Three Essays on Macro-Finance: Robustness and Portfolio Theory
Pedro Henrique Engel Guimarães

Three Essays on Macro-Finance: Robustness and Portfolio Theory

Tese submetida à Escola de Pós-Graduação em Economia como requisito parcial para a obtenção do título de Doutor em Economia.

Área de Concentração: Finanças

Orientador: Caio Ibsen Kleinman Rodrigues de Almeida

Rio de Janeiro
2017
Guimarães, Pedro Henrique Engel


115 f.

Tese (Doutorado) - Fundação Getulio Vargas, Escola de Pós-Graduação em Economia.

Orientador: Caio Ibsen Kleinman Rodrigues de Almeida.

Inclui bibliografia.

PEDRO HENRIQUE ENGEL GUIMARÃES

“THREE ESSAYS IN MACRO-FINANCE: ROBUSTNESS, AND PORTFOLIO THEORY”


Data da defesa: 28/07/2017

Aprovada em:

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

Caiom Ibsen Rodrigues de Almeida
Orientador (a)

Carlos Eugenio Ellery Lustosa da Costa

Luís Henrique Bertolino Braido

André Portela dos Santos

Bruno Cara Giovannetti
Abstract

This doctoral thesis is composed of three chapters related to portfolio theory and model uncertainty. The first paper investigates how ambiguity averse agents explain the equity premium puzzle for a large group of countries including both Advanced Economies (AE) and Emerging Markets (EM). In the second article, we develop a general robust allocation framework that is capable of dealing with parametric and non parametric asset allocation models. In the final paper, I investigate portfolio selection criteria and analyze a set of portfolios out of sample performance in terms of Sharpe ratio (SR) and Certainty Equivalent (CEQ).

Keywords: Risk aversion; Model Uncertainty; Equity premium puzzle; Detection error probability; Costs of model uncertainty; Advanced Economies; Emerging Market
List of Figures

1  Detection error probabilities estimates using the full sample ($\beta_{AE} = \beta_{EM} = 0.995$). .................................................. 43

2  Detection error probabilities estimates using the homogeneous sample ($\beta_{AE} = \beta_{EM} = 0.995$). .................................................. 44

3  Welfare gains from eliminating risk and model uncertainty between 1999 and 2015 ($\beta_{AE} = \beta_{EM} = 0.995$ and $p(\theta^{-1})$ estimated). .................................................. 45

4  Ratio between welfare gains from eliminating only model uncertainty to gains from eliminating only risk between 1999 and 2015 ($\beta_{AE} = \beta_{EM} = 0.995$ and $p(\theta^{-1})$ estimated). .................................................. 46

5  Welfare gains from eliminating risk and model uncertainty using $\beta_{EM} = 0.989$ (full sample and $p(\theta^{-1})$ estimated). .................................................. 47

List of Tables

1  Description of the full data set. .................................................. 48

2  Risk Aversion and Detection Error Probability (Full Sample) .............. 49

3  Risk Aversions and Detection Error Probabilities Coefficient of Variation (Full Sample) .................................................. 50

4  Description of the homogeneous sample data set. ................................ 51

5  Risk Aversion and Detection Error Probability (Homogeneous Sample) ... 52

6  Risk Aversions and Detection Error Probabilities Coefficient of Variation (Homogeneous Sample) .................................................. 53

7  Risk Aversion and Detection Error Probability using $\beta_{EM} = 0.989$ (Homogeneous Sample) .................................................. 54
Contents

1 Introduction 10

2 Risk Aversion or Model Uncertainty? An Empirical Cross-Sectional Analysis Across Countries 12

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2.2 The model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

2.2.1 The transition equation . . . . . . . . . . . . . . . . . . . . . . . . . 15

2.2.2 The agent preferences . . . . . . . . . . . . . . . . . . . . . . . . . . 17

2.2.3 The agent constraints . . . . . . . . . . . . . . . . . . . . . . . . . . 18

2.2.4 The value function . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

2.2.5 The minimizing martingale increment . . . . . . . . . . . . . . . . . . 21

2.2.6 The discounted entropy . . . . . . . . . . . . . . . . . . . . . . . . . 22

2.2.7 The value function for random walk log consumption . . . . . . . . . 22

2.3 The data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

2.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25

2.4.1 Risk and Model Uncertainty Aversion . . . . . . . . . . . . . . . . . . . 25

2.4.2 Welfare Costs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

2.4.3 Robustness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34

2.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34

3 Robust Asset Allocation Problems 55

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

3.2 The Traditional Optimization Problem . . . . . . . . . . . . . . . . . . . . 57
1 Introduction

This doctoral thesis is composed of three chapters related to portfolio theory and model uncertainty. Chapter 1 discusses the equity premium puzzle introduced by Mehra and Prescott. They show that, considering agents with constant relative risk averse preferences, the premium of average return on stocks are too high relative to the risk free asset. That is, they find that the coefficient of risk aversion that would imply this risk premium should be higher than what is expected from economists. Campbell extends this finding by considering a larger set of countries and showing that the risk premium persists for different economies. He also shows that there is a high degree of heterogeneity in the cross section which levels up even further the task of explaining these findings. We consider the alternative framework proposed by Barillas, Hansen and Sargent of uncertainty averse investors. Their idea is that the agents do not know for sure the true distribution of future realizations, and for this reason they incorporate this uncertainty in their preferences by considering the worst case scenario. Barillas et al. (2009) calibrate the model for the US and find that reasonable amounts of model uncertainty can be equivalent to high levels of risk aversion when considering agents’ choices and effects on asset prices. Okubo (2015) expands this analysis to developed economies. According to Okubo (2015), there is also a considerable amount of variability in the model uncertainty aversion levels among developed countries. We build a new data set and expand this work for a large group of countries including both Advanced Economies (AE) and Emerging Markets (EM). As for the US economy, we find that the level of model uncertainty needed to explain the data for most countries remain on the reasonable bound argued by Barillas et al. Additionally, we also find that the levels of model uncertainty present high variability among countries. This dispersion was also present in estimates of the coefficient of RRA of Campbell (2003) and the model uncertainty aversion parameters of Okubo (2015). However, we show that the variability of model uncertainty parameters is much lower than that of the RRA parameters for our series.

The second chapter deals with robust asset allocation problems. The main contribution in this section is to show how a more general theory on robust decision problems applies in the particular case of a portfolio selection problem where agents only consider
the first two moments of the return distribution and take the prior distribution of returns to be Gaussian. We show how robust optimization theory can be used to deal with estimation error and model misspecification, and how available information can be used to improve efficiency by imposing correct constraints in the model. In particular, we derive a covariance constrained robust portfolio model and show that it fits in the class of shrinkage estimators. The results are intuitive. When the agent is uncertainty averse and there is no restrictions in the moments of the worst case distribution, we find that the distorted worst case scenario will keep normality, but with a decreased mean and enlarged variance. On the other hand, when the covariance is constrained to be equal to the prior distribution (e.g., when there is enough precision in the second moment estimation), all uncertainty is dealt to the mean, and the a posteriori distribution will be normal with equal variance, but decreased mean. The size of the diminished mean will depend on the agents’ level of uncertainty aversion. We also show that if there is no uncertainty aversion, that is, the agent has full confidence in the return distribution, the robust allocation will reverse to the traditional mean variance portfolio case. On the other hand, if the agent is extremely uncertain about the return distribution, the estimated mean is considered uninformative, and the robust portfolio converges to the minimum variance portfolio. We also show that this framework is capable of dealing with parametric and non parametric asset allocation models.

The third and concluding chapter deals with portfolio selection criteria. Portfolio selection problems are usually based on a single criterion. The main criteria used in the literature to compare portfolio performance are: out of sample performance in terms of Sharpe ratio (SR) and Certainty Equivalent (CEQ). Other criteria also commonly used are minimization of expected loss, usually achieved by using shrinkage estimators and optimization of conditional moments of return distributions with Bayesian statistics. We explore portfolio models with these different criteria and show that in many ways they point to different directions. In particular, we show that when only considering out of sample performance, none are significantly better than the naive 1/N portfolio as suggested by Garlappi Uppal. On the other hand, we show that if in sample data are to be considered, the 1/N portfolio performs significantly worse than expected. One way to overcome this issue is to combine the 1/N portfolio with the minimum variance portfolio.
2 Risk Aversion or Model Uncertainty? An Empirical Cross-Sectional Analysis Across Countries

Abstract

By analyzing a macro panel data including both Emerging Markets (EM) and Advanced Economies (AE), we show that a reasonable level of model misspecification helps to explain the equity premium existing in these markets. Unsurprisingly, given a fixed common time slot, we also show that fear for model misspecification is in general higher for EMs than it is for AEs. Moreover, the degree of cross-sectional heterogeneity across countries’ estimates is smaller under model ambiguity than when considering the traditional CRRA preference. In addition, we also compute the separate cost effects of risk and ambiguity for these economies in terms of present consumption, and conclude that the most significant effects come from model ambiguity.

2.1 Introduction

There is no doubt that one of the main open questions in the financial literature is the equity premium puzzle introduced by Mehra and Prescott (1985). The combination of high historical average stock return and low risk-free interest rate demands a very high coefficient of relative risk aversion (RRA) to match the historical consumption growth process of the United States. The literature also shows that this debate is not a singularity of the American economy. Campbell (2003) documents high values of implied RRA coefficient for different developed economies, expanding the puzzle to other economies. He also finds a high degree of variability among the coefficients of these countries, making the interpretation of the puzzle using the traditional models even more challenging\(^1\).

Based on the framework proposed by Anderson et al. (2003) and Hansen and Sargent (2008), Barillas et al. (2009) reinterpret the RRA coefficient of Tallarini (2000) into a parameter that controls for model uncertainty aversion. Their interpretation for the agents’ preferences brings new insights to the problem. The idea is that when agents

\(^1\)For a detailed revision of the literature in the Equity Premium Puzzle see ? and the references therein.
choose their consumption path, they are uncertain about possible future states and fear this uncertainty by considering worst case scenarios.

Barillas et al. (2009) calibrate the model for the US and find that reasonable amounts of model uncertainty can be equivalent to high levels of risk aversion when considering agents’ choices and effects on asset prices. Okubo (2015) expands this analysis to developed economies. He finds out that there is also a considerable amount of variability in the model uncertainty aversion levels among developed countries.

Since so far the analysis was focused on the US economy and little was done to evaluate the validity of the model within a broad empirical perspective, we build a new data set and expand this work for a large group of countries including both Advanced Economies (AE) and Emerging Markets (EM). The idea is to test if the model is suitable to explain economic behavior, in particular the equity premium, for a broad set of economies or if there is a feature that makes the model suitable only for some particular set of countries. Moreover, we would like to know if the model is able to capture the expected higher level of uncertainty present in Emerging Markets. The main results are summarized below.

We find that the model uncertainty needed to explain the data for most countries remains on the reasonable bound. Additionally, we also find that the levels of model uncertainty present high variability among countries. This dispersion was also present in the estimates of the coefficient of RRA of Campbell (2003) and the model uncertainty aversion parameters of Okubo (2015). However, we show that the variability of model uncertainty parameters is much lower than of the RRA parameters when we restrict attention to detection error probabilities.

Barillas et al. (2009) also used this framework to reinterpret the large welfare gains eliminating risk found by Tallarini (2000). As Lucas (1987), they show that the elimination of risk only provides a small welfare benefit. According to them, a significant share of the gains found by Tallarini (2000) comes from reducing model uncertainty.\textsuperscript{2} Relying on this

\textsuperscript{2}The welfare gain is computed by considering risk as the exogenous volatility of the consumption path. To compare gains from eliminating risk and model uncertainty we need to evaluate the agent preferences on both the exogenous stochastic consumption path and its deterministic trajectory with the same mean but zero variance. To evaluate the gains from eliminating risk alone, we must do the same procedure, but now setting uncertainty aversion to zero, which can be done by letting $\theta \to \infty$. 

13
framework, we show that this result is also valid for other countries. Most of the welfare gains come from eliminating model uncertainty. In some cases, the benefits are more than 30 times larger than that of only reducing risk.

An apparent counterintuitive result we find from the data is that some Emerging Market countries like Brasil, Mexico, and India bear less risk and model uncertainty in absolute terms than other Advanced Economies like the US, France, and Denmark when considering the full sample. Those results can be attributed to the different time slot of data available for these economies. While most of the Advanced Economies have a long time series data available, Emerging Markets have a restricted sample that starts from 1999. When we restrict the analyses to an homogeneous dataset, all starting from 1999, the results are partially reversed.

The only unexpected result comes from Brazil. We would expect a high level of uncertainty for Brazil but observe the contrary. These results are in line with recent studies that suggest no equity premium in countries like Brazil. One possible explanation for the observed small level of uncertainty in Brazil is that while the interest rate is relatively high, the high volatility in the stock market makes it difficult to evaluate the equity premium when observing only a small time series data. Other possibility is that there are some possible characteristic present in the Brazilian economy that makes the equity premium small despite all uncertainty present. These characteristics are not captured by the model and further investigation is required.

Finally, we do a robustness exercise using an alternative method for calibrating the discount factor of EM and find similar results.

The rest of the paper is organized as follows. Section 2 develops the base model in detail for the interested reader. Section 3 describes the data. In Section 4 we explain the details of the exercises and show the main results. Section 5 concludes.

---

2.2 The model

In this section, we introduce the model by describing the agents, characterized by their objective function and their constraints. The model is dynamic in the sense that the agent’s decision depends on history. The idea is that there is a transition equation that can be derived from equilibrium. This transition equation depends on parameter preferences. Since the preference is well defined by its parameters, we can compute a counter-factual with a different parameter of interest. Using this procedure, we can compute how much an agent would be willing to give up on consumption to avoid undesired uncertainties.

2.2.1 The transition equation

Many econometric papers have already shown that it is difficult to refuse the hypothesis that aggregate log consumption follows a geometric random walk. As used in Tallarini (2000), and later on Barillas et al. (2009), we use one of the following consumption plans:

1. geometric random walk:

\[ c_t = c_0 + t \mu + \sigma \varepsilon (\varepsilon_t + \varepsilon_{t-1} + \ldots + \varepsilon_1), \ t \geq 1 \]

2. geometric trend stationary:

\[ c_t = \rho^t c_0 + t \mu + \sigma \varepsilon (\varepsilon_t + \rho \varepsilon_{t-1} + \ldots + \rho^{t-1} \varepsilon_1), \ t \geq 1 \]

where \( \varepsilon_t \sim i.i.d. N(0,1) \) and \( c_t = \log C_t \).

It is not difficult to see that all these consumption plans are particular cases of the more general multivariate formulation

\[ x_{t+1} = Ax_t + B \varepsilon_{t+1} \]

\[ c_t = Hx_t \]

where \( \varepsilon_{t+1} \sim i.i.d. N(\mu, \Sigma) \) with dimension \( m \times 1 \), \( x_t \) is an \( n \times 1 \) state vector, and the eigenvalues of \( A \) are bounded in modulus by \( \frac{1}{\sqrt{\beta}} \).
Note that this representation implies that the time $t$ element of the consumption plan can be expressed as the following function of $x_0$ and the history of shocks

$$c_t = H(B \epsilon_t + AB \epsilon_{t-1} + ... + A^{t-1} B \epsilon_1) + H A^t x_0$$  \hspace{1cm} (1)

We let $C(A, B, H; x_0)$ denote the set of consumption plans with the broad representation above.

As we can see, the transition path described lacks a vector of control variables, say $u_t$. In the absence of a control, we should understand this consumption path as an equilibrium consumption path. To clarify this idea, suppose the agent solves an optimization problem where he must choose a sequence of controls $\{u_t\}$ in order to maximize a discounted utility that is a function of the consumption path. Suppose also that the transition equation faced by the consumer can be written as

$$x_{t+1} = \hat{A} x_t + \tilde{A} u_t + B \epsilon_{t+1}$$

where $x_t$ is the state vector, $u_t$ is the control vector, and $\epsilon_{t+1}$ is the random vector of shocks.

If the optimal control for the agent can be written as $u_t = -Fx_t$, the equilibrium path for consumption would be

$$x_{t+1} = \hat{A} x_t + \tilde{A} u_t + B \epsilon_{t+1} = (\hat{A} - \tilde{A} F) x_t + B \epsilon_{t+1} = A x_t + B \epsilon_{t+1}$$

Where $A = \hat{A} - \tilde{A} F$.

From here on we will consider that the representative agent is optimizing for his consumption path in the sense presented above. The only layer of optimization will be carried out by nature. The idea is that in order to incorporate model uncertainty we consider an agent that is averse to ambiguity in the sense that his planning problem is distorted by a malevolent agent played by nature. The only role of nature is to distort the distribution of future consumption to the worst possible scenario.
2.2.2 The agent preferences

To incorporate the possibility of mistakes in the model, we consider here an agent with multiplier preferences. An agent is said to have multiplier preferences if his preference ordering over $C$, the set of consumption plans whose time $t$ elements $c_t$ are measurable functions of $(\varepsilon_t, x_0)$, is described by

$$\min \sum_{t=0}^{\infty} \mathbb{E}\{\beta^t G_t[c_t + \beta \theta E(g_{t+1} \log g_{t+1}|\varepsilon^t, x_0)]|x_0]\}$$

(2)

s.t.

$$G_{t+1} = g_{t+1} G_t, \quad \mathbb{E}[g_{t+1}|\varepsilon^t, x_0] = 1, \quad g_{t+1} \geq 0, \quad G_0 = 1$$

Where $g_{t+1}$ is a positive measurable function of the history $X_t = \{(\varepsilon^t, x_0)\}$, $c_t = \log C_t$, $\varepsilon_{t+1} \sim i.i.d.N(\mu, \sigma^2_{\varepsilon})$, and $\varepsilon^t = (\varepsilon_t, ..., \varepsilon_1)$. We restrict attention only to the subset $C(A, B, H; x_0)$ of $C$ described by equation (1). Since this formula seems complicated at first sight, it is usually helpful to break it in parts. Think about the total "welfare", $W$, of the agent as the sum of expected discounted (instantaneous) utilities $U_t$ that the agent obtain from consuming the bundle $c_t$ at time $t$; that is, consider

$$W = \sum_t U_t$$

$$U_t = \beta^t \{\mathbb{E}[G_t c_t|x_0] + \beta \theta E[G_t E(g_{t+1} \log g_{t+1}|\varepsilon^t, x_0)|x_0]\}$$

Defining the distorted expectation of a random variable $X_t$ as $\tilde{E}[X_t] = E[G_t X_t]$, we have that $U_t$ is the discounted sum of two terms:

$$\tilde{E}[c_t|x_0] \text{ and } \theta \tilde{E}[\beta E(g_{t+1} \log g_{t+1}|\varepsilon^t, x_0)|x_0].$$

The first term is just the expected utility of consumption under the worst case scenario. The second term is the expected value of discounted conditional entropy times a parameter $\theta$.

The conditional entropy is the function $E(g_{t+1} \log g_{t+1}|\varepsilon^t, x_0)$, which depends on the history of shocks $\varepsilon^t = (\varepsilon_t, ..., \varepsilon_1)$ and the initial condition $x_0$. The parameter $\theta$ measures
the degree of concern of the agent with respect to model uncertainty. It can be understood as a parameter that constraints the choice set of the distorting sequence $g_{t+1}$ chosen by the nature. Note that if $\theta = \infty$ then $g_{t+1} = 1$ for all $t$, and $U_t = E[c_t]$; that is, the agent has no concern for model specification and his preference ordering is given by the usual representation

$$W(x_0) = \sum_{t=0}^{\infty} \beta^t E\{c_t|x_0\}$$

The entropy is only one possible way to restrict the distribution choices considered by the agent. A more detailed description of entropy and its properties is given below. But first we explain the nature control variable $g_{t+1}$ and its space restrictions.

2.2.3 The agent constraints

To better understand the meaning of $g_{t+1}$ and of the constraints

$$G_{t+1} = g_{t+1}G_t, \quad E[g_{t+1}|\varepsilon^t, x_0] = 1, \quad g_{t+1} \geq 0, \quad G_0 = 1$$

it is useful to think of $g_{t+1}$ as the ratio of two densities that represents a likelihood ratio. Let $f_{t+1}$ be the expected density for next period realization of the exogenous shock $\varepsilon_{t+1}$, and $\tilde{f}_{t+1}$ be its worst case distribution conditional on date $t$ information. With these assumptions, we can write $g_{t+1} = \frac{\tilde{f}_{t+1}}{f_{t+1}}$. Since $\tilde{f}_{t+1}$ and $f_{t+1}$ are both density functions, we must have $g_{t+1} \geq 0$ and $E[g_{t+1}|\varepsilon^t, x_0] = 1$ for all $t$, where the expectation is taken under the expected density $f_{t+1}$. Now, since the density function of $\varepsilon_{t+1}$ is a measurable function of the history $X_t = \{(\varepsilon^t, x_0)\}$, so must be $g_{t+1}$. It is easy to see that $E(g_{t+1} \log g_{t+1}|\varepsilon^t, x_0) = 0$ only if $g_{t+1} = 1$ which means $\tilde{f}_{t+1} = f_{t+1}$, so that the conditional distribution is exactly the expected by the agent.

Now, to take into account the dynamic nature of the problem, it is common to factor a joint density $F_{t+1}$ over an $X_{t+1}$-measurable vector as $F_{t+1} = f_{t+1}F_t$, where $f_{t+1}$ is a one step ahead density conditioned on $X_t$. Following Hansen and Sargent, we factor a random
variable $G_{t+1}$, forming:

$$gt+1 = \begin{cases} 
\frac{G_{t+1}}{G_t} & \text{if } G_t > 0 \\
1 & \text{if } G_t = 0.
\end{cases}$$

Then $G_{t+1} = gt+1 G_t$ and

$$G_t = G_0 \prod_{j=1}^{t} g_j$$

The random variable $G_0$ is set to unity so that $\{G_t : t \geq 0\}$ is a martingale. By construction, $G_t$ is a function of $\varepsilon^t$ and $x_0$ and $E[G_{t+1}|\varepsilon^t, x_0] = G_t$. We then have $E[G_t|x_0] = G_0 = 1$.

With all of this said, we can understand these restrictions as technical conditions to ensure that nature choices are perturbations that are in fact changes of measure. In particular, the optimal choice of nature will be the worst possible measure for the agent. The worst measure is computed by restricting the choices of nature using entropy as a discrepancy measure. We left in the appendix a brief description of entropy and its properties that makes it useful as a discrepancy measure.

### 2.2.4 The value function

The value function associated with the multiplier agent preferences solves the following Bellman equation:

$$GW(x) = \min_{g \geq 0, E g = 1} G \left( c + \beta \int (g(\varepsilon)W(Ax + B\varepsilon) + \theta g(\varepsilon) \log g(\varepsilon))\pi(\varepsilon)d\varepsilon \right)$$

The Bellman equation is a way to transform the sequential problem into a problem in which the agent only needs to solve for a 2 period time schedule. The idea is that solving for the entire sequence is equivalent to choose the best choice for today and the best choice for tomorrow supposing that the agent is optimizing for the remainder.

Dividing by $G$ gives

$$W(x) = c + \min_{g \geq 0} \beta \int [g(\varepsilon)W(Ax + B\varepsilon) + \theta g(\varepsilon) \log g(\varepsilon)]\pi(\varepsilon)d\varepsilon$$

(3)
s.t. 

\[ Eg = 1 \]

By solving this problem, we can express \( W(x) \) as the sum of two components, the first of which is the expected discounted value of log consumption under the worst case scenario, while the second is \( \theta \) times discounted entropy:

\[ W(x) = J(x) + \theta N(x) \quad (4) \]

where

\[ J(x) = c + \beta \int [\hat{g}(\epsilon)J(Ax + B\epsilon)]\pi(\epsilon)d\epsilon \quad (5) \]

and

\[ N(x) = \beta \int [\hat{g}(\epsilon) \log \hat{g}(\epsilon) + \hat{g}(\epsilon)N(Ax + B\epsilon)]\pi(\epsilon)d\epsilon \quad (6) \]

Here

\[ J(x_t) = \hat{E}_t \sum_{j=0}^{\infty} \beta^j c_{t+j} \]

is the expected discounted log consumption under the worst case joint density and

\[ G_t N(x_t) = G_t \beta E \left[ \sum_{j=0}^{\infty} \beta^j \frac{G_{t+j}}{G_t} E[g_{t+j+1} \log g_{t+j+1}]|\epsilon^t, x_0] \right] \]

is continuation entropy.

Substituting the minimizer into the above equation gives the risk-sensitive recursion of Hansen and Sargent:

\[ W(x) = c - \beta \theta \log E \left[ \exp \left( \frac{-W(Ax + B\epsilon)}{\theta} \right) \right] \quad (7) \]

This same recursion is derived in the appendix from an Epstein and Zin preference
relation. The conclusion is that identical consumption plans can be achieved with a
different perspective over the preference parameters. While here $\theta$ is related to the degree
of ambiguity aversion of the agent, in the Epstein-Zin context, the parameter is related
to risk aversion.

2.2.5 The minimizing martingale increment

The minimizing martingale increment is the optimal choice of nature and is given
by

$$\hat{g}_{t+1} = \left( \frac{\exp(-W(Ax_t + B\epsilon_{t+1})/\theta)}{E_t[\exp(-W(Ax_t + B\epsilon_{t+1})/\theta)]} \right)$$  \hspace{1cm} (8)

This is an exponential change of measure. As it is, we can see that it is not necessary to
impose the condition that $g_{t+1} > 0$ since this restriction is naturally satisfied.

For the random walk model, we can show that

$$\hat{g}_{t+1} \propto \exp\left(\frac{-\sigma_\epsilon \epsilon_{t+1}}{(1-\beta)\theta} \right)$$

and the worst-case density for the innovation $\epsilon$ is

$$\hat{\pi}(\epsilon_{t+1}) \propto \exp\left(\frac{-\left(\epsilon_{t+1} + \frac{\sigma_\epsilon}{(1-\beta)\theta}\right)^2}{2} \right)$$

which is the density function of a normal random variable with mean $\frac{-\sigma_\epsilon}{(1-\beta)\theta}$, that is,

$$\hat{\pi}(\epsilon_{t+1}) \sim N(w(\theta), 1)$$ \hspace{1cm} (9)

where

$$w(\theta) = \frac{-\sigma_\epsilon}{(1-\beta)\theta}$$  \hspace{1cm} (10)

It means that the distorted distribution is still normal, but with a lower mean. How much
lower the distorted mean is compared to the prior distribution depends on the level of
concern of the agent with respect to model uncertainty, here represented by the parameter \( \theta \).

### 2.2.6 The discounted entropy

When the conditional densities for \( \epsilon_{t+1} \) under the approximating and worst case models are \( \pi \sim N(0,1) \) and \( \hat{\pi} \sim N(w(\theta), 1) \), respectively, the conditional entropy is

\[
E_t \hat{g}_{t+1} \log \hat{g}_{t+1} = \int (\log \hat{\pi}(\epsilon) - \log \pi(\epsilon)) \hat{\pi}(\epsilon) d\epsilon = \frac{1}{2} w(\theta)' w(\theta)
\]

it then follows that discounted entropy becomes

\[
\beta E \left[ \sum_{t=0}^{\infty} \beta^t \hat{G}_t E(\hat{g}_{t+1} \log \hat{g}_{t+1} | \epsilon^t, x_0) | x_0 \right] = \eta = \frac{\beta}{2(1-\beta)} w(\theta)' w(\theta)
\]

This formula allows us to compute explicitly how distant the worst case measure is from the prior as a function of the ambiguity aversion parameter \( \theta \). It is clear that the greater the \( \theta \), the less ambiguity averse the agent is and the closer the distorted measure is to the prior.

### 2.2.7 The value function for random walk log consumption

Using the formula for \( w(\theta) \) from the random walk model tells us that discounted entropy is

\[
N(x) = \frac{\beta}{2(1-\beta)} \frac{\sigma^2}{(1-\beta)^2 \theta^2}
\]  

(11)

We can then compute the value function for the agent to be

\[
W(x_t) = \frac{\beta}{(1-\beta)^2} \left[ \mu - \frac{\sigma^2}{2(1-\beta)\theta} \right] + \frac{1}{1-\beta} c_t
\]  

(12)

where the value function for the consumption process is
\[ J(x_t) = \frac{\beta}{(1 - \beta)^2} \left[ \mu - \frac{\sigma_t^2}{(1 - \beta)\theta} \right] + \frac{1}{1 - \beta} c_t \]  

(13)

so that \( W(x_t) = J(x_t) + \theta N(x_t) \).

We can interpret \( J(x_t) \) as the value function for an uncertainty constraint agent where discounted entropy is bounded by \( \eta \). To do so, we need to align \( \theta \) and \( \eta \) in a specific fashion (see Hansen and Sargent, pg.159).

We shall use these value functions to construct compensating variations in the initial condition for log consumption \( c_0 \) in an elimination of model uncertainty experiment to be described below.

2.3 The data

Our study focuses on the largest advanced and emerging economies of the world. In this way, our data set is compounded by eight developed countries\(^4\) and ten emerging market economies\(^5\). We do not include a few countries such as Argentina, China, and Saudi Arabia because of the lack of data availability for these economies. We use quarterly data from the beginning of 1970 to the end of 2015. As pointed out by Campbell (2003), there is no significant dispersion among the coefficients of relative risk aversion among developed countries when using annual data. In this way, given our interest in characterizing this variation, we opted to use quarterly data.

We rely on the International Financial Statistics of the International Monetary Fund (IFS-IMF) and the data provided by the Morgan Stanley Capital International (MSCI), obtained using Bloomberg. We use the following series: stock returns, consumer price index (CPI), short-term interest rate, total consumption, and GDP deflator. All data is in local currency and each countries’ data set is composed by the earliest to the latest available data within our predetermined date range.

Following Campbell (2003)\(^6\), we use the monthly MSCI National Price and Gross Return Indexes to build the quarterly stock market data. In this way, our stock series

\(^4\)Australia, Canada, Germany, France, Great Britain, Italy, Japan, and United States.
\(^5\)Brazil, Chile, Colombia, India, Indonesia, Korea, Mexico, Russia, South Africa, and Turkey.
consider not only the stock price increases but also the returns of dividends accumulated within the quarter. It is important to mention that these indexes are representative, but not comprehensive of the whole equity market of the economy. Additionally, this representativeness can vary substantially among countries. These observations should be considered when analyzing the results. The stock indexes are deflated using the CPI, which is obtained from the IFS-IMF. The gross real return on stocks are defined as:

\[ 1 + r_{e,t} = \frac{1 + R_{e,t}}{1 + \pi_t} \]  

where \( r_{e,t} \) is the real return on stocks at time \( t \), \( R_{e,t} \) is the nominal return on stocks, and \( \pi_t \) is the inflation rate.

The short-term interest rate is also obtained from IFS-IMF database, except for Germany, France, Italy, and Turkey which the source is Bloomberg. Besides Italy, for which we use the 3-month Treasury bills, we adopt the money market interest rate as a proxy for the risk-free short-term rate.

The source of the total consumption, GDP deflator, and population data is also the IFS-IMF. Since only the United States provides a reliable and extensive time series of household expenditure on non-durables and services, we decided to adopt total consumption as the proxy for household consumption for all countries. As pointed out by Mankiw (1985), Ogaki and Reinhart (1998b), Ogaki and Reinhart (1998a), Yogo (2006), and Pakos (2011), the treatment of durables can affect the estimates of the coefficient of intertemporal elasticity of substitution and the RRA. We need to consider this possible effect when evaluating the results since the estimation of consumption volatility is affected. Additionally, to keep the same methodology for all countries, we deflate consumption using GDP deflator. The real consumption per capita series were seasonally adjusted using the X-13ARIMA-SEATS Seasonal Adjustment Program from the US Census Bureau. Then, log consumption, \( c_t \), in the model is defined as the log of the seasonally adjusted consumption per capita. Finally, because of the time convention sensitiveness of the correlation between real consumption growth and stock returns showed by Campbell (2003), we define consumption growth for the quarter as the next quarter’s consumption growth.
divided by this quarter’s consumption.

Table 1 describes the data of the 18 countries in our data set. The range of the sample of each country is described in the last two columns. Only three countries’ samples cover the whole time span of 1970 to 2015: Australia, Japan, and United States. On the other hand, India presents the shortest sample, with only 11 years.

Note that even though the Great Recession (2007-2009), when most stock markets suffered a sharp negative hit, is a representative period of most samples, all stock markets present real average return greater than the risk-free rate. In this way, even the countries that present a small sample display positive equity premium.

2.4 Results

2.4.1 Risk and Model Uncertainty Aversion

Table 1 also shows the estimated mean and variance of the consumption processes. They were obtained using maximum likelihood (ML) estimation and assuming that the consumption growth follows the random walk model. The second and third columns of Table 1 show the results of the ML estimation for the 18 countries of our sample. India presents the highest quarterly average consumption growth ($\mu$), around 1.6%, in our estimation. On the other hand, France presents the lowest $\mu$, 0.2%. Once again, it is important to highlight the differences between sample periods among the countries. While India’s sample begins in 2005, the French times series are ten years longer, starting from 1995. Mexican, Japanese, and Colombian consumption processes are the most volatile in our sample. Their $\sigma$’s are higher than 0.03, more than double the average $\sigma$ of the sample, 0.014.

After estimating the consumption processes and calculating equity return and the risk-free rate, we proceed to the computation of the model’s parameters $\gamma$, $\theta = \frac{-1}{(1-\gamma)(1-\beta)}$ (see appendix), and $p(\theta^{\star-1})$. We first focus on how to obtain $\gamma$ that attain the Hansen-Jagannathan bounds:

$$\sigma(m) \geq \sigma^{\star}(m), \quad (15)$$

25
where
\[ \sigma(m) = \beta \exp \left( -\mu + \frac{\sigma}{2} (2\gamma - 1) \right) \left[ \exp \left( \sigma^2 \gamma^2 \right) - 1 \right]^{0.5}, \] (16)
and
\[ \sigma^*(m) = (1 - E[m]E[R])^{1/2} \left[ \exp \left( \sigma^2 \gamma^2 \right) - 1 \right], \] (17)
where \( E[m] = 1/R^f \), \( R^f \) is the real risk-free rate, \( R \) is the real stock return, and \( \Sigma = \text{var}(R) \). The detailed derivation is in the Appendix.

To find \( \gamma \) we first need to calibrate the discount factor, \( \beta \). Following Okubo (2015) and Garcia-Cicco et al. (2010), we set the same \( \beta \) for all AE and also a common \( \beta \) for all EM. Then, for the eight advanced economies in our sample, we follow Okubo (2015)\(^7\) and set \( \beta_{AE} = 0.995 \). On the other hand, for the EM, we calculate the common \( \beta \).

To calibrate \( \beta_{EM} \) we use the steady-state relation \( \beta_{EM} = 1/R^*_EM \), where \( R^*_EM \) is the average risk-free short-term rate of these economies. We define the country risk as the spread of the US dollar denominated five-year bond of the country to the same maturity US bond. After finding the risk estimate, we subtract it from the short-term interest rate to find the risk-free measure. However, a setback of this approach is the lack of US dollar denominated bonds. Only three countries present a sample longer than 15 years: Brazil, Colombia, and Mexico. Then, using these three time series, we found \( \beta_{EM} = 0.995 \). Note that using this approach we find the same value as \( \beta_{AE} \). However, this is not surprising since we expect arbitrage to eliminate the interest rate differences among countries after discounting for risk. In the robustness section, we use an alternative method for calibrating \( \beta_{EM} \).

After computing \( \beta \) and \( \gamma \), we proceed to the computation of \( \theta \) and \( p(\theta^{*-1}) \). We know that an agent with uncertainty aversion \( \theta \) has utility (value function) represented by the recursion
\[ W_t = c_t - \beta \theta \log \mathbb{E} \left[ \exp \left( - \frac{W_{t+1}}{\theta} \right) \right] \] (18)

\(^7\)If we calibrate \( \beta_{AE} \) using its steady-state relation with the risk-free short-term rate of the Advanced Economies since 1970, we find \( \beta_{AE} = 0.996 \), a similar number to the one used in the literature. Then, the results with this new calibration would be the same. Nevertheless, these results are available upon request.
this is the equivalent recursion of a risk-averse investor (see appendix). It means that there is an equivalent observation between risk averse agents and uncertainty averse agents. The difference between both is due to interpretation of the agent preference.

In the case of the uncertainty averse agent, the recursiveness arises from the worst case measure considered by the agent. Suppose that in the case of perfect foresight the consumption trajectory is given by, say model A,

\[ c_{t+1} = c_t + \mu + \sigma \epsilon_{t+1}, \quad \epsilon_{t+1} \sim i.i.d. N(0, 1) \]  

(19)

the consumption path that arises from the worst case scenario, say model B, considered by the agents is

\[ c_{t+1} = c_t + (\mu + \sigma \omega) + \sigma \epsilon_{t+1}, \quad \epsilon_{t+1} \sim i.i.d. N(0, 1), \quad \omega = -\frac{\sigma}{\theta(1 - \beta)} \]  

(20)

that is, the agent considers a path with smaller conditional mean, where the size of the decreased mean is determined by the degree of uncertainty of the agent. The question that must be answered is: what is a reasonable level of uncertainty aversion? To answer this question we follow Hansen and Sargent (2008) who propose the use of detection error probabilities \( p(\theta \, \theta^{-1}) \).

The idea of detection error probabilities is the following. The agent considers the baseline model A, and the alternative model B with lower conditional mean, \( \omega(\theta) \).

When looking at the consumption history, he compares the likelihood of model A, \( L_A \), with the likelihood of model B, \( L_B \).

Note that the likelihood of model B depends on the level of uncertainty aversion \( \theta^{-1} \), where bigger uncertainty implies lower conditional mean for the worst case consumption scenario.

Considering a priori distribution of both consumption paths equally, we can link the uncertainty parameter level \( \theta \) with the probability of incurring a mistake that is, the
probability of considering model A, when model B is the correct model and vice versa.

To compute the detection error probability we only need to calculate the probability of \( L_A > L_B \) when the data is generated by model B and the probability of \( L_B > L_A \) when the model A is the correct one. The detection error probability is then

\[
p(\theta^*-1) = 0.5 p \left( \frac{L_B}{L_A} > 0 \mid \text{model A is correct} \right)
+ 0.5 p \left( \frac{L_A}{L_B} > 0 \mid \text{model B is correct} \right)
\]

It is not difficult to show that

\[
p(\theta^*-1) = \Phi \left( -\sqrt{T} \frac{\sigma}{\theta^*(1-\beta)} \right)
\]

The derivation of the specific formulas for detection error probabilities due to Okubo (2015) is provided in the appendix. This result gives a nice relation between the probability of committing mistakes and the uncertainty level parameter \( \theta^{-1} \). It confirms the intuition that a larger degree of model uncertainty aversion \( \theta^{-1} \) implies in a lower probability of committing mistakes \( p(\theta^{-1}) \), that is, \( p(\theta^{-1}) \) is a decreasing function. This result is expected since a higher \( \theta^{-1} \) implies a smaller conditional mean for the worst case consumption process B, which in turn makes it easier to distinguish it from the approximating model A.

The remaining issue is the level of uncertainty that is acceptable when we consider the possibility of choosing the wrong model. It is clear that no level of uncertainty should make the agent consider a model that is wrong with probability greater than half since this is the probability of committing mistakes when there are no model misspecification concerns (\( \theta^{-1} = 0 \)). Hansen and Sargent (2008) argue that a bound for \( p(\theta^*-1) \) between 0.15 and 0.20 implies a reasonable concern of committing mistakes. Note that this implies an upper bound for misspecification concerns that can be considered reasonable to have \( \bar{\theta}^{-1} = \Phi^{-1}(.15) \times (-2(1 - \beta)/\sigma \sqrt{T}) \).
Table 2 reports $\gamma$, $\theta^{*1}$, and $p(\theta^{*1})$ estimated for each country. The first column reports the risk aversion coefficient value that attains the Hansen-Jagannathan bound. The standard deviation of $\gamma$ among the countries in our sample is 10.89, while the mean is 15.64. This result is similar to Campbell (2003) and Okubo (2015), which also found significant RRA parameter variation across countries, but using a sample including only advanced economies. Looking at individual countries, Brazil presents the lowest $\gamma$, 2.99, while France and the US have the highest, 35.13 and 34.64, respectively. We can also see in this column that the three highest $\gamma$ values are from advanced economies. On the other hand, the four lowest numbers are from emerging markets: Brazil, Turkey, India, and Mexico. These findings suggest not only a significant variation in risk aversion across countries but also between the two country groups. On average, AE’s investors are more risk averse than the ones based on emerging markets.

Given the values of $\gamma$, we find the inverse of the penalty parameter, $\theta^{*1}$, and the detection error probabilities, $p(\theta^{*1})$. The results are reported in the second and third columns of Table 2. Note that the estimates of $p(\theta^{*1})$ also present significant variation across the countries, but the variability is considerably lower. Table 3 shows the coefficients of variation of $\gamma$, $\theta^{*1}$, and $p(\theta^{*1})$. When we consider all countries, the coefficient of variation (CV) of $\gamma$, 0.70, is double the coefficient of $p(\theta^{*1})$, 0.35. The lower variation of detection error probabilities indicates that agents consider similar probabilities of committing mistakes when choosing the optimal allocation. This result stands even if we look within the country groups. The CV of $\gamma$ is 0.59 among AE and the coefficient of $p(\theta^{*1})$ is only 0.33. When we look to EM, the 2:1 ratio also stands, 0.74 for $\gamma$ and 0.38 for the detection error probability.

It is also important to compare the plausibility of the estimated parameters since it helps us to have a perception of the model suitability. On the first column of Table 2, we can see that only Brazil, India, Mexico, and Turkey present $\gamma$ estimates under 5.0, a common threshold in the literature. Even if we consider a looser bound of 10.0, the number of countries inside this limit would not be considerable. Only seven countries of eighteen in our sample would satisfy this bound, corroborating the equity risk premium puzzle to international data.
On the other hand, when we follow Barillas et al. (2009) and use model uncertainty to reinterpret the high level of $\gamma$, we find very different results. Note on Figure 1 that only Colombia and the US present a detection error probability below the 0.15 bound proposed by Hansen and Sargent (2008). Except for these two economies, model misspecification explains the equity premium for all countries in our sample. This is a stark contrast to the results of the specification without model uncertainty and indicates the better suitability of this model to our sample. Additionally, when using the full sample of all countries, we cannot identify any pattern difference between AE and EM, indicating that the model is equivalent for both country groups.

2.4.2 Welfare Costs

After estimating the risk parameters for the countries and analyzing its feasibility, we study the welfare gains of eliminating the “traditional” risk and model uncertainty. Following Barillas et al. (2009), we describe how market prices of uncertainty extracted from data contain information about how much the representative consumer would be willing to pay to eliminate model uncertainty. We use as a point of comparison the certainty equivalent plan

$$c_{t+1} - c_t = \mu + \frac{1}{2} \sigma^2$$

(22)

We seek an adjustment to initial consumption that renders a representative consumer indifferent between the certainty equivalent plan and the original risky consumption plan. For the same initial conditions, the certainty equivalent path of consumption $\exp c_{t+1}$ has the same mean as the original plan $c_{t+1} - c_t = \mu + \sigma \epsilon_{t+1}$, but its conditional variance has been reduced to zero.

Recall the formula for the value function of the representative agent facing a random walk process for log consumption, specified by

$$U(c_0) = \frac{\beta}{(1-\beta)^2} \left[ \mu - \frac{\sigma^2}{2(1-\beta)\theta} \right] + \frac{1}{1-\beta} c_0$$

(23)

we seek a proportional decrease in the certainty equivalent trajectory that leaves $U$ equal
to its value under the risky process. Let $c_0^I$ denote the initialization of the certainty equivalent trajectory for an agent. Evidently, it satisfies equation

$$\frac{\beta}{(1-\beta)^2} \left( \mu + \frac{\sigma^2}{2} \right) + \frac{1}{1-\beta} c_0^I = \frac{\beta}{(1-\beta)^2} \left[ \mu - \frac{\sigma^2}{2(1-\beta)\theta} \right] + \frac{1}{1-\beta} c_0$$  \hspace{1cm} (24)

The left side is the value under the certainty equivalent plan, while the right side is the value under the original risky plan starting from $c_0$. Solving for $c_0 - c_0^I$ gives

$$c_0 - c_0^I = \frac{\beta}{(1-\beta)} \left[ \frac{\sigma^2}{2} + \frac{\sigma^2}{2(1-\beta)\theta} \right] = \frac{\beta\sigma^2}{2(1-\beta)} \left[ 1 + \frac{1}{(1-\beta)\theta} \right] = \frac{\beta\sigma^2}{2(1-\beta)} \gamma (25)$$

Note that this is exactly how much of today consumption the agent would be willing to give up to eliminate all risk and uncertainty from future consumption. As in Barillas et al. (2009), we now consider an agent who does not fear model uncertainty, so that $\theta = \infty$. We ask how much adjustment in the initial condition of a certainty equivalent path a $\theta = +\infty$ type of consumer would require. The compensating variation for the elimination of risk alone must satisfy

$$\frac{\beta}{(1-\beta)^2} \left( \mu + \frac{\sigma^2}{2} \right) + \frac{1}{1-\beta} c_0 = \frac{\beta}{(1-\beta)^2} \mu + \frac{1}{1-\beta} c_0^I(r)$$  \hspace{1cm} (26)

In constructing the right side, we have set $\theta = \infty$ and replaced $c_0$ with $c_0^I(r)$. Solving the above equation for $c_0 - c_0^I(r)$ gives

$$c_0 - c_0^I(r) = \frac{\beta\sigma^2}{2(1-\beta)} \gamma (27)$$

Evidently, the part of the compensation that is accounted for by aversion to model uncertainty is

$$c_0^I(r) - c_0^I = \frac{\beta\sigma^2}{2(1-\beta)} \left[ \frac{1}{1-\beta} \right] = \frac{\beta\sigma^2}{2(1-\beta)^2\theta} \gamma (28)$$
Here, it is important to notice that the results might be affected by the difference of the sample periods of the countries. A period of higher global uncertainty might have different weights in the sample of the countries and affect the results. For example, while Australian data starts in 1970 and ends in 2015, the Brazilian sample has only 17 years, from 1999 to 2015. It is clear that the Great Recession is more relevant to the Brazilian sample than to the Australian one. As a consequence, a significant amount of the differences between the countries’ results might be caused by the different sample periods, which is much larger for most AE countries in our sample. In this way, we now use the same sample period for all countries. We set the sample period from the beginning of 1999 to the end of 2015 and exclude all countries whose data does not fit in this period.

Table 4 gives a brief description of this new data set. Note that now our sample has fewer countries than before. India, Russia, and Turkey were excluded because they do not have available data for the whole period. We can also observe that only the sample of AE countries changed when comparing with the full sample. All EM on Table 4 had their data beginning in 1999 and ending in 2015 in the full sample. Then, we will only have changes in the results of AE. Comparing Tables 4 and 1 we can notice a changing pattern in the sample of the AE. For almost all countries, both $\mu$ and $\sigma$ estimated using the shorter sample are smaller than the previous ones. The biggest decrease in $\mu$ was in the Japanese estimate, in line with their strong GDP growth deceleration in last decades. We can also see a changing pattern of the stock returns. Except for Australia and Canada, whose stock markets benefited from the commodity boom in the last 15 years, all average stock returns decreased. Finally, all AE short term rates are also smaller in this new sample than in the previous one.

We replicate the exercise of the preceding section, assuming $\beta_{AE} = \beta_{EM} = 0.995$ and that the consumption growth process follows a random walk model. Table 5 show the results of the estimation of $\gamma$ and $p(\theta^{*^{-1}})$ that attain the Hansen-Jagannathan bounds. The new $\gamma$’s computed using the 1999-2015 sample are shown in the second column of the Table 5. The most significant changes were in the estimates for Australia, Great Britain, Italy, and Japan. Australian and Japanese estimates, which were lower than ten using the
full sample, increased to well above this level. On the other hand, the estimates of Great Britain fell from 15.9 to 5.2 and of Italy from 11.1 to 0.5. However, the most significant result of our exercise is the general increase in the estimates of $p(\theta^*-1)$. Except for Australia, which the value did not change, all new detection error probabilities estimates of the AE grew, indicating a lower degree of model uncertainty in this period. The estimates for Japan and USA, which were under 0.20 for the full sample, surpassed 0.30 when using the 1999-2015 sample. Using the homogeneous sample, only the Colombian $p(\theta^*-1)$ that attain the HJ bounds is under the lower limit of 0.15 suggested by Hansen and Sargent (2008). Besides most countries’ $p(\theta^*-1)$ remaining on the reasonable bound as in the full sample exercise, the variability of the parameter is still lower when using the homogeneous sample. The coefficient of variation between countries’ estimates is more than two times bigger for $\gamma$ than for $p(\theta^*-1)$ (Table 6).

Figure 2 presents the estimates of $p(\theta^*-1)$ for all countries. Looking at Figure 2 it is easy to see that the $p(\theta^*-1)$ estimates for AE are larger than 0.25 and greater than the figures of EM (except when comparing Canada and Mexico, which have the same values and for Brazil which has the second largest value). Additionally, except for Brazil, the estimates of EM are lower than 0.30 and Colombia and South Africa are the only countries with detection error probabilities under 0.20. Then, it seems that the robust approach assuming a random walk model for the consumption growth process indicates a higher level of uncertainty aversion for EM during the 1999-2015 period than for advanced economies.

Figure 3 compares the potential gains of eliminating risk and model uncertainty for all countries between 1999 and 2015. It is evident when looking at the figure that the welfare gains from reducing uncertainty are much greater in developing countries. All AE are concentrated on the lower left of the graph, while the EM are spread upward and toward the right in the graph. The only exception is Brazil, which we should remember that has a high $p(\theta^*-1)$ parameter, way above the 0.20 bound. The AE’ households would be willing to reduce their average consumption growth by a value between 0 and 10% in order to eliminate model uncertainty. For the EM, these figures are considerably higher. This result is not surprising since EM are usually considered more unstable economically and
politically than AE. It is also noteworthy that even considering only emerging economies, the results of Colombia and Mexico stand out.

We could also ask if the EM has proportionately more model uncertainty than AE. Figure 4 shows the ratio between the welfare gains from eliminating model uncertainty and the gains from just removing risk. Looking at the figure, we notice that, on average, AE bear more model uncertainty proportionately than EM. From the eight AE countries in the sample, six have a ratio greater than 10, while for EM we have four out of seven. Then, we can infer that even though EM economies bear much more model uncertainty in absolute terms, model uncertainty is relatively very important to explain total risk to all countries groups.

2.4.3 Robustness

In this section, we repeat previous exercises but using a different calibration for $\beta_{EM}$. Instead of using the risk-free short-term rate of Brazil, Colombia, and Mexico to calibrate $\beta_{EM}$, we use the average short-term rate of these countries without discounting for risk. Then, the value of $\beta_{EM}$ reduces from 0.995 to 0.989. Since the estimates for the detection error probabilities are almost not affected by the value of $\beta$ (Table 7), we discuss only the results of the welfare analysis.

Figure 5 shows the welfare gains from eliminating risk and model uncertainty for all countries using the homogeneous sample. As expected, the lower $\beta_{EM}$ reduces both risk and model uncertainty elimination gains for all EM. However, emerging economies still present more gains than AE in general. The only exceptions are Brazil and Korea. In this way, even using a lower $\beta_{EM}$, the results that welfare gains from reducing uncertainty are much greater in developing countries still holds.

2.5 Conclusion

We evaluate Barillas et al. (2009) model of ambiguity averse agents in light of a broader dataset by including developing economies and a longer time-span and show

---

8If we use the average short-term rate of all EM, then $\beta_{EM} = 0.994$, which will not significantly change the results from the previous section.
that including model uncertainty aversion also helps to explain the equity premium for countries other than the U.S. Additionally, when considering the same sample period for all countries, we find the fear for model misspecification higher for EM than for AE which is in line of what would be expected. This result is expected since EM usually have more political instability and are more susceptible to economic shocks. Brazil stands out from the other EM, since it possesses a low level of model uncertainty aversion, which can be checked by looking at its 0.468 detection error probability. This value is very close to the .5 achieved for a perfect foresight economy. This result corroborates the analysis of a very low (or none) equity premium observed in Brazil financial markets for this sample period. One possible explanation for this discrepancy is that since the stock market is more volatile when compared to Advanced Economies, the short time span of data available is not able to capture the long run equity premium present in the economy. Other possibility is that there are unidentified factors that make the equity premium low in Brazil. Further investigation is required to understand the low equity premium in Brazil for the time span analyzed here. We also were able to identify model uncertainty as the most important factor of welfare loss when comparing to risk effects alone. This result suggests that a policy to prevent uncertainty may be considered to improve welfare with greater possibility of welfare gains for EM.
Appendix

Computing Welfare Costs

Lucas, Reis and other authors compute the welfare cost of consumption fluctuations in a different fashion. They compute an annual cost instead of computing a present value of costs as we do here. In their computation they define the costs of fluctuations as the scalar \( \lambda \) that solves the equation

\[
E\left[\sum_{t=0}^{\infty} e^{-\beta t} u(C_t(1 + \lambda))\right] = \sum_{t=0}^{\infty} e^{-\beta t} u(\bar{C}_t)
\]

for \( u(C_t) = \ln(C_t) \) it is not difficult to show that

\[
\ln(1 + \lambda) = 0.5(1 - e^{-\beta}) \sum_{t=0}^{\infty} e^{-\beta t} Var(c_t)
\]

Furthermore, if log consumption follows as stationary AR(1) process, \( Var(c_t) = \sigma^2(1 - \eta^2)/(1 - \eta^2) \) for \( t \geq 1 \). Evaluating the sum on the right hand side of the above equation we get

\[
\ln(1 + \lambda) = \frac{0.5\sigma^2}{e^\beta - \eta^2}
\]

Kreps-Porteau-Epstein-Zin preference representation

We use preferences that can be described by a recursive non-expected utility function à la Kreps and Porteus (1978)

\[
V_t = W(C_t, \mu(V_{t+1}))
\]

where \( W \) is an aggregator function. The idea of this kind of representation is that the agent’s preference is a function of today’s consumption and future consumption. The future consumption is uncertain but we may be able to compute an equivalent quantity that the agent would be willing to consume for sure next period in exchange for all possible
future consumption. This quantity $\mu$ is a certainty equivalent function

$$\mu(V_{t+1}) = f^{-1}(E_t f(V_{t+1}))$$

$f$ is a function that determines attitudes toward atemporal risk. It determines how much an individual would be willing to trade off expected future consumption to have a certain quantity of consumption for sure. Note that if $f$ is increasing and concave, then by Jensen inequality we have

$$E(f(X)) \leq f(E(X))$$

And by monotonicity, for any $\mu$ such that

$$f(\mu(X)) = E(f(X)) \leq f(E(X)) \rightarrow \mu(X) \leq E(X)$$

Epstein-Zin use a constant relative risk aversion (CRRA) function to represent individual attitude towards risk

$$f(z) = z^{1-\gamma}$$

$$f(z) = \log(z), \text{ if } \gamma = 1$$

and $\gamma$ is the coefficient of relative risk aversion.

Following Epstein and Zin (1991) it is common to use the CES aggregator $W$

$$W(C_t, \mu) = [(1 - \beta)C_t^{1-\eta} + \beta \mu^{1-\eta}]^{\frac{1}{1-\eta}}$$

$$\lim_{\eta \rightarrow 1} W(C_t, \mu) = C_t^{1-\beta} \mu^\beta$$

Tallarini used a power certainty equivalent function to get the following recursive utility under uncertainty

$$V_t = C_t^{1-\beta}[(E_t(V_{t+1}^{1-\gamma}))^{\frac{1}{1-\gamma}}]^\beta$$

Taking logs gives
\[ \log V_t = (1 - \beta)c_t + \frac{\beta}{1 - \gamma} \log E_t(V_{t+1}^{1-\gamma}) \]

where \( c_t = \log C_t \) or

\[ \log V_t = \frac{c_t}{1 - \beta} + \frac{\beta}{(1 - \gamma)(1 - \beta)} \log E_t(V_{t+1}^{1-\gamma}) \]

define \( U_t = \frac{\log V_t}{1 - \beta} \) and \( \theta = \frac{1}{(1 - \gamma)(1 - \beta)} \), then

\[ U_t = c_t - \beta \theta \log E_t \left( \exp \left( \frac{-U_{t+1}}{\theta} \right) \right) \]

This is the risk-sensitive recursion of Hansen and Sargent (1995). In the special case that \( \gamma = 1 (\theta = +\infty) \) the recursion becomes the standard discounted expected utility recursion

\[ U_t = c_t + \beta E_t U_{t+1} \]

This relation only implies that the total utility of the agent can be expressed as the current utility plus a discounted expected future utility, where \( \beta \) is the intertemporal discounting parameter.

The recursion implies the following Bellman equation for the random walk case

\[ U(c) = c - \beta \theta \log E_t \left( \exp \left( \frac{-U(c + \mu + \sigma \epsilon)}{\theta} \right) \right) \]

the value function that solves this equation is

\[ U(c) = \frac{\beta}{(1 - \beta)^2} \left[ \mu - \frac{\sigma^2}{2\theta(1 - \beta)} \right] + \frac{1}{1 - \beta} c \]

We model the agents’ preferences in a way that we can interpret \( \theta \) as a parameter that measures concern for model specification. To do so, we must describe how the agents evaluate consumption in an environment of model uncertainty. There are two approaches that are equivalent under some conditions. The first uses multiplier preferences and the other uses constraint preferences.

38
The meaning of entropy

Now, to understand a little bit about the penalizing factor, notice that \( h(x) = x \log x \) is a convex function for values of \( x \geq 0 \), so that \( h(y) - h(x) \geq h'(x)(y - x) \). Then we have

\[
h(G_t) - h(1) = G_t \log G_t \geq G_t - 1 = h'(1)(G_t - 1)
\]

so that

\[
E[G_t \log G_t | x_0] \geq 0.
\]

Again, we only have \( E[G_t \log G_t | x_0] = 0 \) if \( G_t = 1 \) in which case there is no probability distortion associated with time \( t \) shock distribution. The factorization \( G_t = \prod_{j=1}^{t} g_j \) implies the following decomposition of entropy

\[
E[G_t \log G_t | x_0] = \sum_{j=0}^{t-1} E[G_j E(g_{j+1} \log g_{j+1} | \varepsilon^j, x_0) | x_0].
\]

We can then compute the discounted entropy over an infinite horizon as

\[
(1 - \beta) \sum_{j=0}^{\infty} \beta^j E(G_j \log G_j | x_0) = \sum_{j=0}^{\infty} \beta^j E[G_j E(g_{j+1} \log g_{j+1} | \varepsilon^j, x_0) | x_0].
\]

This term multiplied by the parameter \( \theta \) is exactly the second term of the summation formula that describes agent preference and summarizes all uncertainty considered by the agent. The parameter \( \theta \) represents how much weight the agent gives to all the uncertainty present in the economy.

Hansen-Jagannathan Bounds

We want to find \( m \) such that \( E[mR] = 1_N \), where \( 1_N \in \mathbb{R}^N \) is a vector of ones and \( R \) is a vector of returns of the risky assets. We claim that any

\[
m = E[m] + (1_N - E[m]E[R])' \Sigma^{-1} (R - E[R]) + \varepsilon
\]
where \( \Sigma = \text{Cov}(R, R') \) with \( \varepsilon \) such that \( E[\varepsilon] = 0 \) and \( E[\varepsilon R] = 0 \) do the job. To see this, first note that

\[
E[mR]' - E[m]E[R]' = E[m(R - E[R])']
\]

substituting \( m \) from the claim we get

\[
E[m(R - E[R])'] = (1 - E[m]E[R])'
\]

We conclude that \( E[mR] = 1_N \). Now since \( \sigma^2(\varepsilon) \geq 0 \) we have

\[
\sigma^2(m) = E[(m - E[m])^2]
\]

\[
= E[((1_N - E[m]E[R])'\Sigma^{-1}(R - E[R]) + \varepsilon)^2]
\]

\[
\geq E[(1_N - E[m]E[R])'\Sigma^{-1}(R - E[R])(R - E[R])'\Sigma^{-1}(1_N - E[m]E[R])]
\]

\[
= (1_N - E[m]E[R])'\Sigma^{-1}(1_N - E[m]E[R])
\]

where, if exists a risk free asset, \( E[m] = 1/R_f \).

**Formulas for Detection Error Probabilities**

Consider the following AR(1) process:

\[
c_{t+1} = c_t + \bar{\mu} + \sigma \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim i.i.dN(0, 1)
\]

The average log-likelihood function for a sample of \( t = 1, ..., T \) takes the form

\[
\ln L = \frac{1}{T} \ln f(c_1) + \frac{1}{T} \sum_{t=2}^{T} \ln f(c_t|c_{t-1})
\]

under the baseline model A we have \( \bar{\mu} = \mu \) and

\[
\ln L_A = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{T} \frac{1}{2\sigma^2} (c_1 - \mu)^2 - \frac{1}{T} \sum_{t=2}^{T} \frac{1}{2\sigma^2} (c_t - c_{t-1} - \mu)^2
\]
and under the worst case model B, \( \bar{\mu} = \mu + \sigma w \), and the log-likelihood is given by

\[
\ln L_B = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{T} \sum_{t=2}^{T} \frac{1}{2\sigma^2} (c_t - c_{t-1} - \mu - \sigma w)^2
\]

Thus, we obtain the log-likelihood ratio

\[
\ln \left( \frac{L_A}{L_B} \right) = -\frac{1}{T} \left[ \frac{1}{2\sigma^2} (c_1 - \mu)^2 + \sum_{t=2}^{T} \frac{1}{2\sigma^2} (c_t - c_{t-1} - \mu)^2 \right] \\
+ \frac{1}{T} \left[ \frac{1}{2\sigma^2} (c_1 - \mu - \sigma w)^2 + \sum_{t=2}^{T} \frac{1}{2\sigma^2} (c_t - c_{t-1} - \mu - \sigma w)^2 \right]
\]

To calculate the detection error probability under model A, we only need to substitute \( c_1 - \mu = \sigma \varepsilon_1 \) and \( c_t - c_{t-1} - \mu = \sigma \varepsilon_t \) for \( t = 2, \ldots, T \) which yields

\[
\ln \left( \frac{L_A}{L_B} \right) | \text{model A is correct} = \frac{1}{T} \sum_{t=1}^{T} \left[ -\frac{1}{2\sigma^2} \varepsilon_t^2 + \frac{1}{2\sigma^2} (\varepsilon_t - w)^2 \right] \\
= \frac{1}{T} \sum_{t=1}^{T} (-w \varepsilon_t) + \frac{1}{2} w^2
\]

Therefore, the detection error probability under model A is

\[
p \left( \ln \left( \frac{L_A}{L_B} \right) < 0 \mid \text{model A is correct} \right) = p \left( \frac{1}{T} \sum_{t=1}^{T} (-w \varepsilon_t) + \frac{1}{2} w^2 < 0 \right) \\
= p \left( \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\sigma}{\theta(1-\beta)} \right) \varepsilon_t < -\frac{1}{2} \left( \frac{\sigma}{\theta(1-\beta)} \right)^2 \right) \\
= p \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t < -\sqrt{T} \frac{\sigma}{2 \theta(1-\beta)} \right) \\
= \Phi \left( -\frac{\sqrt{T} \frac{\sigma}{2 \theta(1-\beta)}} \right)
\]

Where \( \Phi \) is the Normal cumulative distribution function.

On the other hand, if we consider model B as the correct one, we have \( c_1 - \mu = \sigma w + \sigma \varepsilon_1 \) and \( c_t - c_{t-1} - \mu = \sigma \varepsilon_t \) for \( t = 2, \ldots, T \) which yields
Therefore, the detection error probability under model B is

\[
p \left( \ln \left( \frac{L_A}{L_B} \right) > 0 \mid \text{model B is correct} \right) = p \left( \frac{1}{T} \sum_{t=1}^{T} (-w \varepsilon_t) - \frac{1}{2} w^2 > 0 \right)
\]

\[
= p \left( \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\sigma}{\theta(1-\beta)} \right) \varepsilon_t > \frac{1}{2} \left( \frac{\sigma}{\theta(1-\beta)} \right)^2 \right)
\]

\[
= p \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t > \frac{\sqrt{T}}{2} \frac{\sigma}{\theta(1-\beta)} \right)
\]

\[
= 1 - \Phi \left( \frac{\sqrt{T}}{2} \frac{\sigma}{\theta(1-\beta)} \right)
\]
Figures

Full Sample

Figure 1: Detection error probabilities estimates using the full sample ($\beta_{AE} = \beta_{EM} = 0.995$).
Figure 2: Detection error probabilities estimates using the homogeneous sample ($\beta_{AE} = \beta_{EM} = 0.995$).
Figure 3: Welfare gains from eliminating risk and model uncertainty between 1999 and 2015 ($\beta_{AE} = \beta_{EM} = 0.995$ and $p(\theta^{-1})$ estimated).
Figure 4: Ratio between welfare gains from eliminating only model uncertainty to gains from eliminating only risk between 1999 and 2015 ($\beta_{AE} = \beta_{EM} = 0.995$ and $p(\theta^{-1})$ estimated).
Figure 5: Welfare gains from eliminating risk and model uncertainty using $\beta_{EM} = 0.989$ (full sample and $p(\theta^{-1})$ estimated).
<table>
<thead>
<tr>
<th>country</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>Stocks (%)</th>
<th>ST Rate (%)</th>
<th>Min Date</th>
<th>Max Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU</td>
<td>0.005</td>
<td>0.007</td>
<td>1.447</td>
<td>0.613</td>
<td>1970-06-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>BR</td>
<td>0.004</td>
<td>0.010</td>
<td>2.413</td>
<td>2.025</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>CA</td>
<td>0.004</td>
<td>0.005</td>
<td>1.699</td>
<td>0.629</td>
<td>1975-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>CL</td>
<td>0.007</td>
<td>0.011</td>
<td>1.883</td>
<td>0.226</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>CO</td>
<td>0.011</td>
<td>0.032</td>
<td>4.442</td>
<td>0.539</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>DE</td>
<td>0.003</td>
<td>0.005</td>
<td>2.187</td>
<td>0.227</td>
<td>1995-06-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>FR</td>
<td>0.002</td>
<td>0.004</td>
<td>2.012</td>
<td>0.252</td>
<td>1995-06-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>GB</td>
<td>0.005</td>
<td>0.006</td>
<td>1.491</td>
<td>0.667</td>
<td>1988-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>ID</td>
<td>0.005</td>
<td>0.016</td>
<td>3.536</td>
<td>0.440</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>IN</td>
<td>0.016</td>
<td>0.014</td>
<td>0.971</td>
<td>0.201</td>
<td>2005-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>IT</td>
<td>0.001</td>
<td>0.006</td>
<td>1.461</td>
<td>0.591</td>
<td>1995-06-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>JP</td>
<td>0.006</td>
<td>0.034</td>
<td>2.311</td>
<td>0.250</td>
<td>1970-06-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>KR</td>
<td>0.008</td>
<td>0.008</td>
<td>2.708</td>
<td>0.245</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>MX</td>
<td>0.003</td>
<td>0.036</td>
<td>2.850</td>
<td>0.937</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>RU</td>
<td>0.011</td>
<td>0.014</td>
<td>3.270</td>
<td>-1.313</td>
<td>1999-03-01</td>
<td>2014-12-01</td>
</tr>
<tr>
<td>TR</td>
<td>0.006</td>
<td>0.029</td>
<td>3.162</td>
<td>1.512</td>
<td>2002-09-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>US</td>
<td>0.005</td>
<td>0.005</td>
<td>1.833</td>
<td>0.367</td>
<td>1970-06-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>ZA</td>
<td>0.003</td>
<td>0.008</td>
<td>2.751</td>
<td>0.615</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
</tbody>
</table>

This table reports an overview of the full data sample. The second and third columns show the results of ML estimation of the real consumption growth processes assuming a random walk model. Stocks and ST Rate are the quarterly average real return of stocks and short-term interest rate of each country, respectively. The last two columns show the initial and last date of the sample of each country.
<table>
<thead>
<tr>
<th>Country</th>
<th>$\gamma$</th>
<th>$\theta^{*-1}$</th>
<th>$p(\theta^{*-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU</td>
<td>9.378</td>
<td>0.042</td>
<td>0.337</td>
</tr>
<tr>
<td>BR</td>
<td>2.909</td>
<td>0.010</td>
<td>0.468</td>
</tr>
<tr>
<td>CA</td>
<td>19.566</td>
<td>0.093</td>
<td>0.260</td>
</tr>
<tr>
<td>CL</td>
<td>16.353</td>
<td>0.077</td>
<td>0.239</td>
</tr>
<tr>
<td>CO</td>
<td>9.024</td>
<td>0.040</td>
<td>0.145</td>
</tr>
<tr>
<td>DE</td>
<td>31.215</td>
<td>0.151</td>
<td>0.257</td>
</tr>
<tr>
<td>FR</td>
<td>35.427</td>
<td>0.172</td>
<td>0.251</td>
</tr>
<tr>
<td>GB</td>
<td>15.944</td>
<td>0.075</td>
<td>0.323</td>
</tr>
<tr>
<td>ID</td>
<td>11.616</td>
<td>0.053</td>
<td>0.245</td>
</tr>
<tr>
<td>IN</td>
<td>4.308</td>
<td>0.017</td>
<td>0.446</td>
</tr>
<tr>
<td>IT</td>
<td>11.079</td>
<td>0.050</td>
<td>0.389</td>
</tr>
<tr>
<td>JP</td>
<td>5.183</td>
<td>0.021</td>
<td>0.169</td>
</tr>
<tr>
<td>KR</td>
<td>21.284</td>
<td>0.101</td>
<td>0.259</td>
</tr>
<tr>
<td>MX</td>
<td>4.848</td>
<td>0.019</td>
<td>0.285</td>
</tr>
<tr>
<td>RU</td>
<td>14.277</td>
<td>0.066</td>
<td>0.238</td>
</tr>
<tr>
<td>TR</td>
<td>3.966</td>
<td>0.015</td>
<td>0.380</td>
</tr>
<tr>
<td>US</td>
<td>34.383</td>
<td>0.167</td>
<td>0.124</td>
</tr>
<tr>
<td>ZA</td>
<td>29.989</td>
<td>0.145</td>
<td>0.182</td>
</tr>
</tbody>
</table>

This table shows the results of the risk aversion parameter $\gamma$, the penalty parameter $\theta^{*-1}$, and the detection error probability $p(\theta^{*-1})$ of the calibration assuming that the consumption growth process follows a random walk model. The parameter $\gamma$ is chosen to satisfy the minimum of the Hansen-Jagannathan bounds: $\sigma(m) \geq \sigma^*(m)$. The discount factor for all countries is set to $\beta = 0.995$. 


Table 3: Risk Aversions and Detection Error Probabilities Coefficient of Variation (Full Sample)

<table>
<thead>
<tr>
<th>Group</th>
<th>$\gamma$</th>
<th>$\theta^*-1$</th>
<th>$p(\theta^*-1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced economies</td>
<td>0.589</td>
<td>0.620</td>
<td>0.331</td>
</tr>
<tr>
<td>Emerging markets</td>
<td>0.744</td>
<td>0.812</td>
<td>0.374</td>
</tr>
<tr>
<td>All countries</td>
<td>0.698</td>
<td>0.746</td>
<td>0.351</td>
</tr>
</tbody>
</table>

This table shows the coefficient of variation of the risk aversion parameters $\gamma$, the penalty parameters $\theta^*$, and the detection error probabilities $p(\theta^*)$ of the calibration assuming that the consumption growth process follows a random walk model. The parameter $\gamma$ is chosen to satisfy the minimum of the Hansen-Jagannathan bounds: $\sigma(m) \geq \sigma^*(m)$. The discount factor for all countries is set to $\beta = 0.995$. 

Table 4: Description of the homogeneous sample data set.

<table>
<thead>
<tr>
<th>country</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>Stocks (%)</th>
<th>ST Rate (%)</th>
<th>Min Date</th>
<th>Max Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU</td>
<td>0.003</td>
<td>0.006</td>
<td>1.786</td>
<td>0.457</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>BR</td>
<td>0.004</td>
<td>0.010</td>
<td>2.413</td>
<td>2.025</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>CA</td>
<td>0.003</td>
<td>0.004</td>
<td>1.903</td>
<td>0.130</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>CL</td>
<td>0.007</td>
<td>0.011</td>
<td>1.883</td>
<td>0.226</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>CO</td>
<td>0.011</td>
<td>0.032</td>
<td>4.442</td>
<td>0.539</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>DE</td>
<td>0.003</td>
<td>0.005</td>
<td>1.743</td>
<td>0.144</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>FR</td>
<td>0.002</td>
<td>0.004</td>
<td>1.339</td>
<td>0.139</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>GB</td>
<td>0.003</td>
<td>0.006</td>
<td>0.545</td>
<td>0.258</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>ID</td>
<td>0.005</td>
<td>0.016</td>
<td>3.536</td>
<td>0.440</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>IT</td>
<td>0.000</td>
<td>0.006</td>
<td>0.390</td>
<td>0.428</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>JP</td>
<td>0.002</td>
<td>0.008</td>
<td>1.063</td>
<td>0.021</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>KR</td>
<td>0.008</td>
<td>0.008</td>
<td>2.708</td>
<td>0.245</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>MX</td>
<td>0.003</td>
<td>0.036</td>
<td>2.850</td>
<td>0.937</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>US</td>
<td>0.003</td>
<td>0.004</td>
<td>0.835</td>
<td>-0.027</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
<tr>
<td>ZA</td>
<td>0.003</td>
<td>0.008</td>
<td>2.751</td>
<td>0.615</td>
<td>1999-03-01</td>
<td>2015-12-01</td>
</tr>
</tbody>
</table>

This table reports an overview of the 1999-2015 sample. The second and third columns show the results of ML estimation of the real consumption growth processes assuming a random walk model. Stocks and ST Rate are the quarterly average real return of stocks and short-term interest rate of each country, respectively. The last two columns show the initial and last date of the sample of each country.
<table>
<thead>
<tr>
<th>Country</th>
<th>$\gamma$</th>
<th>$\theta^{*\rightarrow 1}$</th>
<th>$p(\theta^{*\rightarrow 1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU</td>
<td>17.945</td>
<td>0.085</td>
<td>0.331</td>
</tr>
<tr>
<td>BR</td>
<td>2.909</td>
<td>0.010</td>
<td>0.468</td>
</tr>
<tr>
<td>CA</td>
<td>34.782</td>
<td>0.169</td>
<td>0.271</td>
</tr>
<tr>
<td>CL</td>
<td>16.353</td>
<td>0.077</td>
<td>0.239</td>
</tr>
<tr>
<td>CO</td>
<td>9.024</td>
<td>0.040</td>
<td>0.145</td>
</tr>
<tr>
<td>DE</td>
<td>23.498</td>
<td>0.112</td>
<td>0.327</td>
</tr>
<tr>
<td>FR</td>
<td>25.899</td>
<td>0.124</td>
<td>0.345</td>
</tr>
<tr>
<td>GB</td>
<td>5.362</td>
<td>0.022</td>
<td>0.460</td>
</tr>
<tr>
<td>ID</td>
<td>11.616</td>
<td>0.053</td>
<td>0.245</td>
</tr>
<tr>
<td>IT</td>
<td>0.525</td>
<td>-0.002</td>
<td>0.504</td>
</tr>
<tr>
<td>JP</td>
<td>14.064</td>
<td>0.065</td>
<td>0.337</td>
</tr>
<tr>
<td>KR</td>
<td>21.284</td>
<td>0.101</td>
<td>0.259</td>
</tr>
<tr>
<td>MX</td>
<td>4.848</td>
<td>0.019</td>
<td>0.285</td>
</tr>
<tr>
<td>US</td>
<td>24.132</td>
<td>0.116</td>
<td>0.345</td>
</tr>
<tr>
<td>ZA</td>
<td>29.989</td>
<td>0.145</td>
<td>0.182</td>
</tr>
</tbody>
</table>

This table shows the results of the risk aversion parameter $\gamma$, the penalty parameter $\theta^{*\rightarrow 1}$, and the detection error probability $p(\theta^{*\rightarrow 1})$ of the calibration assuming that the consumption growth process follows a random walk model. The parameter $\gamma$ is chosen to satisfy the minimum of the Hansen-Jagannathan bounds: $\sigma(m) \geq \sigma^*(m)$. The discount factor for all countries is set to $\beta = 0.995$. 


Table 6: Risk Aversions and Detection Error Probabilities Coefficient of Variation (Homogeneous Sample)

<table>
<thead>
<tr>
<th>Group</th>
<th>$\gamma$</th>
<th>$\theta^{*-1}$</th>
<th>$p(\theta^{*-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced economies</td>
<td>0.618</td>
<td>0.653</td>
<td>0.211</td>
</tr>
<tr>
<td>Emerging markets</td>
<td>0.699</td>
<td>0.753</td>
<td>0.396</td>
</tr>
<tr>
<td>All countries</td>
<td>0.645</td>
<td>0.688</td>
<td>0.323</td>
</tr>
</tbody>
</table>

This table shows the coefficient of variation of the risk aversion parameters $\gamma$, the penalty parameters $\theta^{*-1}$, and the detection error probabilities $p(\theta^{*-1})$ of the calibration assuming that the consumption growth process follows a random walk model. The parameter $\gamma$ is chosen to satisfy the minimum of the Hansen-Jagannathan bounds: $\sigma(m) \geq \sigma^*(m)$. The discount factor for all countries is set to $\beta = 0.995$. 
Table 7: Risk Aversion and Detection Error Probability using $\beta_{EM} = 0.989$
(Homogeneous Sample)

<table>
<thead>
<tr>
<th>Country</th>
<th>$\gamma$</th>
<th>$\theta^{*-1}$</th>
<th>$p(\theta^{*-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU</td>
<td>17.945</td>
<td>0.085</td>
<td>0.331</td>
</tr>
<tr>
<td>BR</td>
<td>2.926</td>
<td>0.021</td>
<td>0.468</td>
</tr>
<tr>
<td>CA</td>
<td>34.782</td>
<td>0.169</td>
<td>0.271</td>
</tr>
<tr>
<td>CL</td>
<td>16.450</td>
<td>0.170</td>
<td>0.237</td>
</tr>
<tr>
<td>CO</td>
<td>9.076</td>
<td>0.089</td>
<td>0.143</td>
</tr>
<tr>
<td>DE</td>
<td>23.498</td>
<td>0.112</td>
<td>0.327</td>
</tr>
<tr>
<td>FR</td>
<td>25.899</td>
<td>0.124</td>
<td>0.345</td>
</tr>
<tr>
<td>GB</td>
<td>5.362</td>
<td>0.022</td>
<td>0.460</td>
</tr>
<tr>
<td>ID</td>
<td>11.685</td>
<td>0.118</td>
<td>0.244</td>
</tr>
<tr>
<td>IT</td>
<td>0.525</td>
<td>-0.002</td>
<td>0.504</td>
</tr>
<tr>
<td>JP</td>
<td>14.064</td>
<td>0.065</td>
<td>0.337</td>
</tr>
<tr>
<td>KR</td>
<td>21.411</td>
<td>0.225</td>
<td>0.258</td>
</tr>
<tr>
<td>MX</td>
<td>4.877</td>
<td>0.043</td>
<td>0.284</td>
</tr>
<tr>
<td>US</td>
<td>24.132</td>
<td>0.116</td>
<td>0.345</td>
</tr>
<tr>
<td>ZA</td>
<td>30.166</td>
<td>0.321</td>
<td>0.181</td>
</tr>
</tbody>
</table>

This table shows the results of the risk aversion parameter $\gamma$, the penalty parameter $\theta^{*-1}$, and the detection error probability $p(\theta^{*-1})$ of the calibration assuming that the consumption growth process follows a random walk model. The parameter $\gamma$ is chosen to satisfy the minimum of the Hansen-Jagannathan bounds: $\sigma(m) \geq \sigma^*(m)$. The discount factor for all countries is set to $\beta = 0.989$. 

54
3 Robust Asset Allocation Problems

Abstract

The objective of this paper is to investigate the benefits of robust decision theory to asset allocation problems. In particular we introduce uncertainty aversion to a mean variance portfolio problem. We show that when agents have uncertainty aversion, they consider a return distribution that has lower expectation and increased variance when compared to the prior distribution. We derive a model where all uncertainty is present in the mean value of the distribution, a model that resemble Garlappi-Uppal Multiprior model. We show that the resulting portfolio is a combination of the mean variance portfolio with the minimum variance portfolio. When no uncertainty is present, the agent chooses the classic mean variance portfolio. On the other hand, when there is too much uncertainty, the agent consider the prior mean uninformative and choose the minimum variance portfolio. We also show that Julliard Information Portfolio can be regarded as a particular case of the more general robust asset allocation problem where no prior information is known about the Utility. Finally, we extend his model to consider different discrepancy functions in the broader class of Cressie Read which leads us to a class of "Information Portfolios".
3.1 Introduction

In constructing optimal portfolios the initial modern theory is generally attributed to Markowitz (1952) with its mean variance portfolio model. Since then, the theory has evolved in different directions. We point out here a few of these directions and where this paper fits in. One direction recognizes the difficulties for estimating the moments of the return distribution with sample moment estimators and try to look for more efficient estimators. In this class of models like in Jorion (1986) we have the class of shrinkage estimators that shrinks the sample average towards a different direction in order to look for an estimator with lower estimation risk. Jagannathan and Ma (2003) study the relationship between short selling constraints with shrinkage estimators. A different approach is based on prior beliefs about the return distribution and update these beliefs with a Bayesian learning. This so called Bayesian models proposed by Klein and Bawa (1976), Kandel and Stambaugh (1996), and Pastor (2000) look for estimators of the conditional expected returns that can be computed from the updated return distribution. Some of this Bayesian models, like Barberis (2000), Pastor and Stambaugh (2000), and Avramov (2004), incorporate beliefs about asset pricing models and how it affects the moments of the predictive distribution. Another set of articles look for measuring the effects of adding a third moment to the agent portfolio problem. The last class of models, the one that we are going to emphasize here is the class of robust models.

Robust models have been gaining increasing importance in the assessment of economic models. An extensive coverage on applications of robust models to economic problems can be found in Hansen and Sargent (2008) and the more recent Hansen and Sargent (2014). The main idea is that the true possible outcomes, that is, the true distribution of further events are not known a priori. The individuals have an idea of the distribution, which is known as the ”a priori” distribution, and must consider different possible distributions, or scenarios that may occur by computing deviations from its prior. An individual is said to have robust preferences if he considers the worst case scenario, that is, the worst possible distribution in terms of utility for a given set of possible alternative scenarios. Considering different scenarios is the main concern for risk evaluations. There are a lot of reasons to do so, such as concerns for model misspecification and difficulties from estimating the true parameters of the model.

The possibilities of different distributions considered in the model is known as model risk. In this case, the individual that consider the worst case distribution is trying to minimize the model risk. Some times model risk is addressed by comparing the results of different models as in comparative statics. We follow here the approach developed in Hansen and Sargent (2008), Hansen et al. (2006), Petersen et al. (2000), and Glasserman and Xu (2014). They treat the robust control problem as a 2 person zero-sum game. The game involve two distinct players, the agent and the nature. The objective of the nature is to change the distribution of events with respect to the prior of the agent and choose the worst possible distribution for the agent. The individual problem is to optimize performance against the worst case measure change imposed by the adversary. The degree of robustness is usually determined through either a constraint or a penalty on relative entropy that limit the nature in making the change of measure arbitrarily bad.
The main contribution in this paper is to show how a more general theory on robust decision problems applies in the particular case of a portfolio selection problem where agents only consider the first two moments of the return distribution and take the prior distribution of returns to be Gaussian. We show how robust optimization theory can be used to deal with estimation error and model misspecification and how available information can be used to improve efficiency by imposing correct constraints in the model. In particular, we derive a covariance constrained robust portfolio model and show that it fits in the class of shrinkage estimators. The results are intuitive. When the agent is uncertainty averse and there is no restrictions in the moments of the worst case distribution, we find that the distorted worst case scenario will keep normality, but with a decreased mean and enlarged variance. On the other hand, when the covariance is constrained to be equal to the prior distribution (e.g. when there is enough precision in the second moment estimation), all uncertainty is dealt to the mean, and the a posteriori distribution will be normal with equal variance, but decreased mean. The size of the diminished mean will depend on the agents level of uncertainty aversion. We also show that if there is no uncertainty aversion, that is, the agent has full confidence in the return distribution, the robust allocation will reverse to the traditional mean variance portfolio case. On the other hand, if the agent is extremely uncertain about the return distribution, the estimated mean is considered uninformative, and the robust portfolio converges to the minimum variance portfolio. We also show that this framework is capable of dealing with parametric and non parametric asset allocation models.

The rest of the paper is organized as follows. Section 2 introduces the traditional approach to optimization problems and its application to portfolio selection in a mean variance framework. Section 3 describes the robust approach to the original problem. In section 4 we show how moments constraints can be inserted into the robust optimization model and how it allows us to derive a robust mean variance portfolio where all uncertainty comes from the estimated mean. Section 5 concludes.

3.2 The Traditional Optimization Problem

Optimization problems are in the core of economic studies. We seek to optimize the scarce allocation of resources in the aim to achieve an optimum that varies depending on the subject of the analysis, that is, on the objective function. In many cases, the objective function depends on random elements. Considering that, $X$ denote the stochastic elements of a model, $a$ its control variables, and $V(a, X)$ denote the objective function that we want to optimize. Since $X$ is random, $V(a, X)$ is also random and the problem is usually specified in terms of its expected value $E[V(a, X)]$, where the expectation is taken with respect to the distribution of $X$. The problem of interest is formulated as

$$\sup_a E[V(a, X)]$$

where $a$ belongs to a certain set $A$ that constraint the possible controls. With some standard requirements the solution to the problem can be found by solving the first order
This first order condition determines the optimal decision $a^*$ that makes the agent achieve the maximum performance $E[V(a^*, X)]$.

Note that to the problem be well defined, we must specify the random elements relevant to the problem, $X$; the control variables, $a$; the set of restrictions that determines the choice set $A$; the objective function $V(a, X)$; and finally, the correct distribution of $X$.

### 3.2.1 The Mean Variance Portfolio Problem

In a mean-variance framework the objective function of an investor is given by a particular expected utility function. In particular, the individual utility derived from a portfolio choice $a$ is given by the expected value of

$$V(a, X) = a'X - \frac{\delta}{2} a'X'Xa$$

where $X$ is the random vector of assets excess returns and $a = (a_1, ..., a_n)^T$ are the portfolio weights. In this case the individual problem is to find the optimal portfolio weights $a^*$ that maximizes its expected utility. Since $a$ represents weights of portfolio its natural to restrict the choice to the set $A = \{a \in \mathbb{R}^n : \sum_{i=1}^n a_i = 1\}$. If we were to consider a no short selling restriction we should require $a$ to be chosen from $A^+ = \{a \in \mathbb{R}^n : \sum_{i=1}^n a_i = 1\}$. It is also clear from this problem that the expected utility depends only on the first 2 moments of the assets excess returns, $E[X]$ and $E[X'X]$ respectively. A common choice for the distribution of $X$ is the normal distribution, which depends only on this 2 parameters to be fully characterized.

If the distribution of $X$ were known for sure by the agent, then the choice $a^* = a(\theta)$ would be easily attained by solving the quadratic optimization problem

$$E[V(a, X)] = a'\mu + \frac{\delta}{2} a'\Sigma a$$

where $\mu = E[X]$ and $\Sigma = E[XX']$ are the two first moments of the random vector of excess returns $X$. To find out the optimal portfolio decision, we need only to apply the first order condition to the problem

$$\partial_a E[V(a^*, X)] = \mu - \delta \Sigma a^* = 0$$

In practice, the agent does not know the true distribution of returns. We advocate he must consider this uncertainty in such a way that allows him to deal with different scenarios than prior considered.
3.3 Incorporating Model Uncertainty

It is not difficult to incorporate model uncertainty into the problem. To better understand the procedure, let again $V(a, X)$ be a random quantity of interest that depends on the control $a$. Suppose the agent wish to maximize its expected value with respect to $a \in A$. The agent problem is

$$
\sup_{a \in A} E[V_a(X)]
$$

where $V(a, X)$ is a borel measurable function of the random vector $X$ and $A$ is a convex and compact set of a normed reflexive vector space. We consider now a modification of the original problem, where the density function $f(x)$ of $X$ is not known by the agent, but instead, a prior is considered. Since the agent knows that this prior may not be the true density, she may consider the true density to be in a neighborhood of the prior. This neighborhood is better defined through an auxiliary variable $m(x) = \tilde{f}(x)/f(x)$. The neighborhood is then defined as $P(\eta) = \{m \in C[-\infty, +\infty] : m > 0, E[m] = 1 \text{ and } R(m) \leq \eta\}$.

Where

$$
R(m) = E[m \log m] = \int \frac{\tilde{f}(x)}{f(x)} \log \frac{\tilde{f}(x)}{f(x)} f(x) dx
$$

It is easy to see that $R(m) \geq 0$ and $R(m) = 0$ if and only if $\tilde{f}$ and $f$ coincide almost everywhere. In this sense, we may consider the quantity $R(m)$ as a measure of deviation of the true distribution with respect to the prior. Note also that this restrictions on $m$ determines the choice set for the possible densities $\tilde{f} = mf$. We left in the appendix the proof that $\tilde{f}$ is in fact a continuous density function of the random variable $Y = g(X)$ where $g : \mathbb{R} \to \mathbb{R}$ is a continuous function.

We consider then the robust version of the original problem as

$$
\sup_{a \in A} \inf_{m \in P(\eta)} E[m(X)V_a(X)]
$$

This is the so called constrained problem defined by Hansen and Sargent. It is clear from the above problem that if $\eta = 0$, $m(X) = 1$ and we are back to the non-robust original problem. This version of the problem can be seen as if the agent has no uncertainty about the distribution of $X$.

Using the Lagrange Duality Theorem from Luenberger (1997)(check appendix), we have that the solution of the inner problem is given by the solution of the dual problem

$$
\min_{\theta \geq 0, \mu \geq 0} \sup_{m \in M} L_a(m, \theta, \mu)
$$

where $L_a(m, \theta, \mu)$ is the Lagrangian of the inner problem, given by

$$
L_a(m, \theta, \mu) = \int [-m(x)V_a(x) - \frac{1}{\theta}(m(x) \log m(x) - \eta) - \mu(m(x) - 1)] f(x) dx
$$
And $M = C[-\infty, +\infty]$ is the vector space of all continuous functions with domain in $\mathbb{R}$. For given values $\theta \geq 0$ and $\mu \geq 0$, let $F(m) = L_a(m, \theta, \mu)$. The inner problem of this dual problem is

$$
\sup_{m \in M} F(m)
$$

Note that $F(m) = \int f(m(x), x)dx$ is a concave functional defined in the vector space $M$, since $f(m, x)$ is concave in $m$. Note also that if $F$ have a Gateaux differential on the vector space $M$, a necessary condition for $F$ to have an extremum at $m^* \in M$ is that $\delta F(m^*, h) = 0$ for all $h \in M$. Where $\delta F(m, h) = \frac{d}{d\alpha} F(m + \alpha h)|_{\alpha=0}$.

Now, let $H(m, x)$ be such that $F(m) = \int_{-\infty}^{+\infty} H(m(x), x)dx$, where we assume $H_m$ exists and is continuous with respect to $m$ and $x$. Then, we show in the appendix that a necessary and sufficient solution for $F$ to achieve a maximum at $m^* \in M$ is given by

$$
H_m(m^*(x), x) = 0, \quad \text{for all } x \in \mathbb{R}
$$

it is not difficult to see that for our particular case,

$$
H(m, x) = [-m(x)V_a(x) - \frac{1}{\theta}(m(x) \log m(x) - \eta) - \mu(m(x) - 1)]f(x)
$$

and the solution of the inner layer of the dual problem is achieved at

$$
m^*_\theta = \exp \frac{-\theta V_a(X)}{E[\exp(-\theta V_a(X))] > 0}
$$

where we consider here that $E[\exp(-\theta V_a(X))] < \infty$. It is clear that the value of $m^*$ does not depend on $\mu$. It is also not difficult to see that the value of $\mu$ is not relevant to the solution of the second layer of the optimization problem, so that we can set it to zero. By doing the obvious redefinition of the function of interest we get

$$
L_a(m^*_\theta, \theta) = \frac{1}{\theta} \log E[\exp(\theta V_a(x))] + \frac{\eta}{\theta}
$$

and the original problem reduces to

$$
\sup_{a \in A} \min_{\theta \geq 0} L_a(m^*_\theta, \theta)
$$

Now, since $A$ is a compact, convex subset of the reflexive normed space. If we assume $\theta \in \Theta = [0, \bar{\theta}]$, which is also a compact and convex subset, then we can interchange the order of the optimization, that is

$$
\sup_{a \in A} \min_{\theta \in \Theta} L_a(m^*_\theta, \theta) = \min_{\theta \in \Theta} \sup_{a \in A} L_a(m^*_\theta, \theta)
$$
This change of order does not alter the resulting extremizing choices of \((a, m)\). For a reference on why this is permissible we refer to Hansen et al. (2006). In our portfolio problem the primary quantity of interest is the random quantity

\[ V_a(X) = a^T X - \frac{\gamma}{2} X^T aa^T X \]

Where \(a\) is the portfolio that must be chosen by the investor, \(A = \{a \in \mathbb{R}^n : a_i \leq a, i = 1, \ldots, n, \text{ and } \sum_{i=1}^n a_i = 1\}\), and \(X\) is the vector of unknown excess return distributions with normal density prior.

As we have shown above, the solution to this problem takes the form

\[ m^* \propto \exp -\theta V_a(X) = \exp \{-\theta a^T X + \theta \frac{\gamma}{2} X^T aa^T X\} \] (34)

Also, since the prior density is a normal density, that is

\[ f \propto \exp -\frac{1}{2}(X - \mu)^T \Sigma^{-1}(X - \mu) \]

by multiplying the density by the solution, we get the modified density,

\[ m^* f \propto \exp -\frac{1}{2}(X - \tilde{\mu})^T \tilde{\Sigma}^{-1}(X - \tilde{\mu}) \]

where

\[ \tilde{\Sigma} = (\Sigma^{-1} - 2\theta \gamma aa^T)^{-1} \]
\[ \tilde{\mu} = \mu - \theta \Sigma a \]

This is the density of the worst case scenario, when the prior is normal with mean \(\mu\) and covariance matrix \(\Sigma\). It is easy to see that the worst case change of measure preserves normality, that is, in the worst-case scenario, \(X \sim N(\tilde{\mu}, \tilde{\Sigma})\). It is important to note that we did not impose the normality as a constraint, the fact that worst case change of measure is still normal is a result of the optimization problem. This illustrates the main idea that we are interested in understand the change in the probability law that is associated with the worst case scenario. It is also important to note that the worst case change of measure is a function of the portfolio weights, which is a common dependence present in a 2 person 2 stage zero-sum game where the optimal solution of one agent depends on the choices to be considered by the adversary.

Finally, by the change of order of the optimization process, the optimal portfolio choice is given by \(a^* = a^*(\theta^*)\), where

\[ a^*(\theta) = \arg \inf_a \frac{1}{\theta} \log E[\exp(\theta V_a(X))] \] (35)
and where $\theta^*$ solves the outer layer of the optimization process, that is,

$$
\theta^* = \arg \inf_{\theta} \frac{1}{\theta} \log E[\exp (\theta V_{\theta}(X))] + \frac{\eta}{\theta}
$$

(36)

It is common in the literature to consider $a^*(\theta)$ as a function of $\theta$ and compute the optimal values numerically by each distinct value of $\theta$.

### 3.3.1 The Worst Case Scenario

Let's now interpret the worst case scenario in this case. Since the portfolio excess return is given by $r_p(a) = a'X$. It is clear that when considering the worst case scenario the agent is considering the case of a lower expected return and increased variance when comparing with the prior case. To see this, just note that the portfolio expected return considered by an uncertainty averse investor is

$$
\tilde{\mu}_p = \tilde{E}[a'X] = a'\tilde{E}[mX] = a'\tilde{\mu} = a'(\mu - \theta\tilde{\Sigma}a) = a'\mu - \theta a'\tilde{\Sigma}a = \mu_p - \theta \tilde{\sigma}^2_p
$$

where the variance of the portfolio return with respect to the adjusted distribution is

$$
\tilde{\sigma}^2_p = a'\tilde{\Sigma}a = a'(\Sigma^{-1} - 2\theta\gamma a\gamma^T)^{-1}a > a'\Sigma a = \sigma^2_p
$$

The formal proof of this last inequality is left in the appendix. It is important to note here that the shadow price $\theta$ can be interpreted as the quantity of uncertainty, where as higher the uncertainty $\theta$, smaller is the portfolio expected return considered by the agent and greater is the portfolio variance when compared to the prior model ($\theta = 0$).

### 3.4 Incorporating Expectation Constraints

In many cases we have more information about some moments of the distribution of the unknown and want to take advantage of this knowledge. As an example, it is usually easier to infer about the covariance of returns then it is to do so about the mean. In these cases it is useful to restrict the uncertainty with respect to the moments that more information is available. It can be done by imposing restrictions of the kind $E[mh_i(X)] \leq \eta_i$ or $E[mh_i(X)] = \eta_i$. As shown by Glasserman and Xu (2013) with results of Peterson et al. (2000), we can solve the following robust constraint problem

$$
\sup_{m \in \mathcal{P}} E[mV(X)]
$$

(37)
where
\[
\mathcal{P} = \{m \in M : E[m \log m] \leq \eta, E[m h_i(X)] \leq \eta_i, i = 1, \ldots, n, \text{ with } \eta_i, \eta \in [0, \infty)\} \quad (38)
\]

by solving its dual multiplier problem
\[
\inf_{\theta > 0, \lambda_i > 0} \sup_{m \in M} E \left[ mV(X) - \frac{1}{\theta} (m \log m - \eta) - \sum_{i=1}^{n} \lambda_i [m h_i(X) - \eta_i] \right] \quad (39)
\]

by combining the terms of \( h_i \) with \( V \) it is easy to see that the worst change of measure given by the solution of the first layer of the optimization problem is
\[
m_\theta \propto \exp \left( \theta \left[ V(X) - \sum_{i=1}^{n} \lambda_i h_i(X) \right] \right) \quad (40)
\]

That means that the optimal decision is given by \( a^*(\theta^*, \lambda^*_i) \), where
\[
a(\theta, \lambda_i) = \arg \inf_{a \in A} \frac{1}{\theta} \log E \left[ \exp \left( \theta \left[ V_a(X) - \sum_{i=1}^{n} \lambda_i h_i(a, X) \right] \right) \right] + \sum_{i=1}^{n} \eta_i \lambda_i \quad (41)
\]

That is, we can solve the first layer of the optimization process, plug it in the solution in the objective function and then solve the second layer of the optimization problem of interest. This is a common procedure in robust optimization problems where the agent must choose the best action possible given the worst case scenario chosen by the nature. To give the complete solution we need to determine the optimal values of the lagrange multipliers. By the same arguments used above, we have that \((\theta^*, \lambda^*_i)\) solves
\[
\inf_{\theta > 0, \lambda_i > 0} \frac{1}{\theta} \log E \left[ \exp \left( \theta \left[ V_a(X) - \sum_{i=1}^{n} \lambda_i h_i(a, X) \right] \right) \right] + \frac{\eta}{\theta} + \sum_{i=1}^{n} \eta_i \lambda_i
\]

3.4.1 Mean Constrained Optimal Portfolio

Glasserman and Xu consider the case where the vector of excess returns \( X \) has a normal density prior and the mean vector is constrained. In this way, all the uncertainty is limited to the covariance matrix, which leads to the robust problem
\[
\sup_a \inf_m E[m(a^T X - \frac{\gamma}{2} a^T (X - \mu)(X - \mu)^T a) + \frac{1}{\theta} (m \log m - \eta)] \quad (42)
\]
Glasserman and Xu show that in this case the worst case likelihood ratio is

\[ m^* \propto \exp\left( \frac{\theta^* \gamma a^* T (X - \mu)(X - \mu)^T a^*}{2} \right) \]  

This follows straight from our previous results. By plugging in the constraint in the objective function through the Lagrangian form, we get that the worst case change of measure is given by

\[ m^* \propto \exp\left( \theta^*(V_{a^*}(X) - \lambda^* T (X - \mu)) \right) \]  

where, by setting \( h(a, X) = X \) and \( \eta = \mu \), \((\theta^*, \lambda^*)\) must solve the optimization problem

\[
\inf_{\theta > 0, \lambda > 0} \frac{1}{\theta} \log E \left[ \exp \left( \theta [V_{a^*}(X) - \lambda^T X] \right) \right] + \frac{\eta}{\theta} + \lambda^T \mu
\]

The trick here is that we don’t need to solve for this problem. Since the term \( \lambda^T X \) is linear in \( X \) and we have constrained the mean so that (14) must hold. By comparing (14), with (15), we find that \( \lambda^* = a \). Note that this implies that the worst case measure \( \tilde{f} \) has \( X \sim N(\mu, \tilde{\Sigma}) \), where \( \tilde{\Sigma}^{-1} = \Sigma^{-1} - \theta^* \gamma a^* a^T \). To find the optimal portfolio they propose to numerically solve

\[
a_{GX}(\theta) = \arg \inf_{a \in A(\theta)} \frac{1}{\sqrt{\det(1 - \theta \gamma aa^T \Sigma)}} + a^T \mu
\]

where the set \( A(\theta) = \{ a : \Sigma^{-1} - \theta \gamma aa^T > 0 \} \) is the set of portfolio weights that ensures that the resulting covariance matrix is positive definite.

### 3.4.2 Covariance Constrained Optimal Portfolio

It is easy to see that model uncertainty results in an augmented variance, and that, this augmented variance is result of uncertainty on both the mean and covariance of returns. We have seen that, if we restrict the mean so that \( \tilde{\mu} = E[mX] = \mu \), we get Glasserman-Xu mean variance robust portfolio problem. This model consider the expected returns as available information and all the uncertainty is due to the unknown covariance. The more interesting situation is though the one that the mean is unknown, since the mean returns is usually harder to estimate. We will consider here this second case, assuming that the covariance is known and show how this optimization problem has
a straight link with Garlappi Uppal Multi prior model.

Intuitively, if we instead constraint the variance, so that $\tilde{\Sigma} = E[XX^T] = \Sigma$, we let all the model uncertainty to the mean. By substituting the restriction in the objective function, letting only the uncertainty to the mean, we get

$$\sup_a \tilde{E}[V(a, X)]$$

where

$$\tilde{E}[V(a, X)] = a^T \tilde{\mu} - \frac{\gamma}{2} a^T \tilde{\Sigma} a = a^T \mu - \frac{\gamma}{2} a^T \Sigma(1 + 2\theta/\gamma) a$$

Since we can interchange the order of the optimization process, we get the robust covariance constrained portfolio

$$a^*(\theta) = \frac{1}{\gamma} \Sigma^{-1} \left( \frac{1}{1 + \theta} \right) \left( \hat{\mu} - \frac{B - \gamma}{A} \frac{1}{1_N} \right)$$

(46)

It is also nice to note that as $\theta \to 0$, the optimal weights converges to the mean-variance portfolio

$$a^*(0) = \frac{1}{\gamma} \Sigma^{-1} \left( \hat{\mu} - \frac{B - \gamma}{A} \right) 1_N$$

(47)

where $\mu^0 = \frac{B - \gamma}{A}$ is the expected return on the zero beta portfolio associated with $a$. On the other hand, as $\theta \to \infty$, the optimal portfolio converges to the minimum variance portfolio

$$a^*(\infty) = \frac{1}{A} \Sigma^{-1} 1_N = a_{MIN}$$

(48)

It is remarkable the resemblance of this model with Garlappi-Uppal multiprior model. Garlappi-Uppal suggested interpreting uncertainty of the mean as a confidence interval for the estimated sample mean $\hat{\mu}$.

To do so, they propose solving the following problem

$$\max_a \min_\mu \{a^T \mu - \frac{\gamma}{2} a^T \Sigma a\}$$

(49)
subject to

\[(\hat{\mu} - \mu)^T \Sigma^{-1} (\hat{\mu} - \mu) \leq \epsilon \]  \hspace{1cm} (50)

which is equivalent to solve (see Garlappi-Uppal)

\[
\max_a \left\{ a^T \hat{\mu} - \frac{\gamma}{2} a^T \Sigma \left(1 + \frac{2\sqrt{\epsilon}}{\gamma \sqrt{a^T \Sigma a}}\right) a \right\}
\]  \hspace{1cm} (51)

This means that, if we are uncertain about the true value for the mean, we can interpret this uncertainty by determining a level of confidence for the true mean \( \mu \) in terms of deviation where the degree of deviation is set by the uncertainty parameter \( \epsilon \). To gain more insight on the degree of confidence we must seek knowledge about the true distribution from the Data. To understand the gain of knowledge from the data, suppose that expected returns are estimated by their sample mean \( \hat{\mu} \). If returns are drawn from a normal distribution, then the quantity

\[
\frac{T(T-N)}{(T-1)N} (\hat{\mu} - \mu)^T \Sigma^{-1} (\hat{\mu} - \mu)
\]

has a \( \chi^2 \) distribution with \( N \) degrees of freedom and an F-distribution with \( N \) and \( T-N \) degrees of freedom if \( \Sigma \) is not known. This means that for a chosen quantile \( \epsilon \) from the F-distribution, the constraint

\[
\frac{T(T-N)}{(T-1)N} (\hat{\mu} - \mu)^T \Sigma^{-1} (\hat{\mu} - \mu) \leq \epsilon
\]

corresponds to the probabilistic statement

\[
P \left( \frac{T(T-N)}{(T-1)N} (\hat{\mu} - \mu)^T \Sigma^{-1} (\hat{\mu} - \mu) \leq \epsilon \right) = 1 - p
\]

for some appropriate level \( p \). We can interpret \( \epsilon \) as how distant is the population mean \( \mu \) from the estimated \( \hat{\mu} \) with some level of confidence. It is also nice to see that this is also the probabilistic restriction of expected loss from estimation risk. To see this, we only need to note that the expected loss from estimation risk when \( \Sigma \) is known, but \( \mu \) is unknown is

\[
\mu' \Sigma^{-1} \hat{\mu} - \frac{1}{2} \hat{\mu}' \Sigma^{-1} \hat{\mu} - \frac{1}{2} \mu' \Sigma^{-1} \mu = 2(\hat{\mu} - \mu)^T \Sigma^{-1} (\hat{\mu} - \mu)
\]

From Garlappi-Uppal, we know that the expression for the optimal portfolio weights can
be written as

$$a_{GU}(\varepsilon) = \frac{1}{\gamma} \Sigma^{-1} \left( \frac{1}{1 + \frac{\varepsilon}{\gamma \sigma_p}} \right) \left( \bar{\mu} - \gamma \left( 1 + \frac{\sqrt{\varepsilon}}{\gamma \sigma_p} \right) \right) \frac{1}{N} \right)$$

(52)

where $A = \frac{1}{N} \Sigma^{-1} 1_N$, $B = \hat{\mu}^T \Sigma^{-1} 1_N$, and $\sigma_p = a^T \Sigma a$ is the standard deviation of the optimal portfolio that can be found by solving for the unique positive real solution $\sigma_p^*$ of the following polynomial equation

$$A \gamma^4 \sigma_p^4 + 2A \gamma \sqrt{\varepsilon} \sigma_p^3 + (A \varepsilon - AC + B^2 - \gamma^2) \sigma_p^2 - 2 \gamma \sqrt{\varepsilon} \sigma_p - \varepsilon = 0$$

It is also easy to see that, after some algebraic manipulation, we can write Garlappi Uppal optimal portfolio as a convex combination between the global minimum portfolio and the market portfolio. To be more precise,

$$a_{GU}(\varepsilon) = \phi_{GU}(\varepsilon) a_{MIN} + (1 - \phi_{GU}(\varepsilon)) a_{MV}$$

where

$$\phi_{GU}(\varepsilon) = \left( \frac{\sqrt{\varepsilon}}{\gamma \sigma_p^* + \sqrt{\varepsilon}} \right)$$

or, if we consider the robust portfolio problem with covariance constraint, we get

$$a(\theta) = \phi(\theta) a_{MIN} + (1 - \phi(\theta)) a_{MV}$$

where

$$\phi(\theta) = \left( \frac{\theta}{\gamma + \theta} \right)$$

With this formulation of the optimal portfolio, we can see that the uncertainty aware with variance constraint portfolio can also be regarded as a shrinkage estimator for the mean, with the mean shrinking towards the minimum variance portfolio as the level of uncertainty $\theta$ goes up. The intuition is simple, since all model uncertainty comes from the estimated mean, when the uncertainty of the estimated mean is large, the interval of confidence is large, then the investor relies less on the estimated mean and reduces the weights invested in the risky asset. If uncertainty goes to infinity, no information about the mean is considered and the investor invests in the minimum variance portfolio. If the investor has total confidence in the estimated value, that is, uncertainty is zero, then he invest in the classical mean-variance portfolio. For a thorough analysis of the class of shrinkage estimators we refer to Wang (2005).
3.4.3 Robust Optimization with Pricing Constraints

So far, despite the flexibility of the model to incorporate any specific objective function, we have restricted ourselves to the specific mean variance utility for its simplicity that allows a better insight of the method. In many cases though, we don’t want to restrict ourselves to any sort of parametric specification of the Utility. To skip the trouble of a misspecification problem we can impose restrictions directly in the stochastic discount factor, that is, we may only impose that whatever is the actual target function, the resulting factor prices the model. We will see now how this semi-parametric model that does not specify an a priori utility can be considered as a particular case of the robust optimization approach. From the multiplier problem

$$\sup_m E \left[ mV(X) - \frac{1}{\theta}(m \log m - \eta) - \sum_{i=1}^{n} \lambda_i [mh_i(X) - \eta_i] \right]$$

with known solution

$$m_\theta \propto \exp \left( \theta \left[ V(X) - \sum_{i=1}^{n} \lambda_i h_i(X) \right] \right)$$

consider the case where the random quantity $X$ is the vector of excess returns. Also, consider $V(X) = 0$, $h_i(X) = X_i$, and $\eta_i = 0$. That is exactly the asset pricing problem considered in Gosh and Julliard (2016) of solving

$$\inf_m E [m \log m]$$

subject to

$$E[mX] = 0$$

or equivalently, its dual problem

$$\inf_m E \left[ m \log m - \sum_{i=1}^{n} \theta_i mX_i \right]$$

where $\theta_i = \theta \lambda_i$. Which yields the known solution

$$m_\theta = \frac{\exp (\theta'X)}{E[\exp (\theta'X)]}$$

Where $\theta \in \mathbb{R}^N$ is the vector of lagrange multiplier that solves the unconstrained convex problem

$$\theta = \arg \max_\theta E[m_\theta \log m_\theta - m_\theta \theta'X]$$

or equivalently

$$\theta = \arg \min_\theta E[\exp (\theta'X)]$$
It is also interesting to note that in this case the radon-nikodyn derivative is interpreted as the individual stochastic discount factor, whereas in the general it may not be the case. Moreover, in the familiar mean variance framework, the stochastic discount factor is linear in the portfolio return, whereas here it takes an exponential form. Since the stochastic discount factor in this case is not a traded portfolio, they consider its projection in the portfolio space which is the portfolio \( \hat{m} = a + b'X \) where \( \{a, b\} \) solves the minimization problem

\[
\{a, b'\} = \arg \min \sum_{t=1}^{T} (m_t - (a + b'X_t))^2 \tag{61}
\]

After obtaining this mimicking stochastic discount factor \( \hat{m} \), they construct a portfolio that is inversely correlated with the stochastic discount factor by setting \( a = -b/|b| \) which they call the Information Portfolio.

As we have mentioned, in our general framework, \( m \) may not be the stochastic discount factor since it only represents a change of measure. We can find the stochastic factor by solving the second layer of the optimization problem, that is

\[
m = \frac{U'(W_{t+1})}{U'(W_t)} g \tag{62}
\]

where \( W_{t+1} \) is the random quantity of interest and \( W_{t+1} = W_t(1 + a'X) \) and \( g \) solves the inner optimization problem, that is,

\[
g_\lambda = \exp \left( \lambda U(W_{t+1}) \right) \frac{\mathbb{E}[\exp (\lambda U(W_{t+1}))]}{\mathbb{E}[\exp (\lambda U(W_{t+1}))]} \tag{63}
\]

Gosh and Julliard investigate this Information Portfolio and show that it has high out-of-sample Sharpe ratio, outperforming both the 1/N naive portfolio and Value and Momentum strategies combined.

### 3.4.4 The Class of Information Portfolios

Almeida and Garcia (2012), consider the case where the change of measure is constrained by a more general discrepancy function. The idea is that the entropy restriction only allows specific changes of measure that may not capture all the possible different scenarios, especially when assets returns possess higher moments dependencies. The idea is to consider a generic convex function \( \phi \). And the objective is to solve

\[
\inf_{m} \mathbb{E}[\phi(m)] \tag{64}
\]

subject to

\[
\mathbb{E}[mX] = 0 \tag{65}
\]
or equivalently, its dual problem

$$\inf_{m} E \left[ \phi(m) - \sum_{i=1}^{n} \theta_{i} m X_i \right]$$  \hspace{1cm} (66)$$

the first order condition to this variational problem is

$$\phi'(m) - \sum_{i=1}^{n} \theta_{i} X_i = 0$$

$$\sum_{i=1}^{n} \theta_{i} E[m X_i] = 0$$

Almeida and Garcia explores the nice properties of a particular discrepancy function, the Cressie Read family described by $$\phi_{\gamma}(m) = \frac{m^{\gamma+1} - 1}{\gamma(\gamma+1)}$$. This family has the Empirical Likelihood ($$\gamma = -1$$), Exponential Tilting ($$\gamma = 0$$), Pearson’s Chi-Square ($$\gamma = -2$$), and Hellinger’s distance ($$\gamma = -1/2$$) as particular cases. By substituting this function on the first order condition we get

$$m_{\gamma} = (\gamma \sum_{i=1}^{n} \theta_{i} X_i)^{1/\gamma}$$  \hspace{1cm} (67)$$

Note that if there is a risk-free asset, then we can simplify the expression to

$$m_{\gamma} = (1 + \gamma \sum_{i=1}^{n} \tilde{\theta}_{i} X_i)^{1/\gamma}$$  \hspace{1cm} (68)$$

It is easy to see that, taking $$\gamma \to 0$$, in which case $$\phi(m) = m \log m$$, we get the special case where

$$m_{\gamma=0} = \exp(\sum_{i=1}^{n} \tilde{\theta}_{i} X_i)$$  \hspace{1cm} (69)$$

Notice now that we can take the stochastic discount factor projection in the portfolio space we find the portfolio $$\hat{m}_{\gamma} = a_{\gamma} + b_{\gamma}' X$$ where $$\{a_{\gamma}, b_{\gamma}'\}$$ solves the minimization problem

$$\{a_{\gamma}, b_{\gamma}'\} = \arg \min \sum_{t=1}^{T} (m_{\gamma t} - (a_{\gamma} + b_{\gamma}' X_t))^2$$  \hspace{1cm} (70)$$

After obtaining this mimicking stochastic discount factor $$\hat{m}_{\gamma}$$, we construct a portfolio that is inversely correlated with the stochastic discount factor by setting $$a = -b/|b|$$ which we call the Class of Information Portfolios, since it depends on the specified value for $$\gamma$$.

70
3.5 Conclusion

The theory of robust decision applied to economic problems has been growing fast for the past few years. A good deal of this expansion in the economic field is due to the efforts of Hansen and Sargent that have already published at least 2 in depth treatment of the subject. We have shown how the basic mathematical tools as the Lagrange Duality Theorem suffices to allow for different discrepancy models. One of our contribution was to show here how to apply the robust decision approach to a simple mean variance portfolio problem. In particular, we showed that by taking relative entropy as discrepancy measure, the worst case model is characterized by an exponential change of measure. That means that if the prior distribution has a normal density, and the objective function is quadratic, the worst case model will also have normal density. The difference from the prior density is that the worst case density has a lower expected mean and an enlarged variance. We also showed that, by constraining uncertainty exclusively to the mean, we can derive a model that resemble Garlappi Uppal multiprior model. Finally, we showed that the same robust setup can be used to account for semi parametric asset allocation models.
Appendix

Proof of Enlarged Variance

In the worst case change of measure, we have \( \hat{\Sigma} = (\Sigma^{-1} - 2\theta \gamma a a^T)^{-1} \). A particular partitioned inverse result gives \( (A + UCV)^{-1} = A^{-1} - A^{-1} U(C^{-1} + VA^{-1})^{-1} V A^{-1} \).

We are interested in the particular case where \( U = V = I \). That is, \( (A + C)^{-1} = A^{-1} - A^{-1}(C^{-1} + A^{-1})^{-1} A^{-1} \). Letting \( A = \hat{\Sigma} \) and \( C = -\Sigma \), we have

\[
(\hat{\Sigma} - \Sigma)^{-1} = \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}(\hat{\Sigma}^{-1} - \Sigma^{-1})^{-1} \hat{\Sigma}^{-1}
\]

\[
= \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}(-2\theta \gamma a a^T)^{-1} \hat{\Sigma}^{-1}
\]

\[
= \hat{\Sigma}^{-1} + \frac{1}{2\theta \gamma} \hat{\Sigma}^{-1}(aa^T)^{-1} \hat{\Sigma}^{-1}
\]

Now, since \( A \) is positive definite, its inverse \( A^{-1} \) is positive definite. Then if \( \hat{\Sigma} \) is positive definite, so is \( \hat{\Sigma} - \Sigma \) and \( d (\hat{\Sigma} - \Sigma)a > 0 \).

Derivation of the Covariance Constrained Robust Portfolio

We have stated that if we use Hansen-Sargent model uncertainty restricting the covariance we get Garlappi Uppal Multi prior model. To see this, consider the robust problem

\[
\sup_{m \in M} -E[m(a^T X - \gamma/2 a^T (X - \mu)(X - \mu)^T a) - 1/\theta (m \log m - \eta)]
\]

s.t.

\[
E[m] = 1
\]

\[
E[m(e_i^T (X - \mu)(X - \mu)^T e_j)] = \sigma_{ij}, \quad i, j = 1, \ldots, n
\]

by the known results presented here, we have

\[
m^* \propto \exp(\theta(V_a(X) - \sum_{i,j} \lambda_{ij}(e_i^T (X - \mu)(X - \mu)^T e_j - \sigma_{ij}))
\]

Now, since \( \lambda_{ij} = \lambda_{ji} \) by symmetry, we write \( \lambda_{ij} = \lambda_i \lambda_j \). Letting \( \lambda = \sum_{i=1}^{n} \lambda_i e_i \), we have

\[
m^* \propto \exp(\theta(V_a(X) - \lambda^T (X - \mu)(X - \mu)^T \lambda))
\]

Now, since we have constrained the covariance, we must have \( \lambda = a \sqrt{\gamma/2} \), so that

\[
m^* \propto \exp(-\theta a^T X)
\]
Considering a normal prior, where
\[ f \propto \exp \left( -\frac{1}{2} (X - \mu)^T \Sigma^{-1} (X - \mu) \right) \]
the worst case density is given by
\[ f^* = m^* f \propto \exp \left( -\frac{1}{2} (X - \hat{\mu})^T \Sigma^{-1} (X - \hat{\mu}) \right) \]
where
\[ \hat{\mu} = \mu - \theta \Sigma a \]

The Lagrange Duality Theorem
To the problem be well defined, we only need to check the Theorem conditions. It is clear that \( M = C[-\infty, +\infty] \) is a vector space. It is also clear that \( \Omega = \{ m \in M : m > 0, E[m] = 1 \} \) is a convex subset of the vector space \( M \). We have that \( G(m) = \int \log m(x)f(x)dx - \eta \) is a convex mapping from \( \Omega \) into the normed space \( Z = \mathbb{R} \) which has a non-empty, closed, positive cone \( P = \mathbb{R}^+ \). There exist \( m_1 \in \Omega \) with \( G(m_1) < 0 \)(That is, \( m_1 \) is an interior point of \( N = -P \)). In fact, \( m_1 = 1 \) clearly satisfies this condition. To check that \( G : \Omega \to \mathbb{R} \) is in fact a convex mapping, note that for all positive values of \( m \in \mathbb{R} \), the function \( g(m) = m \log m \) is convex, since \( g''(m) = 1/m > 0 \). By the property of convex functions we have for all \( \lambda \in (0,1) \)
\[ g(\lambda m'(x) + (1-\lambda)m''(x)) \leq \lambda g(m'(x)) + (1-\lambda)g(m''(x)) \]
for all \( x \) such that \( m(x) > 0 \). Since \( m \in \Omega, m(x) > 0 \) for all \( x \in \mathbb{R} \). Multiplying both sides by \( f(x) \) and integrating over \( \mathbb{R} \), we get the result.

Concave Functionals
Let \( X \) be a vector space and \( F : X \to \mathbb{R} \) be a concave functional. For \( x, h \in X \), and \( \lambda > 0 \), define the functional
\[ g(\lambda; x, h) = \frac{F(x + \lambda h) - F(x)}{\lambda} \]
This functional has the following properties:
1. \( g(\lambda; x, h) \) is non-increasing in \( \lambda \)
2. \( g(\lambda; x, h) + g(\lambda; x, -h) \leq 0 \)
3. \( g(\alpha; x, h) \leq g(\lambda; x, h) \leq -g(\alpha; x, -h) \) for any \( \alpha \geq \lambda \).

4. The limit \( \lim_{\lambda \to 0} g(\lambda; x, h) \) exists and is defined as \( F'(x, h) \). We also have \( F'(x, h) \geq g(\lambda; x, h) \).

5. We also have that \( \lim_{\lambda \to 0} -g(\lambda; x, -h) \) exists and is defined as \( -F'(x, -h) \).

If \( F(x) \) assumes a maximum at \( x_0 \) if and only if \( F'(x_0, h) \leq 0 \) for all \( h \in X \).

**Derivative of Concave Functionals**

We say that the functional \( F(x) \) is Gateaux differentiable at a point \( x \in X \) if for all \( h \in X \), exists:

\[
\delta F(x, h) = \lim_{\lambda \to 0} \frac{F(x + \lambda h) - F(x)}{\lambda}
\]

Note that if we let \( g(\lambda) = F(x + \lambda h) \), for \( \lambda \in \mathbb{R} \). We have that \( \delta F(x, h) = g'(0) \). For the case of concave functionals, they are Gateaux differentiable if and only if \( F'(x, h) + F'(x, -h) = 0 \) for all \( h \in X \). In that case we have \( \delta F(x, h) = F'(x, h) \). Let \( F(x) \) be differentiable in the Gateaux sense at \( x_0 \in X \). Then \( F(x) \) to achieve a maximum at \( x_0 \), if and only if \( \delta F(x_0, h) = 0 \) for all \( h \in X \).

**Integral Concave Functionals**

Let \( f(x, t) \) be a concave and continuous function on \( x, t \in \mathbb{R} \). We assume that \( f_x(x, t) \) exists and is also continuous with respect to \( x \) and \( t \). Let \( X \) be the vector space of continuous functions \( x : [a, b] \to \mathbb{R} \). Define \( F : X \to \mathbb{R} \) as

\[
F(x) = \int_a^b f(x(t), t) dt
\]

Note that \( F(x) \) is a concave functional on \( X \). For any \( x, y \in X \) and \( \alpha \in [0, 1] \), given \( t \in \mathbb{R}, x(t), y(t) \in \mathbb{R} \) and

\[
f(\alpha x(t) + (1 - \alpha)y(t), t) \geq \alpha f(x(t), t) + (1 - \alpha)f(y(t), t)
\]

Integrating both sides with respect to \( t \) gives the result.

**Derivative of Integral Concave Functionals**

It is easy to see that under the above assumptions the Gateaux differential with respect to \( F \) exists and is given by

\[
\delta F(x, h) = \int_a^b f_x(x(t), h(t)) dt
\]
We know that \( F(x) \) assumes a maximum at \( x_0 \) if and only if for all \( h \in X \),

\[
\delta F(x_0, h) = \int_a^b f_x(x_0(t), t)h(t)dt = 0
\]

that is, if and only if

\[
f_x(x_0(t), t) = 0, \quad \forall t
\]

The Important Step

We will show that under the assumptions of the problem, \( \delta F(x, h) = \int_a^b f_x(x, t)h(t)dt \).

To see this note that for given \( x, h \in X \), and \( \lambda \in \mathbb{R} \), we have

\[
\left| g(\lambda; x, h) - \int_a^b f_x(x, t)h(t)dt \right| =
\left| \int_a^b \left[ \frac{f(x(t) + \lambda h(t), t) - f(x(t), t)}{\lambda} - f_x(x, t)h(t) \right] dt \right| \leq
(b - a)\bar{h} \sup_{t \in [a,b], \alpha \in [0,\lambda]} |f_x(x(t) + \alpha h(t), t) - f_x(x(t), t)|
\]

where \( \bar{h} = \sup_{t \in [0,1]} |h(t)| \). Taking \( \lambda \to 0 \), we have the result.

The Improper Case

We consider now the case where \( x : \mathbb{R} \to \mathbb{R} \) and \( F : X \to \mathbb{R} \) is defined as

\[
F(x) = \int_{-\infty}^{+\infty} f(x(t), t)dt
\]

Note that \( F(x) \) is a concave functional since it is the limit of the sequence \( F_n(x) = \int_{-n}^{+n} f(x(t), t)dt \) of concave functionals. For the same reason and for the fact that the sum of concave functionals being concave functionals, \( R_n(x) = F(x) - F_n(x) \) is also a concave functional. From the definition of \( R_n(x) \), we have

\[
F(x) = F_n(x) + R_n(x)
\]

with \( R_n(x) \to 0 \).

The Differentiability of the Improper Case

From the definition of \( F(x) \), \( F_n(x) \) and \( R_n(x) \), we can write the Gateaux differential of \( F \) as

\[
\delta F(x, h) = \delta F_n(x, h) + \delta R_n(x, h)
\]
Since $R_n(x)$ is a concave functional, $\delta R_n(x, h) = R'_n(x, h)$. Now, to show that

$$\delta F(x, h) = \lim \delta F_n(x, h)$$

we just need to show that $\lim R'_n(x, h) = 0$. Since $R_n(x)$ is concave for all $n$ and $R_n(x) \to 0$, the result holds.
4 Optimal Asset Allocation: How Many Miles are we From the Promised Gains?

Abstract

Garlappi Uppal tested different models in a mean variance portfolio problem and found out that with a large dataset, none were consistently better than the 1/N rule in terms of Sharpe ratio, certainty-equivalent return, or turnover. We evaluate here the performance of different portfolio strategies suggested by the literature with an up-to-date dataset and try to achieve a better portfolio strategy that consider not only out of sample performance but take into account the in sample results and estimation risk as possible references for a good portfolio criteria.
4.1 Introduction

In this paper we seek to answer the following questions: Does portfolio optimization models give good results for portfolio managers when compared to “naive” strategies? Is it important to incorporate uncertainty about return distribution when building a portfolio strategy? These are only a few but important questions that we are going to investigate here.

DeMiguel et al. (2009) evaluate the out-of-sample performance of optimal portfolio strategies and considered that none was consistently better than the $1/N$ rule in terms of Sharpe ratio, certainty equivalent return, or turnover. They conclude: "This suggests that there are still many miles to go before the gains promised by optimal portfolio choice can actually be realized out of sample". We evaluate their claim in light of recent data from industry portfolio returns and show that although no single portfolio strategy beat the naive strategy, the combination of the naive strategy with the minimum variance portfolio seems promising.

When evaluating an optimal strategy, there are some basic principles we must assume. What is the individual objective function? What set of information is available at the investment date? What is known and what is unknown and must be estimated from the data. What are the individuals believes about the unknown? All of that must be carefully treated when carrying on an investment analysis. More than that it is important to know how robust the model is with respect to the data. For instance if a model take very different results with only a little more or less of information it cannot be considered robust for a practical purpose. We are going to see that individual believes take a great deal in turning the portfolio model robust in the sense that it takes a great amount of data to change the perception of the investor which in turns bring some desirable robustness to the portfolio problem.

The problem of portfolio allocation in economics is an old and recurrent topic. The first appealing approach was due to Markowitz (1952). Markowitz described a preference ordering where individuals were eager for more return but dislike volatility. Since more returns reflects more consumption and volatility reflects consumption uncertainty, these should be the main concerns for any allocation problem. The partial ordering then implies that for any 2 portfolios, the one that has more expected return given a fixed level of bearable risk (variance) should be preferred, and given a fixed level of return, the portfolio with less variance should be the one preferred. One problem that comes along with this line of reasoning is that when an individual solves for the optimal portfolio he does not know the true mean or covariance of the portfolio assets. Black and Litterman (1992) points out that the resulting optimal portfolio is very sensitive to mean estimation. They propose a more robust estimation for the mean that allows to incorporate investors views. Glasserman and Xu (2014) incorporate individuals uncertainty about parameters estimate into the portfolio problem. Garlappi et al. (2007) Develop a model of Portfolio Selection with parameter and model uncertainty. Others like Jessica and Missaka consider the case if individuals should or should not time the market. The idea for market timing comes
from the fact that when more information is available, better predictability about the return distribution can be made. In this line, Amihud and Hurvich (2004), Ang, Andrew and Geert Bekaert, 2007 and Avramov (2004) show how some market information can be useful to predict stock return.

The article is divided as follows: Section 2 introduces the basic mean variance portfolio model, its known caveats, and extensions that seek to achieve reduced estimation risk; sections 3 presents a more detailed construction of the Bayesian models with particular description of Jorion Bayes-stein portfolio strategy and Pastor Stambaugh model of investors with asset pricing beliefs. Section 4 briefly describes the models to be tested; section 5 the empirical exercise and the performance criteria; section 6 presents the results; and for the last section we conclude.

4.2 The Basic Model and Its Extensions

We start analyzing the standard portfolio problem to assess its weaknesses and evaluate its ramifications. We consider here the case where the individuals maximize their expected utility of the terminal wealth

$$\mathbb{E}[U(W_T)]$$  

(71)

where the terminal wealth is given by

$$W_T = W_0(1 + r_p)$$

Letting

$$r_p = (1 - \sum_{i=1}^{n} a_i)r_f + \sum_{i=1}^{n} a_i r_i$$

or equivalently

$$r_p(a) = r_f + \sum_{i=1}^{n} a_i r_i^*$$

where

$$r_i^* = r_i - r_f.$$ It is easy to see that we can describe the preferences in terms of assets excess return with respect to the risk free rate. In particular, Kandel and Stambaugh (1996), Stambaugh (1999) and Barberis (2000), consider the case where the investor’s preferences over terminal wealth are described by constant relative-risk aversion power utility function of the form

$$v(W) = \frac{W^\gamma}{\gamma}$$  

(72)

Following Jurczenko and Maillet (2006), the individual problem is then

$$\max_a \mathbb{E}[U(W_0(1 + r_p(a)))]$$

which yields as first order condition

$$\mathbb{E}[U'(W)r_i^*] = 0$$
By doing a second Taylor expansion of the marginal utility over the expected wealth \( \bar{W} = W_0 (1 + \mu) \), where \( \mu = E[r_p] \), we get

\[
U'(W) \approx U'(\bar{W}) + U''(\bar{W}) W_0 (r_p - \mu) + \frac{1}{2} U'''(\bar{W}) W_0^2 (r_p - \mu)^2
\]

which yields the approximate first order condition

\[
E[r_t^e] + \frac{U''(\bar{W})}{U'(\bar{W})} W_0 E[r_t^e (r_p - \mu)] + \frac{1}{2} \frac{U'''(\bar{W})}{U'(\bar{W})} W_0^2 E[r_t^e (r_p - \mu)^2] = 0
\]

The term \(-U''(.)W/U'(.)\) is known as the Arrow Pratt relative risk aversion coefficient and the term \((1/2)U'''(.)W^2/U'(.)\) is known as the prudence coefficient. Note that for the CRRA case, where \( U(W) = W^{1-\gamma}/(1-\gamma) \), we have \(-U''(.)W/U'(.) = \gamma\) and \((1/2)U'''(.)W^2/U'(.) = (1/2)\gamma(\gamma + 1)\). So, if we consider just the first order approximation, we get

\[
E[r_t^e] - \gamma E[r_t^e (r_p (a^* - \mu)] = 0
\]

Noting that \( E[r_t^e (r_p (a^* - \mu)] = E[(r_t^e - \mu_t^e)(r_p (a^* - \mu)] \) and that \( r_p - \mu = \sum_{i=1}^n a_i (r_t^e - \mu_t^e) \), we get

\[
E[r_t^e] - \gamma \sum_{i=1}^n a_i^* E[(r_t^e - \mu_t^e)(r_j^e - \mu_j^e)] = 0
\]

that can be conveniently written in vector form

\[
\tilde{\mu} - \gamma \tilde{\Sigma} a^* = 0 \tag{73}
\]

where \( \tilde{\mu} \) is the vector of expected excess returns of the risk assets and \( \tilde{\Sigma} \) is its covariance matrix.

By observing equation (3), we can see that this is the FOC of the quadratic problem

\[
\max_a \{a' \tilde{\mu} - \frac{\gamma}{2} d' \tilde{\Sigma} a\} \tag{74}
\]

and its solution is given by \( a^* = \gamma^{-1} \tilde{\Sigma}^{-1} \tilde{\mu} \).

Even in this simple framework that considers only the first 2 moments of the excess return distribution it is important to note that they are unknown parameters for the portfolio manager. Since the parameters are unknown, they must be estimated from the data available at the investment date. The most simple way to estimate these parameters and make this formula useful for a practical portfolio decision problem is to compute the sample mean and the sample covariance matrix of excess returns \( \hat{\mu} \) and \( \hat{\Sigma} \).
In summary, the feasible problem of an investor is

$$\max_a a' \hat{\mu} - \frac{\gamma}{2} a' \hat{\Sigma} a$$  \hspace{1cm} (75)$$

where \( \hat{\mu} \) and \( \hat{\Sigma} \) must be estimated from the data. The simplest choice is to use the sample mean and the sample covariance of excess returns given by

$$\hat{\mu}_i = \frac{\sum_{t=1}^{T} r^e_{it}}{T}$$  \hspace{1cm} (76)$$

$$\hat{\Sigma}_{ij} = \frac{\sum_{t=1}^{T} (r^e_{it} - \hat{\mu}_i)(r^e_{jt} - \hat{\mu}_j)}{T - 1}$$  \hspace{1cm} (77)$$

One problem with this approach pointed out by Black and Litterman is that the portfolio choice is very sensitive to the mean estimate. That is, it translates into very different portfolios depending on the sample used. Another problem is that if we are concerned about estimation risk as is going to be defined in the next section, the sample mean and sample variance should not be the best estimators to be used.

4.2.1 Problems of the Basic Model

In a mean-variance framework the objective function of an investor is given by the utility function

$$U(r^e) = E[u(a'\tilde{r})] = E[a' \tilde{r} - \frac{\gamma}{2} a' \tilde{\Sigma} a] = a' \hat{\mu} - \frac{\gamma}{2} a' \hat{\Sigma} a$$

where \( r^e = r_p - \mu \) is the portfolio excess return and \( \tilde{r} \) is the random vector of assets excess returns with respect to the risk free rate.

If the true parameters \( \theta = (\hat{\mu}, \hat{\Sigma}) \) were known, then the choice \( a^* = a(\theta) = \gamma^{-1} \hat{\Sigma}^{-1} \hat{\mu} \) would be optimal by construction, and

$$F(\theta, \theta) = a(\theta)' \hat{\mu} - \frac{\gamma}{2} a(\theta)' \hat{\Sigma} a(\theta) = F_{max}$$

Since the parameters are unknown, \( \theta \) must be estimated from the data, \( \hat{\theta}(z) = (\hat{\mu}(z), \hat{\Sigma}(z)) \), which implies a decision rule \( \hat{a} = a(\hat{\theta}) = \gamma^{-1} \hat{\Sigma}^{-1} \hat{\mu} \) that necessarily leads to a lower utility level, that is,

$$F(\theta, \hat{\theta}(z)) = a(\hat{\theta})' \hat{\mu} - \frac{\gamma}{2} a(\hat{\theta})' \hat{\Sigma} a(\hat{\theta}) \leq F_{max}$$

This loss of utility due to parameter uncertainty is known as estimation risk. An estimation loss function is a non-negative function \( L(\theta, \hat{\theta}) \) that achieves its minimum at \( \hat{\theta} = \theta \).
If we define
\[ L(\theta, \hat{\theta}) = F(\theta, \theta) - F(\theta, \hat{\theta}) \]
Since the optimal rule for the problem is by construction \( a(\theta) \), its clear that \( L(\theta, \hat{\theta}) \geq 0 \) for \( \hat{\theta} \neq \theta \) and \( L(\theta, \hat{\theta}) = 0 \) if \( \hat{\theta} = \theta \).

The risk function for an estimator \( \hat{\theta}(z) \) is defined as
\[ R(\theta, \hat{\theta}) = \int L(\theta, \hat{\theta}(z)) f(z|\theta) dz \]
where \( f(z|\theta) \) is the likelihood function of the sample \( z \). It is clear that for a given estimator \( \hat{\theta} \), the risk function is a function of the parameter \( \theta \in \Theta \).

An estimator \( \hat{\theta} \) is said to be inadmissible if there exist another estimator \( \tilde{\theta} \) with at least equal and sometimes lower risk for any possible value of the true unknown parameter \( \theta \). That is, \( \tilde{\theta} \) is inadmissible is exists some \( \tilde{\theta} \) with
\[ R(\theta, \hat{\theta}) \leq R(\theta, \tilde{\theta}) \quad \forall \theta \in \Theta \]
With the inequality strict for some \( \theta \).

If we treat \( \theta \) as random, with an a priori density \( p(\theta) \), then the average loss from use of an estimator \( \hat{\theta} \) is
\[ r(p, \hat{\theta}) = E_p[R(\theta, \hat{\theta})] = \int \left( \int L(\theta, \hat{\theta}(z)) f(z|\theta) dz \right) p(\theta) d\theta \]
\[ = \int \left( \int L(\theta, \hat{\theta}(z)) g(\theta|z) d\theta \right) f(z) dz \]
where
\[ f(z) = \int f(z|\theta)p(\theta)d\theta \]
and
\[ g(\theta|z) = \frac{f(z|\theta)p(\theta)}{f(z)} \]
that is, the marginal density of \( z \) and the a posteriori density of \( \theta \) given \( z \).

Given a priori density \( p \), the estimator \( \hat{\theta} \) that minimizes \( r(p, \hat{\theta}) \) is the Bayes estimator, and the resulting minimum is the Bayes risk. The estimator that minimizes \( r(p, \hat{\theta}) \) is one that for each \( z \) minimizes the expression in parenthesis, that is, the expectation of \( L(\theta, \hat{\theta}(z)) \) with respect to the a posteriori distribution. If, for instance, \( \theta \) and \( \hat{\theta}(z) \) for a given \( z \) are
vectors and \( L(\theta, \hat{\theta}(z)) = (\theta - \hat{\theta})' Q (\theta - \hat{\theta}) \), where \( Q \) is positive definite, then

\[
E_{\theta \mid z}[L(\theta, \hat{\theta}(z))] = E_{\theta \mid z}[(\theta - E[\theta \mid z])' Q (\theta - E[\theta \mid z])] \\
+ (E[\theta \mid z] - \hat{\theta}(z))' Q (E[\theta \mid z] - \hat{\theta}(z))
\]

which is minimized at \( \hat{\theta}(z) = E[\theta \mid z] \), the mean of the a posteriori distribution.

Going back to our particular case we have \( a^* = a(\theta) = \gamma^{-1} \hat{\Sigma}^{-1} \hat{\mu} \), which yields a resulting utility of

\[
F(\theta, \theta) = a(\theta)' \hat{\mu} - \frac{\gamma}{2} a(\theta)' \hat{\Sigma} a(\theta) = \frac{1}{2\gamma} \hat{\mu} \hat{\Sigma}^{-1} \hat{\mu} \\
= \frac{S^2}{2\gamma}
\]

where \( s^2 \) is the squared Sharpe ratio of the ex ante tangency portfolio of the risky assets. We have seen that given \( \theta \), the risk function is given by

\[
R(\theta, \hat{\theta}(z)) = E_{z \mid \theta}[L(\theta, \hat{\theta}(z))] = F(\theta, \theta) - E_{z \mid \theta}[F(\theta, \hat{\theta}(z))]
\]

Considering \( z = \{\tilde{r}_1, ..., \tilde{r}_T\} \) is the matrix of \( T \) observations of vector excess returns and \( \hat{\theta} = (\hat{\mu}, \hat{\Sigma}) \) is the sample mean and sample covariance computed as in (6) and (7). Following Kan and Zoh, it is interesting to consider the case when \( \tilde{r}_t \) are independent and normally distributed random vectors each of which has common mean and variance \( \hat{\mu} \) and \( \hat{\Sigma} \). Under these assumptions, it is well known that \( \hat{\mu} \) and \( \hat{\Sigma} \) are independent of each other and have the following exact distributions,

\[
\hat{\mu} \sim N(\hat{\mu}, \hat{\Sigma}), \\
T \hat{\Sigma} \sim W(T - 1, \Sigma),
\]

where \( W(T - 1, \Sigma) \) denotes a Wishart distribution with \( T - 1 \) degrees of freedom and covariance matrix \( \Sigma \). Since \( E[\hat{\Sigma}] = T \Sigma^{-1} / (T - N - 2) \), we have

\[
E[\hat{\mu}] = E[\gamma^{-1} \hat{\Sigma}^{-1} \hat{\mu}] = \frac{T}{T - N - 2} a(\theta)
\]

(78)
When $\Sigma$ is known, $\hat{a} = \Sigma^{-1} \mu / \gamma$, and since $T\hat{\mu}'\Sigma^{-1}\hat{\mu} \sim \chi^2(T\mu'\Sigma\mu)$, we have

$$E_{z|\theta}[F(\theta, \hat{\theta}(z))] = E[\hat{a}']\mu - \frac{\gamma}{2} E[\hat{a}'\Sigma\hat{a}]$$

$$= \frac{\mu'\Sigma\mu}{\gamma} - \frac{1}{2\gamma} E[\hat{\mu}'\Sigma^{-1}\hat{\mu}]$$

$$= \frac{\mu'\Sigma\mu}{\gamma} - \frac{1}{2\gamma} \left( \frac{N + T\mu'\Sigma\mu}{T} \right)$$

$$= \frac{S^2}{2\gamma} - \frac{N}{2\gamma T}$$

As a result, the risk function from using $\hat{\theta}$ instead of $\theta$ is

$$R(\theta, \hat{\theta}(z)) = F(\theta, \theta) - E_{z|\theta}[F(\theta, \hat{\theta}(z))] = \frac{N}{2\gamma T}$$  \hspace{1cm} (79)$$

The result is intuitive. As large is the sample size $T$, more is learned about the true parameter $\theta$ and the estimation loss is reduced. In the limiting case where $T \to \infty$ the estimator $\hat{\theta} \to \theta$ and the loss from estimation risk goes to zero. On the other hand, as big is the size $N$ of the vector of unknown parameters, more uncertainty is present, since more parameters must be estimated, bigger is the loss. Finally, the bigger $\gamma$, more risk averse is the investor, so he invest less in the risky assets and the loss due to uncertainty is smaller.

To consider the more general case where $\mu$ and $\Sigma$ are unknown, we must use 2 known results about the Wishart distribution. Let $W = \Sigma^{-1/2} \Sigma \Sigma^{-1} \sim W_N(T - 1, I_N)/T$. The inverse moments of $W$ are

$$E[W^{-1}] = \left( \frac{T}{T - N - 2} \right) I_N$$

$$E[W^{-2}] = \left[ \frac{T^2(T - 2)}{(T - N - 1)(T - N - 2)(T - N - 4)} \right] I_N$$

With these results and the fact that $\hat{\mu}$ and $\hat{\Sigma}$ are independent, we can easily compute the expected out-of-sample performance

$$E_{z|\theta}[F(\theta, \hat{\theta}(z))] = \frac{1}{\gamma} E[\hat{\mu}'\hat{\Sigma}^{-1}\mu] - \frac{1}{2\gamma} E[\hat{\mu}'\hat{\Sigma}^{-1}\Sigma \hat{\Sigma}^{-1}\hat{\mu}]$$

$$= \frac{1}{\gamma} E[\hat{\mu}'\hat{\Sigma}^{-1/2}W^{-1}\Sigma^{-1/2}\mu] - \frac{1}{2\gamma} E[\hat{\mu}'\hat{\Sigma}^{-1/2}W^{-2}\Sigma^{-1/2}\hat{\mu}]$$
Now, it is easy to see that

$$E[\hat{\mu}'\Sigma^{-1/2}W^{-1}\Sigma^{-1/2}\mu] = \left(\frac{T}{T-N-2}\right)S^2$$

(80)

and with a little extra work, we get

$$E[\hat{\mu}'\Sigma^{-1/2}W^{-2}\Sigma^{-1/2}\hat{\mu}] = E[\text{tr}(\Sigma^{-1/2}\hat{\mu}\hat{\mu}'\Sigma^{-1/2}W^{-2})]$$

$$= \text{tr}(E[\Sigma^{-1/2}\hat{\mu}\hat{\mu}'\Sigma^{-1/2}W^{-2}])$$

$$= \left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\right]\text{tr}(E[\Sigma^{-1/2}\hat{\mu}\hat{\mu}'\Sigma^{-1/2}W^{-2}])$$

$$= \left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\right]E[\hat{\mu}'\Sigma^{-1}\hat{\mu}]$$

$$= \left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\right]\left(\frac{N + T\mu'\Sigma^{-1}\mu}{T}\right)$$

(81)

combining (10) with (11), we have

$$E_z[F(\theta, \hat{\theta}(z))] = k_1 S^2 - \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)}$$

(82)

where

$$k_1 = \left(\frac{T}{T-N-2}\right)\left[2 - \frac{T(T-2)}{(T-N-1)(T-N-4)}\right]$$

Hence, the out-of-sample performance of the MLE for $\theta$ when $\mu$ and $\Sigma$ are unknown is

$$R(\theta, \hat{\theta}(z)) = F(\theta, \theta) - E_z[F(\theta, \hat{\theta}(z))]$$

$$= (1 - k_1)\frac{S^2}{2\gamma} + \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)}$$

(83)

This is a closed form solution that relates the expected loss with respect to $N, T, \gamma$, and $S^2$. The intuition remains, as $N$, or $S^2$ increases, the loss increases. On the other hand, as $T$ or $\gamma$ increases, the loss decreases.

Next we are going to show that the sample covariance mean and Covariance is inadmissible in the sense that there exist another estimator that provides a smaller loss, regardless of the values of the true parameters.

4.2.2 The Bayesian Solution

The Bayesian approach assumes that the investor cares about the expected utility under the predictive distribution $p(r_{t+1}|z)$, which is determined by the historical data and the prior. Brown (1976), Klein and Bawa (1976), and Stambaugh (1997) show under the
diffuse prior on $\mu$ and $\Sigma$, 

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{N + 1}{2}}$$

The optimal portfolio weights computed by the predictive moments in this case is

$$\tilde{\alpha} = \frac{1}{\gamma} \left( \frac{T - N - 2}{T + 1} \right) \hat{\Sigma}^{-1} \hat{\mu} = \left( \frac{T - N - 2}{T + 1} \right) \hat{\mu}$$

The question is if this approach actually improves out of sample performance in terms of reduced loss.

$$E_{\theta | \theta}[F(\theta, \tilde{\theta}(z))] = \left( \frac{T - N - 2}{T + 1} \right) \frac{1}{\gamma} E[\hat{\mu}' \hat{\Sigma}^{-1} \mu] - \left( \frac{T - N - 2}{T + 1} \right)^2 \frac{1}{2\gamma} E[\hat{\mu}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\mu}]$$

and using (10) and (11) we get

$$E_{\theta | \theta}[F(\theta, \tilde{\theta}(z))] = k_2 \frac{S^2}{2\gamma} - \frac{NT(T - 2)(T - N - 2)}{2\gamma(T + 1)^2(T - N - 1)(T - N - 4)} \quad (84)$$

where $T > N + 4$ and

$$k_2 = \left( \frac{T}{T + 1} \right) \left[ 2 - \frac{T(T - 2)(T - N - 2)}{(T + 1)(T - N - 1)(T - N - 4)} \right]$$

The resulting Bayesian risk in this case is

$$R(\theta, \tilde{\theta}(z)) = F(\theta, \theta) - E_{\theta | \theta}[F(\theta, \tilde{\theta}(z))]$$

$$= (1 - k_2) \frac{S^2}{2\gamma} + \frac{NT(T - 2)(T - N - 2)}{2\gamma(T + 1)^2(T - N - 1)(T - N - 4)} \quad (85)$$

Now we can compare the risk of the sample mean and covariance $\tilde{\theta}$ with the risk of the Bayes estimator $\hat{\theta}$. We have,

$$R(\theta, \hat{\theta}(z)) - R(\theta, \tilde{\theta}(z)) = (k_2 - k_1) \frac{S^2}{2\gamma} + \frac{NT(T - 2)(2T - N - 1)(N + 3)}{2\gamma(T + 1)^2(T - N - 1)(T - N - 2)(T - N - 4)}$$

It is easy to see that

$$k_2 - k_1 = \frac{T^2(T - 2)(2T - N - 1)(N + 3)}{(T + 1)^2(T - N - 1)(T - N - 2)(T - N - 4)} - \frac{2}{(T + 1)^2(T - N - 1)(T - N - 2)(T - N - 4)} > 0$$

86
because $T^2 > 2(T + 1)$ for any $T > 2$. Thus the Bayesian portfolio rule always strictly outperforms the sample mean and sample covariance by yielding lower expected loss regardless of the value of true parameters $\theta$. That means that $\hat{\theta}$ is inadmissible as defined before.

### 4.3 Kan-Zhou Optimal Two-Fund Rule

Kan and Zhou (2007) consider the class of Two-funded portfolio rules that have weights

$$a(\hat{\theta}) = \frac{c}{\gamma} \hat{\Sigma}^{-1} \hat{\mu}$$

where $c$ is a constant scalar. Clearly, the previous rules are particular cases of this class. In the first case we have $c_1 = 1$ and in the second case $c_2 = (T - N - 2)/(T + 1)$. This rule can be viewed as a plug-in estimator that estimates $\Sigma$ by using $\hat{\Sigma}^* = \hat{\Sigma}/c$.

It is easy to compute the expected out of sample performance of this class of portfolio rules. Let $\hat{\theta}^* = (\hat{\mu}, \hat{\Sigma}^*)$. Using equations (10) and (11), we have

$$E_{z|\theta}[F(\theta, \hat{\theta}^*(z))] = \frac{cS^2}{\gamma} \left( \frac{T}{T - N - 2} \right) - \frac{c^2}{2\gamma} \left( S^2 + \frac{N}{T} \right) \left[ \frac{T^2(T - 2)}{(T - N - 1)(T - N - 2)(T - N - 4)} \right]$$

differentiating with respect to $c$, the optimal $c$ is

$$c^* = \left[ \frac{(T - N - 1)(T - N - 4)}{T(T - 2)} \right] \left( \frac{S^2}{S^2 + \frac{N}{T}} \right)$$

Although $c^*$ is optimal, it is not a feasible strategy since $\theta$ is unknown in practice. Nevertheless, the second factor is close to 1 for high values of $S^2$ and or high values of $T$. It suggests taking the suboptimal rule

$$\hat{a}^* = \frac{c_3}{\gamma} \hat{\Sigma}^{-1} \hat{\mu},$$

where

$$c_3 = \frac{(T - N - 1)(T - N - 4)}{T(T - 2)}$$

This rule is parameter independent and suggests investing $\hat{a}^*$ in the risky assets ans $1 - 1_N' \hat{a}^*$ in the risk free asset. It is not difficult to show that this strategy dominates over the Bayesian strategy. In fact, since $f(c) = E_{z|\theta}[F(\theta, \hat{\theta}^*(z))]$ is a quadratic function of $c$, the expected out of sample performance is a decreasing function of $c$ for $c \geq c^*$. So we
have the result by checking that $c_2 > c_3 > c^*$. To see this, just note that

$$c_2 = \frac{T - N - 1}{T + 1} > \frac{T - N - 4}{T}$$

$$> \left( \frac{T - N - 4}{T} \right) \left( \frac{T - N - 1}{T - 2} \right) = c_3$$

$$> \left[ \frac{(T - N - 1)(T - N - 4)}{T(T - 2)} \right] \left( \frac{S^2}{S^2 + \frac{N}{T}} \right) = c^*$$

4.3.1 The Three-Fund Rule and Mixed Portfolio Strategies

Kan and Zhou propose a mixed portfolio strategy to reduce estimation risk. The idea is that since if two portfolio have estimation errors that are not perfect correlated, estimation risk can be diversified. As the alternative risk portfolio they propose using the global minimum variance portfolio since it depend only on $\hat{\Sigma}$ but not on $\hat{\mu}$. The idea is to consider the portfolio rule:

$$\hat{a} = \hat{a}(c, d) = \frac{1}{\gamma} (c \hat{\Sigma}^{-1} \hat{\mu} + d \hat{\Sigma}^{-1} 1_N) \quad (89)$$

and find $c$ and $d$ that minimizes estimation risk. We show in the appendix that this strategy suggests choosing

$$c^* = c_3 \left( \frac{\psi^2}{\psi^2 + \frac{N}{T}} \right)$$

$$d^* = c_3 \left( \frac{\frac{N}{T}}{\psi^2 + \frac{N}{T}} \right)$$

where

$$\psi^2 = \mu' \Sigma^{-1} \mu - \left( \frac{\mu' \Sigma^{-1} 1_N}{1_N' \Sigma^{-1} 1_N} \right)^2 = (\mu - \mu_g 1_N)' \Sigma^{-1} (\mu - \mu_g 1_N)$$

and

$$\mu_g = \frac{1_N' \Sigma^{-1} \mu}{1_N' \Sigma^{-1} 1_N}$$

is the expected excess return of the ex-ante global minimum-variance portfolio. The problem with this strategy is that the values of $c$ and $d$ in this case depend on population parameters that are unknown. So we are going to analyze the a posteriori naive strategy of $c = d = 1/2$. 88
4.4 The Bayesian Framework

4.4.1 The Bayesian Prior-Posterior Approach to Model Uncertainty

Some important concepts of Bayesian statistics are the definitions of the prior, posterior and the predictive distributions. The idea is that the investor may consider a number of possible return distributions. An investor who ignores the uncertainty in the model parameters uses the distribution of future returns conditional on both past data and fixed parameter values \( f(r_{t+T}|\theta, z) \), where \( z = (z_1, ..., z_t)’ \). In contrast, the investor who takes parameter uncertainty into account samples from the predictive distribution, conditional only on past data and not on the parameters, \( f(r_{t+T}|z) \). To be more precise, suppose that the initial prior pdf for a parameter vector \( \theta \) is \( p(\theta) \) and the investor observes a set of data \( z \) with pdf \( p(z|\theta) \). Let \( p(z, \theta) \) the joint probability density function for a random observation vector \( z \) and a parameter vector \( \theta \), also considered random. Since

\[
p(z, \theta) = p(z|\theta)p(\theta) = p(\theta|z)p(z)
\]

we have that

\[
p(\theta|z) \propto p(\theta)p(z|\theta)
\]

The term \( p(\theta) \) is the prior distribution, \( p(\theta|z) \) is the posterior pdf for the parameter vector \( \theta \), and \( p(z|\theta) \), viewed as a function of \( \theta \), is the likelihood function. As pointed out by Zellner (1996), the joint posterior p.d.f. \( p(\theta|z) \), has all the prior and sample information incorporated in it.

An important prior to be considered in some cases is the prior of ignorance, also known, as an uninformative prior in Bayesian statistics. Jeffreys’ prescription for representing ignorance about a value of \( \theta \), which can assume values from \(-\infty\) to \(+\infty\), is to take

\[
p(\theta)d\theta \propto d\theta
\]

When considering parameters, like the standard deviation \( \sigma \), which by their nature, can assume values from 0 to \(+\infty\), Jeffreys suggests taking its logarithm uniform; that is, considering \( \theta = \log \sigma \), where \( \theta \) takes values from \(-\infty\) to \(+\infty\). Since \( d\theta = d\sigma/\sigma \), we have

\[
p(\sigma)d\sigma \propto \frac{d\sigma}{\sigma}, \quad 0 < \sigma < +\infty
\]

A more detail explanation about Jeffreys’rule will be given in the following section. Now, suppose that, given our sample information \( z \), we are interest in making inferences about other observations that are still unobserved. Let \( \tilde{z} \) represent a vector of yet unobserved observations. We have

\[
p(\tilde{z}, \theta|z) = p(\tilde{z}|\theta, z)p(\theta|z)
\]
Note that on the right we have the conditional pdf for $\tilde{z}$, given $\theta$ and $z$, whereas $p(\theta|z)$ is the posterior pdf for $\theta$. To obtain the predictive pdf, $p(\tilde{z}|z)$, we merely integrate the above equation with respect to $\theta$; that is

\begin{equation}
    p(\tilde{z}|z) = \int p(\tilde{z}, \theta|z)d\theta
\end{equation}

\begin{equation}
    = \int p(\tilde{z}|\theta, z)p(\theta|z)d\theta
\end{equation}

To contextualize this framework, we analyze here how much a portfolio manager decides to invest in each of the available risky assets. Let $U(R)$ be the utility function, where $R$ is the return distribution from the investment, and $g(R|\theta)$ the conditional density of asset returns given the set of parameters $\theta$. If $\theta$ is known, the conditional expected utility of the investor is

$$
\mathbb{E}[U(R)|\theta] = \int U(R)g(R|\theta)dR
$$

In practice, however, $\theta$ is unknown and needs to be estimated from data. In the presence of parameter uncertainty, we need to infer the posterior density, $p(\theta|z)$, from the data, where $z = (r_1, ..., r_T)$ is the vector of past returns. The expected utility is then given by

$$
\mathbb{E}[U(R|z)] = \mathbb{E}[\mathbb{E}[U(R)|\theta]|z] = \int \int U(R)g(R|\theta)p(\theta|z)d\theta dR
$$

Let $p(\theta)$ be the unconditional prior about the unknown parameter. Then the posterior density given $z$ is

$$
p(\theta|z) = \frac{g(z|\theta)p(\theta)}{p(z)} = \frac{g(z|\theta)p(\theta)}{\int g(z|\theta)p(\theta)d\theta}
$$

(100)

If the returns are i.i.d., then

$$
p(\theta|z) = \frac{\prod_{t=1}^{T} g(r_t|\theta)p(\theta)}{\int g(z|\theta)p(\theta)d\theta}
$$

(101)

and the predictive density, given $z$, is

$$
g(R|z) = \int g(R|\theta)p(\theta|z)d\theta = \int g(R|\theta) \left( \frac{\prod_{t=1}^{T} g(r_t|\theta)p(\theta)}{\int g(z|\theta)p(\theta)d\theta} \right) d\theta
$$

(102)

Using the predictive density, the expected utility of the investor is given by

$$
\mathbb{E}[U(R|z)] = \int U(R)g(R|z)dR
$$

(103)
For instance, if we consider a quadratic utility in the investments returns, that is, if $U(R) = R - \frac{\gamma}{2} R^2$, then using the predictive density, the expected utility is

$$\mathbb{E}[U(R|z)] = E[R|z] - \frac{\gamma}{2} E[R^2|z]$$

That makes clear that in a Bayesian framework with unknown return distribution a straightforward generalization of Markowitz original problem urges us to compute the first and second predictive moments of the assets returns which are, respectively

$$E[R|z] = \int R g(R|z) dR,$$
$$E[R^2|z] = \int R^2 g(R|z) dR$$

We follow Barberis (2000) and analyze 2 distinct portfolio problems: a static buy-and-hold problem and a dynamic problem with optimal rebalancing. We define regular intervals for rebalancing the portfolio. We treat here uncertainty about the parameters, known as estimation risk, using a Bayesian approach. To understand the Bayesian approach it is useful to think that we can solve the portfolio in 2 different ways. The simple way is to construct the distribution of future returns conditional on fixed parameter estimates. The second is to consider the parameters as functions of the Data and integrate the parameters over the posterior distribution. This allows for the construction of the predictive distribution for future returns, conditional only on observed data. Comparing the results obtained from considering the parameters fixed with the one obtained with the predictive distribution is the way considered in the literature to evaluate the effect of parameter uncertainty. Barberis analyzes the separate effects of model uncertainty and predictability by considering uncertainty in an i.i.d. case first. Secondly he analyzes the effect of buy and hold by comparing it with an optimal rebalancing problem. It is important to point out here that our analyzes differ from Barberis in the sense that He considers only the problem where the investor must decide how much to invest in the risk free and how much to invest in the risky asset.

### 4.4.2 Jeffreys’ Rule for Prior Ignorance

Since the prior distribution of $\theta$, $p(\theta)$ plays an important role in Bayesian models, we are going to give here a brief discussion based on Box and Tiao (2011).

Let $y = (y_1, ..., y_n)$ be a random sample from a distribution $p(y|\theta)$. With certain regularity conditions, for sufficiently large $n$, the likelihood function of $\theta$ is approximately Normal, and remains Normal under mild one-to-one transformations of $\theta$. In such case, the log-
likelihood is quadratic, so that

\[ L(\theta|y) = L(\hat{\theta}|y) - \frac{n}{2}(\theta - \hat{\theta})^2 \left( -\frac{1}{n} \frac{\partial^2 L}{\partial \theta^2}(\hat{\theta}) \right) \]

Where \( \hat{\theta} \) is the Maximum Likelihood Estimator of \( \theta \). In general, the quantity

\[ J(\hat{\theta}) = -\frac{1}{n} \frac{\partial^2 L}{\partial \theta^2}(\hat{\theta}) \]

is a function of all the data \( y \). But it is interesting to note that for a large \( n \), \( \hat{\theta} \) converges to \( \theta_0 \) and the average of the above function converges to the true mean of the function

\[ E \left[ -\frac{\partial^2 L}{\partial \theta^2}(\theta_0) \right] = -\int \frac{\partial^2 L}{\partial \theta^2}(\theta_0) p(y|\theta_0) dy \]

Then we have \( J(\hat{\theta}) \) is approximately \( \mathcal{F}(\hat{\theta}) \), where \( \mathcal{F}(\theta) \) is the function

\[ \mathcal{F}(\theta) = -E \left[ \frac{\partial^2 L}{\partial \theta^2}(\theta) \right] = E \left[ \frac{\partial L}{\partial \theta}(\theta) \right]^2 \]

Consequently, we use \( \mathcal{F}(\hat{\theta}) \), that depends on \( \hat{\theta} \) only, to approximate \( J(\hat{\theta}) \). Now suppose \( \phi(\theta) \) is a one to one transformation. Then,

\[ J(\hat{\phi}) = \left( -\frac{1}{n} \frac{\partial^2 L}{\partial \phi^2}(\hat{\phi}) \right) = \left( -\frac{1}{n} \frac{\partial^2 L}{\partial \theta^2}(\hat{\theta}) \right) \left( \frac{d\theta}{d\phi} \right)^2 = J(\hat{\theta}) \left( \frac{d\theta}{d\phi} \right)^2 \]

or by using the approximation,

\[ \mathcal{F}(\hat{\phi}) = \mathcal{F}(\hat{\theta}) \left( \frac{d\theta}{d\phi} \right)^2 \]

It follows that if \( \phi(\hat{\theta}) \) is chosen such that

\[ \left| \frac{d\theta}{d\phi} \right| \propto \mathcal{F}^{-1/2}(\hat{\theta}) \]

Then \( J(\hat{\theta}) \) will be a constant independent of \( \hat{\phi} \), and the likelihood will be approximately data translated in terms of \( \phi \). That is, we find that the metric \( \phi(\theta) \) for which a locally uniform prior is approximately noninformative is such that

\[ \frac{d\phi}{d\theta} \propto \mathcal{F}^{1/2}(\theta) \]
Equivalently, the noninformative prior for \( \theta \) should be chosen so that, locally,

\[ p(\theta) \propto \mathcal{F}^{1/2}(\theta) \]

Now, for the multiple parameters case, if the distribution of \( y \), depending on \( k \) parameters \( \theta \), obeys certain regularity conditions, then, for sufficiently large samples, the likelihood function for \( \theta \) and for mild transformations of \( \theta \) approaches a Multivariate Normal distribution. The log likelihood is thus quadratic,

\[ L(\theta|y) = \log l(\theta|y) = L(\hat{\theta}|y) - \frac{n}{2}(\theta - \hat{\theta})'D_\theta(\theta - \hat{\theta}), \]

where \( \hat{\theta} \) is the vector of Maximum Likelihood Estimates of \( \theta \) and \( -nD_\theta \) is a \( k \times K \) matrix of second derivatives of the parameters evaluated at \( \hat{\theta} \),

\[
D_\theta = \left\{ -\frac{1}{n} \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} (\hat{\theta}) \right\}, \quad i, j = 1, ..., k
\]

which, for a large \( n \), can be approximated by \( \mathcal{F}_n(\hat{\theta}) \) which is a function of \( \hat{\theta} \) only, where

\[
\mathcal{F}_n(\theta) = E \left[ \frac{\partial^2 L}{\partial \theta_i \partial \theta_j}(\theta) \right]
\]

where the expectation is taken with respect to the data distribution \( p(y|\theta) \). That is, \( \mathcal{F}_n(\theta) \) is the information matrix associated with the sample \( y \).

Now, the idea is to find a transformation \( \phi \) which ensures that the content of the approximate likelihood region of \( \phi \),

\[ (\phi - \hat{\phi})'\mathcal{F}_n(\hat{\phi})(\phi - \hat{\phi}) < \text{const.} \]

remains constant for different \( \hat{\phi} \). This is equivalent to asking for a transformation for which the \( |\mathcal{F}_n(\hat{\phi})| \) is independent of \( \hat{\phi} \). Since

\[
\mathcal{F}_n(\phi) = A\mathcal{F}_n(\theta)A'
\]

where \( A \) is the \( k \times k \) matrix of partial derivatives of \( \theta \) with respect to \( \phi \). Thus,

\[
|\mathcal{F}_n(\phi)| = |A|^2|\mathcal{F}_n(\theta)|
\]

Then, the above requirement is fulfilled if \( \phi \) is such that

\[
|A| \propto |\mathcal{F}_n(\theta)|^{-1/2}
\]

93
The corresponding noninformative prior in $\theta$ is then

$$p(\theta) = p(\phi)|A|$$

that is,

$$p(\theta) \propto |\mathcal{F}_n(\theta)|^{1/2}$$

So, we can conclude the Jeffreys’ rule for multiparameter problems is to take the prior distribution for a set of parameters to be proportional to the square root of the determinant of the information matrix.

To illustrate the argument, we are going to show that the determinant of the information matrix for $\Sigma$ and for $\Sigma^{-1}$ in the multivariate regression model are respectively proportional to $|\Sigma|^{-(m+1)}$ and $|\Sigma|^{m+1}$. To see this first we have that the $m$-dimensional Normal distribution has density

$$p(y|\mu, \Sigma) = (2\pi)^{-m/2}|\Sigma^{-1}|^{1/2}\exp\left[-\frac{1}{2}tr\Sigma^{-1}(y-\mu)(y-\mu)\right]$$

and its corresponding log likelihood is given by

$$L(\mu, \Sigma) = -\frac{m}{2} \log 2\pi + \frac{1}{2} \log |\Sigma^{-1}| - \frac{1}{2} tr\Sigma^{-1}(y-\mu)(y-\mu)'$$

Differentiating with respect to $\sigma^{ij}$ we get

$$\frac{\partial L}{\partial \sigma^{ij}} = \frac{1}{2} \frac{\partial |\Sigma^{-1}|}{\partial \sigma^{ij}} - (y_i - \mu_i)(y_j - \mu_j)$$

Since $\partial |\Sigma^{-1}|/\partial \sigma^{ij}$ is just the cofactor of $\sigma^{ij}$ (see Appendix), the first term on the right hand side is $(1/2)\sigma_{ij}$. Thus, the second derivatives are

$$\frac{\partial^2 L}{\partial \sigma^{ij}\partial \sigma^{kl}} = \frac{1}{2} \frac{\partial \sigma_{ij}}{\partial \sigma^{kl}}$$

so the determinant of the information matrix is proportional to

$$|\mathcal{F}(\Sigma^{-1})| = \left| E \left[ \frac{\partial^2 L}{\partial \sigma^{ij}\partial \sigma^{kl}} \right] \right| \propto \left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right|$$

since we know that (see appendix)

$$\left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right| = |\Sigma|^{m+1}$$
we have the second result. To get the first, we only need to remind that

\[ |\mathcal{F}(\Sigma)| = |\mathcal{F}(\Sigma^{-1})| \left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right|^{-2} \]  

(111)

We conclude that the uninformative prior for a covariance matrix is

\[ p(\Sigma) \propto |\mathcal{F}(\Sigma)|^{1/2} \propto |\Sigma|^{-(m+1)/2} \]  

(112)

### 4.4.3 A Simple Bayesian Portfolio Model

In a simple framework, the problem facing a Bayesian investor is to estimate the \( N \)-dimensional vector of means \( \mu \) from i.i.d. population \( y_t \sim N(\mu, \Sigma), t = 1, ..., T \). The key result in Jorion (1986) can be summarized as follows. Assume the following 3 conditions:

(i) Investors have an informative prior on \( \mu \) of the form

\[
p(\mu|\bar{\mu}, \nu_\mu) \propto \exp \left[ -\frac{1}{2} (\mu - \bar{\mu}1_N)^T (\nu_\mu \Sigma^{-1}) (\mu - \bar{\mu}1_N) \right]
\]  

(113)

with \( \bar{\mu} \) being the grand mean and \( \nu_\mu \) giving an indication of prior precision; (ii) Investors have diffuse prior on the grand mean \( \bar{\mu} \); (iii) The density \( p(\nu_\mu|\mu, \bar{\mu}, \Sigma) \) is a Gamma function with mean at \((N + 2)/d\) where \( d \) is defined as

\[
d = (\mu - \bar{\mu}1_N)^T \Sigma^{-1} (\mu - \bar{\mu}1_N)
\]

and is replaced by its sample estimate

\[
(\hat{\mu} - \mu_{MIN}1_N)^T \Sigma^{-1} (\hat{\mu} - \mu_{MIN}1_N)
\]

Where

\[
\mu_{MIN} = \frac{1_N^T \Sigma^{-1} \hat{\mu}}{1_N^T \Sigma^{-1} 1_N}
\]  

(114)

Then, the predictive density for the return distribution \( g(r|y, \Sigma, \nu_\mu) \), conditional on \( \Sigma \) and the precision \( \nu_\mu \) is a multivariate normal with predictive Bayes-Stein mean, \( \mu_{BS} \), equal to

\[
\mu_{BS} = (1 - \phi_{BS})\bar{\mu} + \phi_{BS}\mu_{MIN}1_N
\]

where \( \bar{\mu} \) is the sample mean, \( \mu_{MIN} \) is the minimum variance portfolio,

\[
\phi_{BS} = \left( \frac{\nu_\mu}{T + \nu_\mu} \right) = \frac{N + 2}{(N + 2) + T(\hat{\mu} - \mu_{MIN}1_N)^T \Sigma^{-1} (\hat{\mu} - \mu_{MIN}1_N)}
\]

95
and covariance matrix

\[ \text{Var}(r) = \Sigma \left( 1 + \frac{1}{T + \nu_{\mu}} \right) + \frac{\nu_{\mu}}{T(T + 1 + \nu_{\mu})} 1_N 1_N^T \Sigma^{-1} 1_N \]

The term \( \phi_{BS} \) is known as the shrinkage coefficient, since it shrinks the sample mean towards the mean of the minimum variance portfolio. It is easy to see that the case of zero precision, \( \nu_{\mu} = 0 \), corresponds to the Bayes diffuse prior case considered in Bawa, Brown, and Klein (1979) and Zellner and Chetty in which the sample mean is the predictive mean but the covariance matrix is inflated by the factor \((1 + 1/T)\). Finally, for \( \nu_{\mu} \to \infty \) the predictive mean is the mean of the minimum variance portfolio and the covariance matrix is given by \( \Sigma + (1/T)1_N 1_N^T (1/\Sigma^{-1} 1_N) \).

Substituting the Bayes Stein estimator \( \mu_{BS} \), it is easy to see that the optimal portfolio weights can be written as

\[ \omega_{BS}(\nu_{\mu}) = (1 - \phi_{BS}(\nu_{\mu})) \omega_{MV} + \phi_{BS}(\nu_{\mu}) \omega_{MIN} \]  \hspace{1cm} \text{(115)}

where the global minimum variance portfolio is given by

\[ \omega_{MIN} = \frac{1}{1_N \Sigma^{-1} 1_N} \Sigma^{-1} 1_N \]  \hspace{1cm} \text{(116)}

and the mean variance portfolio is given by

\[ \omega_{MV} = \frac{1}{\gamma} \Sigma^{-1} (\hat{\mu} - \hat{\mu}^0 1_N) \]

In practice, \( \Sigma \) is unknown, and is usually replaced by

\[ \hat{\Sigma}_{BS} = \frac{T - 1}{T - N - 2} \hat{\Sigma} \]  \hspace{1cm} \text{(117)}

where \( \hat{\Sigma} \) is the usual unbiased sample covariance matrix.

4.5 Incorporating Pricing Model Restrictions

Pastor and Stambaugh (2000) consider a framework where individuals can invest in assets that represent benchmark positions and non-benchmark assets. It suggests partition the return distribution \( r_t = (r_{1,t}, r_{2,t}) \) where \( r_{2,t} \) contains the payoffs on \( k \) benchmark positions and \( r_{1,t} \) the payoff of the remaining assets. They consider then a multivariate regression,

\[ r_{1,t} = \alpha + Br_{2,t} + u_t \]  \hspace{1cm} \text{(118)}
where $B$ is a $(n - 3) \times 3$ matrix where each of its rows is a vector of betas of the first $(n - 3)$ non-benchmark positions, $(\beta_{1,i}, \beta_{2,i}, \beta_{3,i})$, and

$$u_t \sim N(0, \Sigma) \quad (119)$$

This regression equation implies moments restrictions that the parameters must obey. In particular,

$$\alpha = E[r_{1,t}] - BE[r_{2,t}] \quad (120)$$

and

$$\Sigma = V_{11} - BV_{22}B' \quad (121)$$

The approach they used impose restrictions on $\alpha$ but no restrictions on $B$, $\Sigma$, $E_2$ and $V_{22}$. In Bayesian statistics that means to use uninformative prior distributions to this set of parameters. The prior distribution for $\Sigma$ is specified as inverted Wishart,

$$\Sigma^{-1} \sim W(H^{-1}, v) \quad (122)$$

with degrees of freedom $v = 15$. From the properties of the inverted Whishart distribution, the prior expectation of $\Sigma$ equals $H/(v - m - 1)$. We specify $H = s^2(v - m - 1)I_m$, so that $E[\Sigma] = s^2I_m$. The value of $s^2$ is set equal to the average of the diagonal elements of the sample estimate of $\Sigma$. The joint prior distribution for the remaining parameters $(B, E_2, V_{22})$ is assumed to be diffuse and independent of $\alpha$ and $\Sigma$. The density function for these parameters are

$$p(\Sigma) \propto |\Sigma|^{-(v+m+1)/2} \exp \left\{ -\frac{1}{2} trH \Sigma^{-1} \right\}$$

$$p(B) \propto 1$$

$$p(E_2) \propto 1$$

$$p(V_{22}) \propto |V_{22}|^{-k+1/2}$$

Following their approach, we consider as factor based the CAPM, in which $k = 1$, and the 3 factor FF model, in which $k = 3$. The pricing model impose $\alpha = 0$. To allow for mispricing uncertainty, they consider a prior distribution for $\alpha$ specified as a normal distribution,

$$\alpha|\Sigma \sim N(0, \sigma^2_{\alpha}(1/s^2\Sigma)) \quad (123)$$

so, its density function is represented by

$$p(\alpha|\Sigma) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} \alpha'(\frac{\sigma^2_{\alpha}}{s^2\Sigma}^{-1}) \alpha \right\}$$

Where $\sigma^2_{\alpha}$ reflects the investor’s prior degree of mispricing uncertainty. When $\sigma_{\alpha} = 0,$
the investor believes dogmatically in the model. When $\sigma_\alpha = \infty$, the investor regards the model as useless.

Letting $\theta = (\alpha, B, \Sigma, E_2, V_{22})$, the joint prior distribution of the parameters can be easily obtained by the factor

$$p(\theta) = p(\alpha|\Sigma)p(\Sigma)p(B)p(E_2)p(V_{22})$$

To understand how these benchmark models alter our original mean variance problem, it is useful to consider the moments partitioned as follows

$$E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, \quad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

(124)

The vector of betas for each of the non-benchmark positions is a row of the matrix

$$B = V_{12}V_{22}^{-1}$$

Also, the pricing restriction implies that $\alpha = 0$ or equivalently

$$E_1 = BE_2$$

(125)

Substituting in the portfolio optimization formula, we have

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{A} \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}^{-1} \begin{bmatrix} V_{12}V_{22}^{-1}E_2 \\ E_2 \end{bmatrix} = \frac{1}{A} \begin{bmatrix} 0 \\ V_{22}^{-1}E_2 \end{bmatrix}$$

(126)

Since from a partitioned inverse result due to Duncan (1944) we have

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (V_{11} - V_{12}V_{22}^{-1}V_{21})^{-1} - (V_{11} - V_{12}V_{22}^{-1}V_{21})^{-1}V_{12}V_{22}^{-1} \\ -(V_{22} - V_{21}V_{11}^{-1}V_{12})^{-1}V_{21}V_{11}^{-1} & (V_{22} - V_{21}V_{11}^{-1}V_{12})^{-1} \end{bmatrix}$$

That is, if the investor fully believes in the pricing model, $\sigma_\alpha = 0$, the optimal portfolio involves only the benchmark positions. Then, these would be the exact moments he should estimate from the data. Writing in terms of the predictive distribution, the optimal portfolio in this case is

$$w^* = (1/A)V_{22}^{-1}E_2^*$$

On the other hand, if the investor consider the pricing model useless, $\sigma_\alpha = \infty$. It would be sound to consider the case in between. In the next subsection we are going to show how to replace $E$ and $V$ by moments of the Bayesian predictive distributions corresponding to varying degrees of prior confidence in the pricing model.
4.5.1 Finding the Predictive Moments of the model

To find the predictive moments, we follow Pastor and Stambaugh and define $Y = (r_{1,1}, ..., r_{1,T})'$, $X = (r_{2,1}, ..., r_{2,T})'$, and $Z = (1_T, X)$, where $1_T$ denotes a $T$ vector of ones. Also define the $(k+1) \times m$ matrix $A = (\alpha, B)'$, and let $a = vec(A)$. For the $T$ observations $t = 1, ..., T$, the regression model can be written as

$$Y = ZA + U, \quad vec(U) \sim N(0, \Sigma \otimes I_T), \quad (127)$$

Where $U = (u_1, ..., u_T)'$. The matrix $R = (Y, X)$ contains the entire sample. Define the statistics $\hat{A} = (Z'Z)^{-1}Z'Y$, $\hat{a} = vec(\hat{A})$, $\hat{\Sigma} = (Y - Z\hat{A})(Y - Z\hat{A})/T$, $\hat{E}_2 = X'1_T/T$, and $\hat{V}_{22} = (X - 1_T\hat{E}_2')(X - 1_T\hat{E}_2')/T$.

We factor the likelihood function as

$$p(R|\theta) = p(Y, X|\theta) = p(Y|\theta, X)p(X|\theta), \quad (128)$$

where

$$p(Y|\theta, X) \propto |\Sigma|^{-T/2}\exp\left\{ -\frac{1}{2}tr(Y - ZA)'(Y - ZA)\Sigma^{-1} \right\}$$

$$\times |\Sigma|^{-T/2}\exp\left\{ -\frac{T}{2}tr\hat{\Sigma}\Sigma^{-1} - \frac{1}{2}tr(A - \hat{A})'Z'Z(A - \hat{A})\Sigma^{-1} \right\}$$

$$\times |\Sigma|^{-T/2}\exp\left\{ -\frac{1}{2}trT\hat{\Sigma}\Sigma^{-1} - \frac{1}{2}(a - \hat{a})'(\Sigma^{-1} \otimes Z'Z)(a - \hat{a}) \right\}$$

where the second line follows from the fact that $(Y - XB)'(Y - XB) = (Y - X\hat{B})'(Y - X\hat{B}) + (B - X\hat{B})'X'X(B - X\hat{B})$, where $\hat{B} = (X'X)^{-1}X'Y$ and the third line follows from the facts $trA'B = vec(A)'vec(B)$ and $vec(ABC) = (C' \otimes A)vec(B)$. And

$$p(X|\theta) \propto |V_{22}|^{-T/2}\exp\left\{ -\frac{1}{2}tr(X - 1_TE_2)'(X - 1_TE_2)V_{22}^{-1} \right\}$$

$$\times |V_{22}|^{-T/2}\exp\left\{ -\frac{T}{2}tr\hat{V}_{22}V_{22}^{-1} - \frac{T}{2}tr(E_2 - \hat{E}_2)(E_2 - \hat{E}_2)V_{22}^{-1} \right\}$$

As we have seen, the joint prior distribution of all parameters is

$$p(\theta) = p(\alpha|\Sigma)p(\Sigma)p(B)p(E_2)p(V_{22}) \quad (129)$$
where

\[
p(\alpha | \Sigma) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} \alpha' \left( \frac{\sigma^2}{s^2} \Sigma \right)^{-1} \alpha \right\}
\]
\[
p(\Sigma) \propto |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} tr H \Sigma^{-1} \right\}
\]
\[
p(B) \propto 1
\]
\[
p(E_2) \propto 1
\]
\[
p(V_{22}) \propto |V_{22}|^{-\frac{k+1}{2}}
\]

The priors of \(B, E_2,\) and \(V_{22}\) are diffuse. The prior of \(\Sigma\) is inverted Wishart with a small number of degrees of freedom which is uninformative. The prior on \(\alpha\) given \(\Sigma\) is normal and centered at the pricing restriction.

Note that

\[
\alpha' \left( \frac{\sigma^2}{s^2} \Sigma \right)^{-1} \alpha = a' (\Sigma^{-1} \otimes D) a
\]

where \(D\) is a \((k+1) \times (k+1)\) matrix whose \((1,1)\) element is \(\frac{s^2}{\sigma^2}\) and all other elements are zero. So, we can rewrite the density of \(\alpha|\Sigma\) as

\[
p(\alpha | \Sigma) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} a' (\Sigma^{-1} \otimes D) a \right\}
\]

Combining the prior with the likelihood function we get the posterior distribution of \(\theta\):

\[
p(\theta | R) \propto p(R|\theta)p(\theta)
\]

An interesting fact is that the posterior distribution can be factored into two parts, one that involves the regression parameters \((a, \Sigma)\) and another that involves the benchmark moments \((E_2, V_{22})\). To see this, note that

\[
p(\theta | R) \propto p(R|\theta)p(\theta)
\]
\[
\propto [p(Y|\theta, X)p(X|\theta)]\left[ p(\alpha|\Sigma)p(\Sigma)p(B)p(E_2)p(V_{22}) \right] (130)
\]
\[
\propto [p(Y|\theta, X)p(\alpha|\Sigma)p(\Sigma)p(B)][p(X|E_2, V_{22})p(E_2)p(V_{22})] (132)
\]
\[
\propto p(a, \Sigma|R)p(E_2, V_{22}|X) (133)
\]

The joint posterior of the regression parameters is just proportional to the product
\[ p(Y|\theta, X)p(\alpha|\Sigma)p(\Sigma)p(B), \] that is
\[
p(a, \Sigma| R) \propto |\Sigma|^{-\frac{k+1}{2}} \exp \left\{ -\frac{1}{2} \left[ a'(\Sigma^{-1} \otimes D)a + (a - \hat{a})'(\Sigma^{-1} \otimes Z'Z)(a - \hat{a}) \right] \right\} \\
\times |\Sigma|^{-\frac{T+v+k+m+1}{2}} \exp \left\{ -\frac{1}{2} tr(H + T\hat{\Sigma})\Sigma^{-1} \right\}, \]

Letting \( \hat{a} = (I_m \otimes F^{-1}Z'Z)\hat{a}, \ F = D + Z'Z, \ Q = Z'(I_T - ZF^{-1}Z')Z \) we have
\[
p(a, \Sigma| R) \propto |\Sigma|^{-\frac{k+1}{2}} \exp \left\{ -\frac{1}{2} [(a - \hat{a})'(\Sigma^{-1} \otimes F)(a - \hat{a})] \right\} \\
\times |\Sigma|^{-\frac{T+v+k+m+1}{2}} \exp \left\{ -\frac{1}{2} tr(H + T\hat{\Sigma} + \hat{A}'Q\hat{A})\Sigma^{-1} \right\}, \]

Now, from \( p(a, \Sigma| R) = p(a|\Sigma, R)p(\Sigma| R), \) it is clear that \( \Sigma| R \) have a IW\((H + T\hat{\Sigma} + \hat{A}'Q\hat{A}, T + v - k, m)\). It follows that
\[
a|\Sigma, R \sim N(\hat{a}, \Sigma \otimes F^{-1}) \\
\Sigma^{-1}|R \sim W((H + T\hat{\Sigma} + \hat{A}'Q\hat{A})^{-1}, T + v - k, m) \]

Therefore,
\[
\hat{a} = E(a| R) = (I_m \otimes F^{-1}Z'Z)\hat{a} \\
\hat{\Sigma} = E(\Sigma| R) = \frac{1}{T + v - k - m - 1}(H + T\hat{\Sigma} + \hat{A}'Q\hat{A}) \\
Var(a| R) = \hat{\Sigma} \otimes F^{-1}. \]

Where the third equality follows from the Law of Total Conditional Variance which states that \( Var[a| R] = E[Var(a|\Sigma, R)| R] + Var[E(a|\Sigma, R)| R]. \)

Now, the joint posterior of the benchmark moments is just proportional to
\[
p(E_2, V_{22})p(X| E_2, V_{22}) \\
, \] that is
\[
p(E_2, V_{22}|X) \propto |V_{22}|^{-\frac{T+v+1}{2}} \exp \left\{ -\frac{T}{2} tr\hat{V}_{22}V_{22}^{-1} - \frac{T}{2} tr(E_2 - \hat{E}_2)(E_2 - \hat{E}_2)'V_{22}^{-1} \right\} \]
It follows that

\[ E_2 | V_{22}, R \sim N(\hat{E}_2, \frac{1}{T} V_{22}) \]
\[ V_{22}^{-1} | R \sim W((T\hat{V}_{22})^{-1}, T - 1, k). \]

Therefore,

\[ \hat{E}_2 = E(E_2 | R) = \hat{E}_2 \]
\[ \hat{V}_{22} = E(V_{22} | R) = \frac{T}{T - k - 2} \hat{V}_{22} \]
\[ Var(E_2 | R) = E[Var(E_2 | V_{22}, R) | R] = \frac{1}{T} E[V_{22} | R] \]
\[ = \frac{1}{T - k - 2} \hat{V}_{22}. \]

Where the third equality, again, follows the Law of Total Conditional Variance. Now we only need to note that the predictive mean obeys the relation,

\[ E^* = E(r_{T+1} | R) = E(E(r_{T+1} | \theta, R) | R) = E(E | R) = \tilde{E}. \]

Since \( B \) and \( E_2 \) are independent in the posterior, the mean of the predictive distribution is

\[ E^* = \tilde{E} = E \begin{pmatrix} \alpha + BE_2 \\ E_2 \end{pmatrix} = \begin{pmatrix} \tilde{\alpha} + \tilde{B}\tilde{E}_2 \\ \tilde{E}_2 \end{pmatrix} \tag{134} \]

where \( \tilde{\alpha} \) and \( \tilde{B} \) are obtained from

\[ \tilde{a} = (I_m \otimes F^{-1} Z'Z) \hat{a} \]

where \( \tilde{a} = vec((\tilde{\alpha} \quad \tilde{B})') \). To compute \( V_{22}^* \) we only need to note that

\[ V^* = Var(r_{T+1} | R) = E(V | R) + Var(E | R) = \tilde{V} + Var(E | R), \]

We can partition the predictive covariance matrix as

\[ V^* = \begin{bmatrix} V_{11}^* & V_{12}^* \\ V_{21}^* & V_{22}^* \end{bmatrix} \]
Applying this rule to the lower right submatrix gives

\[ V_{22}^* = \tilde{V}_{22} + Var(E_2|R), \]

And applying it to the off diagonal submatrices gives

\[ V_{12}^* = V_{21}^* = E(BV_{22}|R) + Cov(\alpha + BE_2, E_2'|R) = \tilde{B}\tilde{V}_{22} + BVar(E_2|R) \]

A little bit of extra work is needed for the \( V_{11}^* \) elements. We use the decomposition

\[ Cov(y_{i,T+1}, y_{j,T+1}|R) = E(Cov(y_{i,T+1}, y_{j,T+1}|a, R)|R) + Cov(E(y_{i,T+1}|a, R), E(y_{j,T+1}|a, R)|R) \]

where

\[ E(Cov(y_{i,T+1}, y_{j,T+1}|a, R)|R) = \tilde{b}_i V_{22}^* \tilde{b}_j + \text{Tr}[V_{22}^* Cov(b_i, b_j'|R)] + \tilde{\sigma}_{ij} \]

and

\[ Cov(E(y_{i,T+1}|a, R), E(y_{j,T+1}|a, R)|R) = [1 \tilde{E}_2'] Cov(a_i, a_j'|R) [1 \tilde{E}_2']' \]

Here, \( Cov(b_i, b_j'|R) \) and \( Cov(a_i, a_j'|R) \) are submatrices of \( Var(a|R) \).

It is interesting to note that if the investor believes dogmatically in the model (\( \sigma_\alpha = 0 \)), then the predictive mean is the sample mean and the covariance is inflated to

\[ V_{22}^* = \left( \frac{T + 1}{T - k - 2} \right) \hat{V}_{22} \approx \left( 1 + \frac{1}{T} \right) \hat{V}_{22} \]

which is the Bayes Stein estimator with zero precision. The prescription to find the others submatrix from the predictive second moment is given in the appendix.

### 4.6 Brief Description of Asset Allocation Strategies Considered

In this section we summarize the asset allocation models considered. The main difference between this models is how to estimate the unknown population parameters \( \mu \) and \( \Sigma \) from the data.

#### 4.6.1 Naive Risk Portfolio

The naive risk portfolio is defined as the portfolio that invest equal weights, \( 1_N/N \), in the risk assets. This strategy is considered naive since it disregard the data and any sort of optimization problem. Note that in this case no amount is invested in the risk free asset. We later consider a naive strategy that takes the risk free asset under consideration.
4.6.2 Sample-based mean-variance portfolio

The sample mean-variance portfolio follows the classic "plug-in" approach, that is, it substitutes the population parameters $\mu$ and $\Sigma$ by its sample counterparts in the optimization rule. This strategy ignores the possibility of estimation errors. In this case we have $a_{mv} = (1/\gamma)\Sigma^{-1}\hat{\mu}$.

4.6.3 Bayesian diffuse-prior portfolio

Barry (1974), Klein and Bawa (1976), and Brown (1979) show that if the prior for $\Theta = (\mu, \Sigma)$ is diffuse, and the conditional likelihood is normal, then the predictive distribution is a student-t with mean $\hat{\mu}$ and variance $\hat{\Sigma}(1 + 1/M)$, where $M$ is the estimation window. Since in our study $M = 120$ months, the difference from the sample mean-variance portfolio is negligible and the results of this strategy are not reported.

4.6.4 Jorion-Bayes-Stein portfolio

As described before, this strategy consider correcting estimation error by setting

$$\hat{\mu}^{JBS} = (1 - \hat{\phi})\hat{\mu} + \hat{\phi}\hat{\mu}^{min}1_N$$

$$\hat{\phi} = \frac{N + 2}{(N + 2) + T(\hat{\mu} - \hat{\mu}^{min}1_N)^T\hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}^{min}1_N)}$$

where

$$\hat{\Sigma}^{JBS} = \frac{T - 1}{T - N - 2} \hat{\Sigma}$$

and $\mu_{MIN} = a_{MIN}^T\hat{\mu}$ is the average excess return on the sample global minimum variance portfolio. This sample estimator is called a shrink estimator because it shrink the sample mean towards the mean of the minimum variance portfolio.

4.6.5 Pastor-Stambaugh Bayesian portfolio

As we described here, Pastor Stambaugh adress the arbitrariness of the choice of a shrinkage target $\bar{\mu}$, and of the shrinkage factor, $\phi$, by using the investor’s belief about the validity of an asset pricing model. As for the asset pricing model we consider the Capital Asset Pricing Model (CAPM) and the Fama and French (1993) three-factor model.

4.6.6 Global Minimum-Variance portfolio

As its name suggests, the minimum variance portfolio strategy is the one that minimizes the variance of returns. To implement this strategy we only need to estimate the covariance matrix of asset returns. This strategy suggests $a_{MIN} = \Sigma^{-1}1_N/1_N\Sigma^{-1}1_N$. Note that this strategy does not take the risk free asset under consideration.
4.7 Kan and Zhou two-fund portfolio

The Kan and Zhou optimal two fund rule suggest choosing

\[ a(\hat{\theta}) = \frac{c}{\gamma} \hat{\Sigma}^{-1} \hat{\mu} \tag{137} \]

where the value of \( c \) is the one that minimizes expected loss. Since the exact value of \( c \) is not feasible, we set \( c \) to the approximate value that does not depend of the unknown population values. That is, we set

\[ c = \frac{(T - N - 1)(T - N - 4)}{T(T - 2)} \tag{138} \]

We favor here the Kan-Zhou two portfolio rule with respect to the three fund rule because of its simplicity and for the fact that the three fund rule is not feasible, so the sample analog add estimation error that turns its optimality questionable.

4.8 Mixture of equally weighted and minimum-variance portfolios

Garllapi Uppal consider the case where investors combine the equally weighted with the minimum-variance portfolio. Their motivation is that because expected returns are more difficult to estimate than covariances, one way to ignore the estimates of mean returns but not the estimates of covariances is to consider this strategy. This portfolio strategy is

\[ \hat{a}^{ew-min} = c \frac{1}{N} 1_N + d \hat{\Sigma}^{-1} 1_N \tag{139} \]

where \( c \) and \( d \) are chosen to maximize expected out of sample performance.

4.8.1 Naive Mixture of Portfolios

We consider a new portfolio strategy that has not been studied in the existing literature. This is strategy is a naive combination of two portfolio rules. Our motivation is that although mean and variance are difficult to estimate, both may be useful to reduce estimation uncertainty, that is, different believes may result in different estimation risks that maybe diversified. Since we do not want to favor any of the chosen models, and since usually, the optimal weights that maximize out of sample performance are not feasible, we set equal weights to both portfolio rules.

The natural candidates are combinations of the sample based optimal mean variance
portfolio with the global minimum variance portfolio, that is

$$\hat{a}_{nmv-min} = \frac{1}{2} \left( \frac{\hat{\Sigma}^{-1} \hat{\mu}}{\gamma} + \frac{\hat{\Sigma}^{-1}1_N}{1_N \hat{\Sigma}^{-1}1_N} \right)$$

and the naive risk portfolio with the global minimum variance portfolio,

$$\hat{a}_{new-min} = \frac{1}{2} \left( \frac{1}{N}1_N + \frac{\hat{\Sigma}^{-1}1_N}{1_N \hat{\Sigma}^{-1}1_N} \right)$$

We also consider the Jorion-Bayes-Stein optimal portfolio with the global minimum variance portfolio,

$$\hat{a}_{njbs-min} = \frac{1}{2} \left( \frac{\hat{\Sigma}^{-1} \hat{\mu}_{JBS}}{\gamma} + \frac{\hat{\Sigma}^{-1}1_N}{1_N \hat{\Sigma}^{-1}1_N} \right)$$

4.9 Data Description and Methodology for Evaluating Performance

We consider a data set of 12 industry portfolios as the risk assets and 1 risk-free asset that here is the 1 month T-bill. We use 120 months of data retrieved from french.com for the industry portfolios and the T-bill retrieved from fed.com.

We follow here the approach used in the literature that relies on a "rolling sample". This approach consider an estimation window of size M. In each month t, starting from t=M+1, we use the data in the previous M months to estimate the parameters needed to implement a particular strategy. The outcome of this rolling-window approach is a series of T-M monthly out of sample returns generated by each of the portfolio strategies listed.

Given the time series of monthly out of sample returns generated by each strategy and in each dataset, we compute 3 quantities. The out of sample Sharpe ratio of strategy k:

$$\hat{SR}_k = \frac{\hat{\mu}_k}{\hat{\sigma}_k} \quad (140)$$

The certainty equivalent return, defined as the risk free rate that an investor is willing to accept rather than adopting a particular risky portfolio strategy:

$$\hat{CEQ}_k = \hat{\mu}_k - \frac{\gamma}{2} \hat{\sigma}_k^2 \quad (141)$$

In which $$\hat{\mu}_k$$ and $$\hat{\sigma}_k^2$$ are the mean and variance of out of sample portfolio returns for strategy k, and $$\gamma$$ is the risk aversion. For the third evaluation method we compute the
return-loss with respect to the 1/N strategy. The return-loss is defined as the additional return needed for strategy \( k \) to perform as well as the 1/N strategy in terms of the Sharpe ratio. Let \( \mu_{ew} \) and \( \sigma_{ew} \) be the monthly out of sample mean and volatility of the returns from 1/N strategy. The return-loss from strategy \( k \) is

\[
\text{return-loss}_k = \frac{\mu_{ew} - \mu_k}{\sigma_{ew}} - \sigma_k
\]

It is important to point out that to compute the welfare loss of uncertainty, we must compare what should be the individual utility when the investor is certain about the pricing model. That is, if \( \sigma_\alpha = 0 \) in the Bayesian framework. We can interpret the results in the following manner. How much money an investor would be willing to give up in order to be sure about the true pricing model.

4.10 Results

4.10.1 Results from Empirical Datasets

The first strategy considered is the one the individual invest in the naive risk portfolio. In this case the individual disregard the risk free asset and invest all in the risk assets with equal amounts in each asset.

The second strategy considered is the one the individual invests in the global minimum variance portfolio. In this case the investor disregard the risk free asset and invest all the amount in the portfolio that minimizes the variance of the risk portfolio return. Performance, read, utility, is computed for \( \gamma = 10 \).

The third strategy we consider is the mean variance portfolio with the sample mean and sample variance as estimates for the populational moments. The table below reports the sharpe ratio in sample and out of sample of the resulting portfolio for each value of \( t \) ranging from 61 to 119 for the selected value of \( \gamma = 10 \).

The fourth strategy we consider is the Jorion-Bayes Stein estimate for the mean. The idea is to shrink the mean toward the grand mean that here is the minimum variance mean. With this strategy, a smaller estimation risk is achieved.

The fifth strategy is Kan-Zhou two fund portfolio. This strategy attains an even lower estimation risk than Jorion suggestion.

The next strategy we are now going to consider is Pastor and Stambaugh framework that consider individual beliefs in the asset pricing model. The beliefs can be summarized in the value of the parameter \( \sigma_\alpha \). If \( \sigma_\alpha = 0 \) we say that the individual truly believes in the asset pricing model, if \( \sigma_\alpha = \infty \) we say that the individual consider the asset pricing
model as useless. We follow Garlappi-Uppal and consider $\sigma_\alpha = 1\%$. The benchmark asset considered here is based in the CAPM model where we consider the SP500 as the benchmark.

As a matter of comparison, we are going to seek combinations of each individual strategy with the minimum variance portfolio. Since the minimum variance portfolio only depends on sample estimation of the covariance matrix, but not on the individual assets mean, it is a good portfolio to be considered for combined strategies.

### 4.10.2 Table of results

<table>
<thead>
<tr>
<th>Strategy/t</th>
<th>61</th>
<th>71</th>
<th>81</th>
<th>91</th>
<th>101</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1/N$</td>
<td>0.0681</td>
<td>0.0881</td>
<td>0.1358</td>
<td>0.1560</td>
<td>0.1659</td>
<td>0.1485</td>
</tr>
<tr>
<td>$min$</td>
<td>0.3058</td>
<td>0.3899</td>
<td>0.3445</td>
<td>0.3214</td>
<td>0.3500</td>
<td>0.3334</td>
</tr>
<tr>
<td>$mv$</td>
<td>0.5939</td>
<td>0.5404</td>
<td>0.4963</td>
<td>0.4423</td>
<td>0.4648</td>
<td>0.4494</td>
</tr>
<tr>
<td>$JBS$</td>
<td>0.5599</td>
<td>0.5075</td>
<td>0.4584</td>
<td>0.4007</td>
<td>0.4333</td>
<td>0.4187</td>
</tr>
<tr>
<td>$KZ$</td>
<td>0.5939</td>
<td>0.5404</td>
<td>0.4963</td>
<td>0.4423</td>
<td>0.4648</td>
<td>0.4494</td>
</tr>
<tr>
<td>$PS$</td>
<td>0.6129</td>
<td>0.5715</td>
<td>0.5417</td>
<td>0.5064</td>
<td>0.5164</td>
<td>0.4904</td>
</tr>
<tr>
<td>$mv - min$</td>
<td>0.5633</td>
<td>0.5212</td>
<td>0.4724</td>
<td>0.4196</td>
<td>0.4452</td>
<td>0.4288</td>
</tr>
<tr>
<td>$ew - min$</td>
<td>0.1643</td>
<td>0.2117</td>
<td>0.2356</td>
<td>0.2416</td>
<td>0.2604</td>
<td>0.2417</td>
</tr>
<tr>
<td>$JBS - min$</td>
<td>0.4710</td>
<td>0.4559</td>
<td>0.4055</td>
<td>0.3585</td>
<td>0.3930</td>
<td>0.3782</td>
</tr>
</tbody>
</table>
Out of Sample Sharpe ratios for empirical data

<table>
<thead>
<tr>
<th>Strategy/t</th>
<th>61</th>
<th>71</th>
<th>81</th>
<th>91</th>
<th>101</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/N</td>
<td>0.3720</td>
<td>0.3724</td>
<td>0.2746</td>
<td>0.2123</td>
<td>0.1580</td>
<td>0.6741</td>
</tr>
<tr>
<td>min</td>
<td>0.2887</td>
<td>0.2114</td>
<td>0.2304</td>
<td>0.2355</td>
<td>0.0469</td>
<td>−0.0814</td>
</tr>
<tr>
<td>mv</td>
<td>0.0727</td>
<td>0.1135</td>
<td>0.0570</td>
<td>0.0708</td>
<td>−0.1017</td>
<td>−0.2672</td>
</tr>
<tr>
<td>JBS</td>
<td>0.1558</td>
<td>0.1586</td>
<td>0.1362</td>
<td>0.1575</td>
<td>−0.0409</td>
<td>−0.2072</td>
</tr>
<tr>
<td>KZ</td>
<td>0.0727</td>
<td>0.1135</td>
<td>0.0570</td>
<td>0.0708</td>
<td>−0.1017</td>
<td>−0.2672</td>
</tr>
<tr>
<td>PS</td>
<td>0.1434</td>
<td>0.1675</td>
<td>0.1138</td>
<td>0.1048</td>
<td>−0.0371</td>
<td>−0.1843</td>
</tr>
<tr>
<td>mv − min</td>
<td>0.1519</td>
<td>0.1490</td>
<td>0.1203</td>
<td>0.1349</td>
<td>−0.0548</td>
<td>−0.2200</td>
</tr>
<tr>
<td>ew − min</td>
<td>0.3832</td>
<td>0.3333</td>
<td>0.2881</td>
<td>0.2593</td>
<td>0.1203</td>
<td>0.3184</td>
</tr>
<tr>
<td>JBS − min</td>
<td>0.2181</td>
<td>0.1812</td>
<td>0.1788</td>
<td>0.1942</td>
<td>−0.0041</td>
<td>−0.1624</td>
</tr>
</tbody>
</table>

In Sample Certainty Equivalent for empirical data

<table>
<thead>
<tr>
<th>Strategy/t</th>
<th>61</th>
<th>71</th>
<th>81</th>
<th>91</th>
<th>101</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/N</td>
<td>−0.0123</td>
<td>−0.0097</td>
<td>−0.0062</td>
<td>−0.0044</td>
<td>−0.0033</td>
<td>−0.0041</td>
</tr>
<tr>
<td>min</td>
<td>0.0046</td>
<td>0.0067</td>
<td>0.0057</td>
<td>0.0051</td>
<td>0.0059</td>
<td>0.0054</td>
</tr>
<tr>
<td>mv</td>
<td>0.0176</td>
<td>0.0146</td>
<td>0.0123</td>
<td>0.0098</td>
<td>0.0108</td>
<td>0.0101</td>
</tr>
<tr>
<td>JBS</td>
<td>0.0139</td>
<td>0.0115</td>
<td>0.0098</td>
<td>0.0077</td>
<td>0.0088</td>
<td>0.0083</td>
</tr>
<tr>
<td>KZ</td>
<td>0.0147</td>
<td>0.0128</td>
<td>0.0111</td>
<td>0.0090</td>
<td>0.0101</td>
<td>0.0096</td>
</tr>
<tr>
<td>PS</td>
<td>0.0184</td>
<td>0.0156</td>
<td>0.0145</td>
<td>0.0128</td>
<td>0.0132</td>
<td>0.0120</td>
</tr>
<tr>
<td>mv − min</td>
<td>0.0141</td>
<td>0.0124</td>
<td>0.0105</td>
<td>0.0084</td>
<td>0.0094</td>
<td>0.0088</td>
</tr>
<tr>
<td>ew − min</td>
<td>−0.0007</td>
<td>0.0013</td>
<td>0.0021</td>
<td>0.0024</td>
<td>0.0031</td>
<td>0.0024</td>
</tr>
<tr>
<td>JBS − min</td>
<td>0.0096</td>
<td>0.0090</td>
<td>0.0077</td>
<td>0.0062</td>
<td>0.0073</td>
<td>0.0068</td>
</tr>
</tbody>
</table>
4.10.3 Results from Simulated Data

In this section, we follow the approach in the literature and evaluate how the performance of the portfolio strategies described above depends on the number of assets, N, and the length of the estimation window, M. The usefulness of simulated data comes from the fact that we know exactly their economic and statistical properties. We simulate returns in the same fashion as Garlappi et al. (2007) and MackinlayPastor (2000). As before, we assume that there are N risky assets and 1 risk-free asset. K of the N risk assets are treated as factors, which leaves N-K remaining risky assets generated by the factor model

\[ R_{a,t} = \alpha + BR_{b,t} + \varepsilon_t, \]

where \( R_{a,t} \) is the vector of excess return of assets, \( \alpha \) is the mispricing coefficients vector, \( B \) is the factor loadings matrix, \( R_b \) is the vector of excess returns on the factor portfolios. We assume that the returns on the factor are normally distributed, that is,

\[ R_b \sim N(\mu_b, \Omega_b) \]

and the vector of noise is normally distributed with mean zero, that is,

\[ \varepsilon \sim N(0, \Sigma_\varepsilon) \]

and independent with respect to the factor portfolios. We consider that there is only one factor, \( K = 1 \), whose annual excess return has an annual average of 8% and standard deviation of 16%. The mispricing coefficient is set to zero, \( \alpha = 0 \), the factor of loadings
are set to be evenly spread between .5 and 1.5. The risk free rate is also assumed to be normally distributed with annual mean of 2% and standard deviation of 2%. We consider cases with number of assets $N = \{10, 25, 50\}$, and estimation window lengths $M = \{120, 360, 6000\}$ months, which correspond to 10, 30, and 500 years. We use Monte Carlo sampling to generate monthly return data for $T = 24,000$ months.

### 4.11 Conclusion

By analyzing the monthly returns from the 12 industry portfolio for the 10 years period from January 2006 to January 2017 we were able to form portfolio strategies based on expected loss minimization and predictive moments from Bayesian beliefs. When comparing this models with the naive portfolio, that invests equal amount in each of the risk assets, none were able to outperform it. This confirms Garlappi-Uppal findings that we are still miles away when looking at only single portfolio strategies. Besides this fact, we were able to form a simple mixture of the naive portfolio with the minimum variance portfolio that improves in sample performance whether maintaining similar out of sample performance. This suggests an investigation of possible optimal strategies that combine the risk free asset with this mixture of naive risk asset and minimum variance portfolio.
.1 Appendix

.1.1 The derivative of the determinant as a cofactor

Let $B = b_{ij}$ be a $p \times p$ matrix. Then

$$\frac{\partial |B|}{\partial b_{ij}} = B_{ij}$$

To see this we only need to expand $|B|$ by the elements of the $i$th row

$$|B| = \sum_{h=1}^{p} b_{ih} B_{ih}$$

since $B_{ih}$ does not contain $b_{ij}$, the result follows.

.2 The Jacobian of the Inverse

Let $A = G^{-1}$ be a PDS matrix. The Jacobian of the transformation can be shown to be $|A|^{-(m+1)}$. To see this, write $AG = I$. Then

$$\frac{\partial A}{\partial \theta} G + A \frac{\partial G}{\partial \theta} = 0$$

or

$$\frac{\partial G}{\partial \theta} = -G \frac{\partial A}{\partial \theta} G$$

If $\theta = a_{ij}$, we have $\partial g_{\alpha\beta}/\partial a_{ij} = -g_{\alpha i} g_{\beta j}$, for $\beta \leq \alpha$ and $j \leq i$, since $A$ and $G$ are symmetric and the transformation from elements of $G$ to those of $A$ involves just $m(m+1)/2$ distinct elements of $G$. On forming the Jacobian matrix and taking its determinant, we have $|G|^{m+1} = |A|^{-(m+1)}$.

.2.1 Proof of the First Moment of the Inverted Whishart

We show here that if $A$ has a $W(\Sigma, n, p)$ distribution, then

$$E[A^{-1}] = (1/(n - p - 1))\Sigma^{-1}$$

To see this, let $C$ be such that $\Sigma = CC'$. Then we can write

$$A = CBC'$$
where $B$ has a $W(I, n)$ distribution and

$$E[A^{-1}] = C'^{-1}E[B^{-1}]C^{-1}$$

By symmetry we can write

$$E[B^{-1}] = k_1 I + k_2 ee'$$

Now note that for every orthogonal matrix $Q$, $QBQ'$ has a $W(I, n)$ distribution. This is true, since if $B = ZZ'$ where $Z$ is $N(0, I)$, then $QBQ' = QZZ'Q'$ where $QZ$ is also $N(0, I)$. This imply that $E[(QBQ')^{-1}] = QE[B^{-1}]Q' = E[B^{-1}]$ for every orthogonal matrix $Q$. For that, we must have $k_2 = 0$ and

$$E[B^{-1}] = k_1 I$$

Now we just need to note that the diagonal terms of $B^{-1}$ are $I\chi^2(n-p+1)$. Since $E[I\chi^2(n-p+1)] = (n-p-1)^{-1}$, we have

$$E[B^{-1}] = (n-p-1)^{-1}I$$

We conclude that

$$E[A^{-1}] = C'^{-1}E[B^{-1}]C^{-1}$$
$$= (n-p-1)^{-1}(CC')^{-1}$$
$$= (n-p-1)^{-1}\Sigma^{-1}$$
References


