systems (NIPS*2004. MIT Press.

Samuelson, P. (1969). Lifetime portfolio selection by dynamic stochastic programming. Review of Economics and Statistics 51(3), 239–246.

Shiller, R. (2002). Bubbles, human judgment, and expert opinion. Financial Analysts Journal, 18–26.

Taylor, M. and H. Allen (1992). The use of technical analysis in the foreign exchange market. Journal of International Money and Finance 11(3), 304–314.

Weil, P. (1989). The equity premium puzzle and the risk-free rate puzzle. Journal of Monetary Economics 24(3), 401–421.

Wenzelburger, J. (2004). Learning to predict rationally when beliefs are heterogeneous. Journal of Economic Dynamics and Control 28(10), 2075 – 2104.

Wold, H. (1975). Soft modeling by latent variables: the nonlinear iterative partial least squares approach. London: Academic Press.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68(1), 49–67.

Zhao, P., G. Rocha, and B. Yu (2009). The composite absolute penalties family for grouped and hierarchical variable selection. The Annals of Statistics 37(6A), 3468–3497.

Zhao, P. and B. Yu (2004). Boosted lasso. Technical report, Journal of Machine Learning Research.