DISCUSSION PAPER SERIES

No. 6161

ROBUST PORTFOLIO OPTIMISATION WITH MULTIPLE EXPERTS

Frank Lutgens and Peter C Schotman

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP6161.asp**

ISSN 0265-8003

ROBUST PORTFOLIO OPTIMISATION WITH MULTIPLE EXPERTS

Frank Lutgens, Maastricht University and NETSPAR
Peter C Schotman, Maastricht University, NETSPAR and CEPR

Discussion Paper No. 6161
March 2007

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Frank Lutgens and Peter C Schotman

CEPR Discussion Paper No. 6161

March 2007

ABSTRACT

Robust Portfolio Optimisation with Multiple Experts*

We consider mean-variance portfolio choice of a robust investor. The investor receives advice from J experts, each with a different prior for the distribution of returns. Confronted with these multiple priors the investor follows a min-max portfolio strategy. We study the structure of the robust mean-variance portfolio and empirically compare its performance with a variety of alternative portfolio strategies. The empirical tests are based on bootstrap simulations on the 25 Fama-French portfolios and on 81 European country and value portfolios. We find that the robust portfolio performs well in both settings. Robust portfolios do not exhibit the extreme weights typically observed in naive mean-variance portfolios. Robust portfolios are also better diversified than portfolios that impose short-sell constraints to suppress the symptoms of extreme weights.

JEL Classification: C11 and D80
Keywords: mean-variance, model uncertainty and portfolio choice

Frank Lutgens
Department of Finance
Maastricht University
PO Box 616
NL-6200 MD Maastricht.
THE NETHERLANDS
Email: f.lutgens@finance.unimaas.nl

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=166135

Peter C Schotman
LIFE
Maastricht University
PO Box 616
NL-6200 MD Maastricht.
THE NETHERLANDS
Email:
P.Schotman@finance.unimaas.nl

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=117320

* An early version of this paper was presented at the London Business School, the Swedish Institute for Financial Research (Stockholm), the Econometric Study Group (London), ABP Investments and at the Inquire Europe conference in Hamburg. We thank Mark Salmon for an insightful discussion of our results. We thank Inquire Europe for their financial support of this research.

Submitted 12 February 2007

1 Introduction

Quantitative portfolio management requires estimates of expected return and risk. Unfortunately analysts often disagree about both the general outlook of the stock market and the relative risks and returns of individual stocks. They employ different models to estimate expected return and risk. The decision maker is thus confronted with uncertain inputs over which to optimise the portfolio.

We analyse the decisions and performance of a robust decision maker. In our setting the robust portfolio is defined as the portfolio that performs best under the worst case scenario formed by the recommendations of J expert advisers. The robust expected utility of a portfolio is the minimum expected utility over each of the alternative analyst recommendations. The optimal robust portfolio achieves the highest minimum expected utility.

There are two ways to motivate robust portfolio decisions. First, from a behavioral viewpoint the decision maker can be ambiguity averse. Gilboa and Schmeidler (1989) and Klibanoff, Marinacci and Mukerji (2005) provide an axiomatic foundation for the maxmin utility optimisation of the decision maker.

The second motivation comes from an operations research perspective. Historical average returns are very noisy estimates of the mean return. With a 20% annual standard deviation and 25 years of historical data, the standard error on the mean is of an order of magnitude $20/\sqrt{25} = 4\%$. A large historical average return for a stock could either indicate truly high returns or just result from positive estimation error. Estimation error is thus a large component of estimated expected returns. The problem is most acute in situations where many assets are positively correlated. The mean-variance optimisation attempts to arbitrage on small profit opportunities, and often selects a portfolio with extreme positive and negative weights. Such portfolios are known to perform poorly out-of-sample.[1] The result is what Michaud (1989) has

[1] Early references include Jobson and Korkie (1981) and Jorion (1985). Most recently DeMiguel, Garlappi and Uppal (2006) find that optimised portfolios rarely outperform a naive equally weighted benchmark. Kan and Zhou (2006) provide further analytical treatment and more references.

called 'error maximisation'.

For the investor the performance of the mean-variance portfolio is disappointing for two reasons. The portfolio is not only suboptimal, but the estimated means and covariances grossly overestimate the true investment opportunities. Since the robust investor looks at different sets of estimates, and for each potential portfolio evaluates the worst case, the robust investor is less likely to overestimate the available investment opportunities, especially when the different experts disagree. For that reason we conjecture that the robust portfolio will actually exhibit good out-of-sample performance. The robust decision rule advocates a portfolio that is not necessarily optimal for a given set of inputs, but that performs reasonably well in many different circumstances. Robustness also refers to a different desirable property: small changes in the input have only little impact on the portfolio choice, in contrast to the standard mean-variance portfolio.

To deal with the estimation error and alternative investment views many studies have suggested a Bayesian approach. An informative prior reduces the estimation noise and this improves the actual performance of the portfolios. In most applications the expected returns are shrunk to ex-ante reasonable means. For example, Jorion (1986) shrinks the individual sample means towards a common grand mean. In this setup all stocks have the same prior expected return, and the mean-variance optimiser has less scope to form extreme portfolios. In Black and Litterman (1990) the prior means are the expected returns for which the observed market portfolio would be optimal. In Pàstor and Stambaugh (2000) the prior is centered around the expected returns implied by the CAPM. A special type of implicit prior is defined by restricting the portfolio weights like the short-sell constraints in Jagannathan and Ma (2003).

In many ways our setting is similar to Black and Litterman (1990). Like the Bayesian investor in Black and Litterman, our robust investor combines the distinct views of a finite set of experts. The important difference is that the Bayesian investor must attach exogenous prior probabilities to each of the views. The robust investor

is not able to attach weights to the different views, but determines the weights of the different views endogenously as part of the portfolio optimisation. More fundamentally, the robust investor dislikes the uncertainty generated by the dispersion in expert opinions.

Closely related to our work are Garlappi, Uppal and Wang (2007) and Goldfarb and Iyengar (2003). Both studies also consider robust solutions to the static mean-variance problem. Unlike us, these papers deal with estimation uncertainty. The investor recognises that the parameter estimates are subject to estimation error and chooses the worst case within a classical confidence set. Our approach considers model uncertainty, where the investor is confronted with alternative models of returns. Each prior represents a different model of risk and return. In that sense our approach also deviates from DeMiguel et al. (2006). Whereas DeMiguel et al. (2006) consider the performance of individual models, our model robust portfolios combine the inputs from different models. When the different models produce markedly distinct estimates of expected returns and covariances, the robust portfolio will use a weighted average of the different estimates. If different models produce similar estimates, however, the robust portfolio will not perform better than any of the single models.

For the more elaborate priors we will need numerical methods to construct the optimal portfolios and evaluate the performance. Analytical results are available, however, for a number of interesting special cases. The first special case shows that the robust portfolio is not always the most conservative portfolio. At times the robust solution features a larger risky investment than the pessimistic recommendations would. This result is interesting, since robust decision rules (in particular the maxmin approach) are often criticised for being extremely conservative.

The second case determines under which conditions a robust investor performs better than investors who base their decisions on a single model. Assuming that the covariance matrix is known, the robust portfolio outperforms if the differences in opinion among experts have weak correlations with the true expected returns.

We analytically also study a third special case to explain why the robust investor largely relies on the most structured view. Again assuming a known covariance matrix, we find that the most informative prior on an asset pricing model will be binding in determining the robust portfolio. This implies that if one of the experts has reasonably strong beliefs in the CAPM, then the CAPM will determine the robust expected returns. Since portfolios weights are usually well-behaved under asset-pricing conditions, the robust portfolio will not have extreme weights in this case.

The only advice that matters for the decision maker is the worst case expected utility associated with each portfolio. This means that the approach will only work well in practice when the expert opinions are reasonable. Therefore we avoid dogmatic priors in the empirical analysis. Experts have different priors, but they share the same five or ten years of data. If the data are informative, they must be able to alter the prior views.

For the empirical tests we analyse the performance of the robust portfolio choice using two different data sets. The first data set contains the 25 Fama-French portfolios sorted by size and book-to-market. The second data set considers 81 portfolios from 9 European countries sorted by size and book-to-market. The experts have different prior views on the relation between expected returns and risk, and on the structure of the covariance matrix. Their views are based on factor models with different priors on the validity of the CAPM, the International CAPM and the Fama-French 3-factor model. Each expert uses the Pàstor and Stambaugh (2000) model to obtain their estimates. The only difference between the experts are their prior views. The empirical analysis confirms that the robust portfolio in many cases has better performance than portfolios based on each of the individual priors held by the different experts.

The remainder of the paper is structured as follows. Section 2 develops the model robust portfolio choice. In sections 3 and 4 we outline the form of the expert's prior views. Section 5 describes the numerical simulation that we use to measure the empirical performance. Section 6 reports the results and section 7 concludes.

2 Robust mean-variance portfolio choice

We consider an investor with mean variance preferences over single-period portfolio choice. Her problem is to find the optimal allocation to N risky assets and a riskfree asset. The expected excess returns (relative to the riskfree rate) and covariances are denoted by the N-vector μ and $N \times N$ covariance matrix Σ respectively. Portfolio weights are referred to by the N-vector w. The investor maximizes the mean variance utility

$$Q(w) = \mu' w - \tfrac{1}{2}\gamma w' \Sigma w, \tag{1}$$

with γ measuring risk aversion. Since a riskfree asset exists, portfolio weights are not restricted to sum to one. The fraction of wealth not invested in risky assets, $1 - \iota' w$, is invested in the riskfree asset. As $\iota' w$ can be larger than one, we allow for borrowing. When μ and Σ are known, the optimal investment in the risky assets is

$$w_0 = \frac{1}{\gamma}\Sigma^{-1}\mu. \tag{2}$$

In practice, the investor does not know the true values of μ and Σ. She obtains advice from J experts who recommend their personal estimate for the expected returns and variance matrices (μ_j, Σ_j). The set of all recommendations is denoted as $\Theta = \{(\mu_j, \Sigma_j); j = 1, \ldots, J\}$. The investor does not have information on the reliability of the experts. She does not want to disregard anyone's recommendation, just in case that particular expert turns out to be right. In terms of the framework by Klibanoff et al. (2005), our investor is infinitely averse to the ambiguity in the rival experts' recommendations. Therefore the investor maximizes the minimum expected utility implied by the various experts,

$$Q_R(w) = \min_j Q_j(w) \tag{3}$$

where Q_j refers to the mean variance utility according to expert j''s recommendation,

$$Q_j(w) = \mu_j' w - \tfrac{1}{2}\gamma w' \Sigma_j w. \tag{4}$$

This corresponds to what Rustem, Becker and Marty (2000) refer to as a model with rival return and rival risk scenarios. The worst case analysis implied by the minimum condition in (3) is called robust decision making. Theorem 1 provides the optimal portfolio that maximises the robust utility (all proofs are in Appendix A).

Theorem 1 *Consider a robust decision maker who obtains recommendations* (μ_j, Σ_j) *from* J *experts. Her optimal portfolio maximising* (3) *is*

$$w_R = \frac{1}{\gamma} S^{-1} m \tag{5}$$

where

$$S = \sum_{j=1}^{J} \lambda_j \Sigma_j,$$

$$m = \sum_{j=1}^{J} \lambda_j \mu_j,$$

and λ_j *are constants satisfying* $0 \leq \lambda_j \leq 1$ *and* $\sum \lambda_j = 1$.

The structure of the robust portfolio is similar to (2), but the expected returns m and covariance matrix S are a weighted average of the rival experts' estimates. Since the Lagrange multipliers λ_j have the properties of probabilities, the solution also resembles a Bayesian approach with λ_j as prior probabilities on the reliability of the experts. The crucial difference, however, is that the Lagrange multipliers are endogenously determined in the robust optimisation in contrast the exogenous Bayesian prior probabilities.[2] Positive Lagrange multipliers indicate the most pessimistic estimates in view of the optimal portfolio.

Theorem 1 is not a complete solution, since it does not provide explicit formulas for the Lagrange multipliers λ_j. For general (μ_j, Σ_j) the active Kuhn-Tucker constraints have to determined numerically. Efficient numerical solutions are, however,

[2] The solution in (5) can not be fully interpreted as the result of a Bayesian decision problem. Apart from the difference between interpreting λ_j as endogenous Lagrange multipliers versus exogenous prior probabilities, a Bayesian decision maker would use the predictive covariance matrix $\tilde{S} = \sum \lambda_j \left(\Sigma_j + (\mu_j - m)(\mu_j - m)'\right)$ leading to the optimal portfolio $w_B = \tilde{S}^{-1} m$. The dispersion in beliefs about the expected returns increases the unconditional variance in the Bayesian setting.

available from the operations research literature (cf. Ben-Tal and Nemirovski (2001)). The same numerical techniques are also effective under additional restrictions on the portfolio weights such as short-sale constraints.

For general true values (μ, Σ) the *loss function* expresses the utility loss that results from selecting a portfolio w which differs from the optimal portfolio w_0 based on the true parameters,

$$\begin{aligned} L(\mu, \Sigma|w) &= Q(w_0) - Q(w) \\ &= \frac{1}{2\gamma} (\mu - \gamma\Sigma w)' \Sigma^{-1} (\mu - \gamma\Sigma w) \end{aligned} \tag{6}$$

Since we can interpret $Q(w)$ as the risk adjusted expected return of a portfolio, the loss function measures the expected return difference between the optimal portfolio and a sub-optimal portfolio, both after risk adjustment. Expected loss is equal to zero if w is the optimal portfolio under the true parameters. More insight in the loss function is obtained by rewriting the equation in two different ways. Let $\tilde{\mu} = \gamma\Sigma w$ be the implied expected returns under which portfolio w would be optimal. Then

$$\begin{aligned} L(\mu, \Sigma|w) &= \frac{1}{2\gamma} (\mu - \tilde{\mu})' \Sigma^{-1} (\mu - \tilde{\mu}) \\ &= \frac{\gamma}{2} (w_0 - w)' \Sigma (w_0 - w) \end{aligned} \tag{7}$$

The loss is large if implicit expected returns $\tilde{\mu}$ are very different from the true expected returns μ. Equivalently, the second line of (7) states that the loss is large when the chosen portfolio w is far from the optimal portfolio w_0. For the interpretation of the loss function we view it as a function of the true parameters (μ, Σ) and compare the loss of each individual portfolio $w_j = \Sigma_j^{-1}\mu_j$ with the loss of the robust portfolio w_R. By construction the robust portfolio will never be the worst performing portfolio. For some (μ, Σ) combinations it can be the best, though.

If the investor strongly believes in the estimates of expert i, she would evaluate her portfolio decisions using the utility function $Q_j(w)$. Ex-ante a portfolio w promises a performance $Q_j(w)$. Ex-post, when the investor learns the true parameters, she

evaluates her portfolio using the true parameters and obtains utility $Q(w)$. We call the difference between promised and realised utility the disappointment of the investor,

$$D_j(w) = Q_j(w) - Q(w). \tag{8}$$

The disappointment can be both positive and negative. If expert j has been very pessimistic, while the actual parameters support much better investment opportunities, the investor is pleasantly surprised by the performance. For the robust investor the analogue of disappointment is

$$D_R(w) = Q_R(w) - Q(w) \tag{9}$$

If the true parameters (μ, Σ) are in the set of expert estimates Θ, then by construction $D_R(w) \leq 0$, since $Q_R(w)$ is the worst case evaluation of the portfolio. The ambiguity aversion of the robust investor implies that she will not be disappointed by the performance of her decisions.

Below we discuss two special instances for which we can analytically derive the optimal robust portfolio choice. The analytical form will help us understand some aspects of robust portfolio choice. The first example shows that a robust portfolio is not necessarily the most conservative portfolio. The second example delineates the range of parameter values for which the robust portfolio will outperform the portfolio choice based on each of the individual experts.

2.1 One risky asset, two experts

As a first example, consider an investor whose choice is limited to one risky asset and the risk free asset. In this case the expected return, the variance and the optimal investment in the risky assets are scalars. The investor does not know the true values of the expected return and variance. She obtains advice from two experts who recommend different estimates (μ_1, σ_1^2) and (μ_2, σ_2^2).

If one expert is more optimistic than the other, both in terms of higher expected returns $(\mu_1 > \mu_2)$ and lower variance $(\sigma_1^2 < \sigma_2^2)$, then the robust portfolio relies entirely

on the most prudent expert's recommendation and $w_R = \mu_2/\gamma\sigma_2^2$. The optimistic expert's advice does not affect the robust decision at all.

The interesting case is when neither expert is more optimistic. Expert 1 expects a larger return but also a higher variance than expert 2. The optimal portfolio is given in proposition 1.

Proposition 1 *Suppose* $\mu_1 > \mu_2 > 0$ *and* $\sigma_1^2 > \sigma_2^2$. *Let* w_j *be the optimal portfolio according to expert* j*'s advice. Define* $w_{12} > 0$ *as the portfolio weight for which* $Q_1(w) = Q_2(w)$. *Let* $c = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$. *A robust investor will have an optimal portfolio*

$$w_R = \begin{cases} w_2 & \text{if} \quad \mu_2 \leq 2c\mu_1, \\ w_{12} & \text{if} \quad 2c\mu_1 < \mu_2 < \frac{1}{2}\mu_1/(1-c), \\ w_1 & \text{if} \quad \mu_2 \geq \frac{1}{2}\mu_1/(1-c), \end{cases} \tag{10}$$

Figure 1 depicts this situation in terms of mean variance utility. Each panel shows the mean-variance utility anticipated by the experts. The objective of the robust investor is to maximize the minimum of the two. The optimal portfolio is located at the maximum of the robust objective function.

Insert figure 1 about here

The figure shows the three situations described in proposition 1. The interrelationship between the expected returns and variances recommended by the experts determine which situation applies. The robust portfolio is not necessarily the most conservative portfolio. In the middle panel the robust solution allocates a larger share to the risky asset than expert 1 would.

Insert figure 2 about here

Figure 2 depicts the loss function of the alternative portfolio for the situation depicted in the middle panel of figure 1. None of the three portfolios dominates the others for all parameter values. For small expected returns and large variance, the portfolio advocated by expert 1 would be best. For large expected returns and small variances, expert 2's recommendation would be best. For intermediate parameter values the robust portfolio would be best. Figure 2 also shows that the robust portfolio never has a larger loss than both 'expert' portfolios. Indeed, the robust portfolio is a strictly convex combination of the expert portfolios and the loss function (6) is strictly convex. This implies that the performance of the robust portfolio is strictly better than the worst performance alternative. Moreover, since the recommendations follow from alternative estimates of returns, both with their specific estimation error, these different beliefs possibly outline a set of parameter values which contain the true parameter values. The robust portfolio is likely to outperform either of the expert portfolios.

2.2 Multiple risky assets with equal covariance matrix

For the second example we assume that the expert opinions only differ in their expected return but not in the covariance matrix. As shown by Chopra and Ziemba (1993) uncertainty about the mean is usually more important than uncertainty about the covariance matrix. since variances and covariances can be estimated more precisely from high frequency data if necessary, the assumption of a known covariance matrix will often be relevant.

In this case, theorem 1 implies that the robust portfolio is a convex combination of the individual optimal portfolios. When the investor consults only two experts, the optimal portfolio allocation is summarized in proposition 2 below.

Proposition 2 *Consider a robust portfolio optimisation problem with multiple assets and two experts with different expected returns* μ_1 *and* μ_2*, but a known covariance*

matrix Σ. *Define* $\rho_{ij} = \mu_i' \Sigma^{-1} \mu_j$. *The optimal robust portfolio is*

$$w_R = \lambda w_1 + (1 - \lambda) w_2 \tag{11}$$

with $w_j = \frac{1}{\gamma} \Sigma^{-1} \mu_j$ *and*

$$\lambda = \begin{cases} 0 & \text{if} \quad \rho_{22} \leq \rho_{12} \\ \frac{\rho_{22} - \rho_{12}}{(\rho_{11} - \rho_{12}) + (\rho_{22} - \rho_{12})} & \text{if} \quad \min(\rho_{11}, \rho_{22}) > \rho_{12} \\ 1 & \text{if} \quad \rho_{11} \leq \rho_{12}. \end{cases} \tag{12}$$

The optimal solution is characterized by the Sharpe ratios (ρ_{11} and ρ_{22}) anticipated by the experts and the cross sectional correlation between the expert's portfolio recommendations ($\frac{\rho_{12}^2}{\rho_{11}\rho_{22}}$).[3] A lower correlation implies more disparate expert recommendations and more incentive for the robust investor to combine the recommendations. The robust investor tends to downweight the advice of an expert who promises a high Sharpe ratio.

For the performance we compute the loss function (6) of the robust portfolio and compare it to the portfolios that would have been recommended by the individual experts.

Proposition 3 *Assume* $\rho_{12} < \min(\rho_{11}, \rho_{22})$. *The robust portfolio* w_R *dominates the portfolios* w_1 *and* w_2 *for a range of true expected returns* μ_0 *satisfying*

$$\tfrac{1}{2}(\rho_{12} - \rho_{11}) < \rho_{02} - \rho_{01} < \tfrac{1}{2}(\rho_{22} - \rho_{12}), \tag{13}$$

where $\rho_{0j} = \mu_0' \Sigma^{-1} \mu_j$ *(*$j = 1, 2$*).*

The robust portfolio has lower loss than either of the two expert portfolios for an important set of returns illustrated in figure 3. This set contains the expected returns that lie (geometrically delimited by parallel planes) between the expert's estimates for the expected returns. Outside this set, the robust portfolio still features a better

[3] We call the ratio $\rho_{12}^2/\rho_{11}\rho_{22}$ the cross sectional squared correlation, since it can be computed as the R^2 of a GLS regression of μ_i on μ_j (or vice versa) using Σ as the GLS covariance matrix.

performance than the worst of the alternatives. Condition (13) states that whenever the difference between the two views has a low enough correlation with the true expected returns, then the robust portfolio will outperform the portfolios that are optimal under a single model. In the ideal case, the difference of opinion between the experts is orthogonal to the true parameters.

3 Expert posterior views

So far we have not been specific about how the experts arrive at their estimates (μ_j, Σ_j). From here on we will assume that all of them use the same sample data, but may have different prior views on the parameters. The experts observe a sample of T observations, which they combine with their individual prior views to obtain the posterior expectation of the mean and covariance matrix. Each expert reports his posterior expectations $(\hat{\mu}_j, \hat{\Sigma}_j)$ to the investor. We use the notation $\hat{\mu}_j$ instead of μ_j to express the dependence of the posterior mean on the sample information Y, i.e.

$$\hat{\mu}_j = \mu_j(Y) = \int_{\mu} p_j(\mu|Y)\mu \, d\mu, \tag{14}$$

where $p_j(\mu|Y)$ is the posterior of μ using the prior of expert j. $\hat{\Sigma}_j$ is defined analogously. Ex-ante the investor does not know how the experts arrived at their estimates $(\hat{\mu}_j, \hat{\Sigma}_j)$. She neither observes the prior nor the full historical sample. As a result she can not assess the precision of the estimates.

Sampling variation in the posterior expectations leads to sampling variation in the utility functions $Q_j(w)$ and thus in the optimal portfolios $\hat{w}_j$. Consequently also the loss and disappointment of the optimal portfolios become subject to sampling variation. The corresponding expectations follow from integrating over the sampling variation. Let $p(Y|\mu, \Sigma)$ denote the true density, and let $\hat{w} = w(Y)$ be the robust portfolio rule depending on the set $\hat{\Theta} = \{(\hat{\mu}_j, \hat{\Sigma}_j); j = 1, \ldots, J\}$. The expected loss is defined as

$$\mathsf{E}\left[L(\mu, \Sigma|\hat{w})\right] = \int_{Y} L(\mu, \Sigma|w(Y)) \; p(Y|\mu, \Sigma) \; dY \tag{15}$$

The expected disappointment is

$$\mathsf{E}\left[D_j(\hat{w})\right] = \int_Y D_j(w(Y)|\hat{\mu}_j, \hat{\Sigma}_j)\; p(Y|\mu, \Sigma)\; dY \tag{16}$$

Each portfolio strategy corresponds to a specific estimator of the expected returns and covariances. Adding prior information reduces the sampling variance of these estimators, but introduces bias (or misspecification). The performance of the portfolio based on the recommendation of an individual expert will exhibit the usual tradeoff between bias and variance. A tight prior will lead to low variation in the estimated parameters. If, however, the prior is far from the true values, it will at the same time have a large bias.

Analytical expressions for the expected loss are only available in special cases. For the versatile priors we develop in section 4 we need bootstrap simulations to compute the expected loss and disappointment. As in section 2 we use a few specific examples to gain understanding of important aspects of the prior. The examples build on the example in section 2.2. They consider an investor consulting two experts, who agree on the covariance matrix. The examples illustrate the trade-off between bias and estimation error and the way the robust approach balances the two.

3.1 Known covariance matrix

Both experts know the covariance matrix Σ. The first expert has an uninformative prior $p(\mu) \propto 1$ for the mean. His posterior expectation of the mean is the sample average $\hat{\mu}_1 = \bar{y} = \frac{1}{T}\sum y_t$.

The second expert consulted by the investor relies on the CAPM. In his view expected returns are linearly related to beta times a market risk premium. In this example we assume that he knows the N-vector of factor loadings β, but that he does not have information on the market risk premium ν,

$$\mu \sim \mathsf{N}\left(\beta\nu, \Sigma/k\right),$$

$$p(\nu) \propto 1.$$

The parameter k measures the prior confidence of the expert in the CAPM. The posterior shrinks the sample mean of each asset towards the CAPM implied expected returns $\beta\bar{\bar{y}}$,

$$\hat{\mu}_2 = a\bar{y} + (1-a)\beta\bar{\bar{y}}, \tag{17}$$

where $\bar{\bar{y}} = (\beta'\Sigma^{-1}\beta)^{-1}(\beta'\Sigma^{-1}\bar{y})$ is the GLS estimator of ν, and $a = 1/(1+k/T)$ is the shrinkage factor.[4] If $k = 0$, the prior reduces to the uninformative prior. The optimal portfolio follows from substituting $\hat{\mu}_2$ in (2).

The robust investor consults both experts and chooses the optimal robust portfolio. Proposition 4 states that a more informative prior dominates an uninformative prior in the robust portfolio.

Proposition 4 *Consider two experts who agree on the covariance matrix Σ, but report different expected returns. Expert 1 has an uninformative prior; expert 2 has an informative prior centered on the CAPM. The robust portfolio follows the advice of expert 2 and is thus equal to $\hat{w}_R = \frac{1}{\gamma}\Sigma^{-1}\hat{\mu}_2$.*

When $k \to \infty$, the expert has a dogmatic belief in the CAPM. In this case, expert 2 imposes the asset pricing model's restrictions on the expected returns. Proposition 4 then implies that the robust portfolio will be based on the CAPM expected returns.[5] In other words, if the expert models are nested, the robust investor will follow the advice from the most restricted model. The CAPM would continue to dominate the portfolio, no matter how much data evidence would be available against it. Since such a lack of learning from the data would be a strong disadvantage of robust portfolio choice, we assume that the experts are disciplined by data evidence, meaning that k is finite.

[4] For a derivation of this result see, for example, the appendix of Jorion (1986). Instead of β Jorion (1986) uses the vector of ones ι and shrinks towards a common mean.

[5] The result holds more generally for any number of factors. If expert 2's advice is formed by a projection of the sample means on the factor space with $(N \times K)$ matrix of factor loadings B, i.e. $\mu_2 = B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\bar{y}$, then the optimal robust portfolio coincides with the portfolio advice of expert 2. The proof is a straightforward application of proposition 2.

The expected loss of the two portfolios is very different. For the uninformative prior the loss function becomes[6]

$$L(\mu|\hat{w}_1) = \frac{1}{2\gamma}(\mu - \hat{\mu}_1)\Sigma^{-1}(\mu - \hat{\mu}_1). \tag{18}$$

The expected loss of this portfolio is obtained by integrating over the sampling variation using $\bar{y} \sim \mathsf{N}(\mu, \Sigma/T)$,

$$\mathsf{E}[L(\mu|\hat{w}_1)] = \frac{1}{2\gamma}\frac{N}{T}. \tag{19}$$

Even though the expert's model is correctly specified, the expected loss is not equal to zero due to estimation uncertainty. Estimation uncertainty increases with the investment universe N and decreases with the number of observations T.

If only a small number of observations is available, the magnitude of the loss is tremendous. A mean-variance portfolios of $N = 25$ assets with expected returns based on $T = 60$ monthly observations and risk aversion $\gamma = 5$ (all typical numbers in portfolio management) leads to an expected loss of $\frac{25}{600} = 4.1\%$ per month in terms of certainty equivalent loss. This explains why unconstrained mean-variance portfolios exhibit such poor performance. The loss further increases if estimation error in Σ is added. Using the sample covariance matrix $\hat{\Sigma}$, Kan and Zhou (2006) obtain an analytic expression for the loss. They show that uncertainty about Σ becomes important, relative to estimation error in μ, when $\frac{N}{T}$ is large. For the combination $N = 25$ and $T = 60$, the expected loss increases by a factor six, when both μ and Σ are unknown.[7]

Proposition 5 states the expected loss and disappointment for the robust (CAPM) portfolio.

Proposition 5 *The optimal portfolio based on the CAPM prior* (17) *has expected*

[6] When we assume that the covariance matrix is known we omit Σ as an argument in the loss function.

[7] See table 1 in Kan and Zhou (2006).

loss and disappointment

$$\mathsf{E}\left[L(\mu|\hat{w}_R)\right] = \tfrac{1}{2\gamma}\left((1-a)^2\mathsf{Sh}^2(1-R^2) + \tfrac{1}{T}(a^2N+1-a^2)\right) \tag{20a}$$

$$\mathsf{E}\left[D_R(\hat{w}_R)\right] = \tfrac{1}{\gamma}\left(-a(1-a)\mathsf{Sh}^2(1-R^2) + \tfrac{1}{T}(a^2N+1-a^2)\right) \tag{20b}$$

where

$$\mathsf{Sh}^2 = \mu'\Sigma^{-1}\mu$$

denotes the maximal Sharpe ratio under the true parameters (μ, Σ)*, and*

$$R^2 = \frac{(\beta'\Sigma^{-1}\mu)^2}{(\beta'\Sigma^{-1}\beta)(\mu\Sigma^{-1}\mu)}$$

is a measure for the correlation between β *and* μ *(the* R^2 *of the GLS regression of* μ *on the factor loadings* β*).*

The expected loss has two components: bias and variance. Since $a < 1$, the variance term $\frac{1}{T}(a^2N+1-a^2)$ is strictly less than for the unconstrained mean-variance portfolio. The bias term depends on the true investment opportunities indicated by the maximal Sharpe ratio (Sh) and the misspecification of the prior $(1-R^2)$. If the pricing model is well specified, the expected returns can be perfectly explained by the factor loadings and $R^2 = 1$. In this case there is no bias and the prior improves the estimator of the expected returns and reduces the loss. The bigger the misalignment between the factor loadings and the true expected returns, the larger the bias term in (20a). When $T \to \infty$, the shrinkage factor a converges to 1 and $\hat{\mu}_2$ goes to the unrestricted sample mean $\bar{y}$. As the confidence of the expert in the CAPM increases, meaning larger k, the variance in (20a) decreases, whereas the bias due to misspecification increases.

Disappointment exhibits a different trade-off between bias and variance. The robust investor is prudent and holds - by definition - the lowest expectations among all experts. Imposing structure tempers the investor's expectations. This leads to the negative bias in (20b). Ex-post the true value of μ will be more likely to imply better investment opportunities than anticipated by the robust investor. If the variance is

small, the disappointment could even be negative, meaning that the investor ex-post finds herself performing better than expected.

In contrast the disappointment of the uninformed mean-variance portfolio $\hat{w}_1$ is just twice the expected loss: $\mathsf{E}[D(\hat{w}_1)] = 2\mathsf{E}[L(\mu|\hat{w}_1)]$. The investor holds great expectations and invests expediently only to discover ex-post that the true investment opportunities are very moderate. Her opportunistic but sub-optimal portfolio choice further aggravates her disappointment. Indeed, the portfolio weights are typically extreme to exploit the seemingly great opportunities, but under the realised moments of the returns the weights are far from optimal.

3.2 Conflicting views

As a variation on the previous example suppose that the two experts use conflicting asset pricing models. This could happen in the case of international diversification, when one expert estimates expected returns using a local CAPM with beta's relative to a domestic market index, while the other expert uses the international CAPM with beta's relative to a global market index. Formally, expert j ($j = 1, 2$) assumes that expected returns are proportional to the N-vector β_j. As before they agree on the covariance matrix Σ. For simplicity, let us assume that they have the same prior confidence k in their respective asset pricing models. The posterior expectations are as in (17)

$$\hat{\mu}_j = a\bar{y} + (1-a)\beta_j\bar{\bar{y}}_j, \tag{21}$$

where $\bar{\bar{y}}_j = (\beta_j'\Sigma^{-1}\beta_j)^{-1}\beta_j'\Sigma^{-1}\bar{y}$ are the estimates of the market premium. The divergence of the two experts can be measured by the cross sectional correlation of their prior expected returns

$$R_{12}^2 = \frac{(\beta_1'\Sigma^{-1}\beta_2)^2}{(\beta_1'\Sigma^{-1}\beta_1)(\beta_2\Sigma^{-1}\beta_2)} \tag{22}$$

The model does not have an analytical solution for the robust portfolio. From proposition 2 we know that the robust portfolio is a convex combination of the two individual

portfolios $\hat{w}_1$ and $\hat{w}_2$. In expectation, the robust solution does not coincide with the optimal portfolio of either of expert since

$$\begin{aligned} \mathsf{E}\left[Q_1(\hat{w}_1) - Q_2(\hat{w}_1)\right] &= \mathsf{E}\left[\hat{\mu}_1'\hat{w}_1 - \hat{\mu}_2'\hat{w}_1\right] \\ &= \frac{(1-a)^2}{T\gamma}(1 - R_{12}^2) > 0 \end{aligned} \tag{23}$$

(see Appendix A for a proof). Inequality (23) states that at the local optimum for expert 1 the robust utility value is expected to be the utility of expert 2 (and vice versa), in which case the local optimum of expert 1 can not be the robust optimum. The probability of an internal optimum (a strictly convex combination $\hat{w}_R = \lambda\hat{w}_1 + (1-\lambda)\hat{w}_2$) increases with the divergence of their opinions on expected returns. For a precise quantitative evaluation of the performance we need simulation methods, which will be introduced in the next section, when we also introduce a more general class of priors.

4 Multiple Priors

For an empirically relevant setting we move to a more general prior that considers both expected returns and, beta's and the covariance matrix. We assume that each expert maintains a subjective view on the distribution of μ and Σ. The expert combines his prior with the sample information to arrive at posterior estimates of the mean and covariance matrix. These posterior moments are the expert's advice to the investor.

We use the prior structure suggested by Pàstor and Stambaugh (2000). To estimate the moments of the excess returns the experts consider the linear factor model

$$y_t = \alpha + Bx_t + u_t, \tag{24}$$

where y_t is the N-vector of excess returns considered by the experts and the investor; x_t is an M-vector of observed factors; and u_t is an error term with mean zero and $(N \times N)$ covariance matrix D. Both the vector of abnormal returns α, the factor

loadings B and the error covariance matrix D are unknown parameters. The factors themselves follow the process

$$x_t = \nu + e_t, \tag{25}$$

with ν an M-vector of factor means and e_t a vector of shocks with mean zero, covariance matrix Ψ, and uncorrelated with u_t. The prior on the set of parameters $(\alpha, B, \nu, D, \Psi)$ is taken from Pàstor and Stambaugh (2000). For (B, ν, Ψ) all experts use an uninformative prior. Priors for α and D differ across investors. Still following Pàstor and Stambaugh (2000) we assume

$$\begin{aligned} D^{-1} &\sim W(df, H^{-1}) \\ \alpha \mid D &\sim N(0, \frac{\sigma_\alpha^2}{s^2} D) \end{aligned} \tag{26}$$

where $W(df, \cdot)$ denotes the Wishart distribution with df degrees of freedom. The covariance matrix prior is calibrated by setting $H = s^2(df - N - 1)I_N$. Following an empirical Bayes approach the scaling parameter s^2 is set equal to the average of the diagonal elements of the sample estimate of D. Observe that the prior for D approximately contains the equivalent information of df observations and the expectation of D equals $s^2 I_N$. The larger df the more confidence an expert has in the factor structure for the covariance matrix.

All priors are centered on the underlying factor pricing model with $\alpha = 0$. The strength of the expert's belief in the model is quantified by σ_α. A dogmatic follower of the asset pricing conditions will set $\sigma_\alpha = 0$. An expert with serious doubts about expected returns in the factor model will set σ_α large. We assume that all experts have non-dogmatic priors, i.e. $\sigma_\alpha^2 > 0$. If the number of observations in the sample grows to infinity, all posteriors will converge to the same true values and all experts will agree on the true parameters. This also implies that the robust investor's decision converges to the truly optimal portfolio.

The experts combine their model prior with sample information to obtain a posterior return model. This posterior is their advice to the investor. Appendix B computes the posterior, mostly following Pàstor and Stambaugh (2000). An analytical solution

for the robust portfolio is not available, but the numerical solution can be computed extremely fast.

5 Bootstrap Simulations

Robust portfolios have the potential to outperform naive mean-variance portfolios. Whether they do so in empirically relevant settings must be tested. We will compare expected loss for different implementations of mean-variance portfolios: naive unrestricted portfolios, robust portfolios, and portfolios subject to short-selling constraints. For the simulations we will use parameters obtained from empirical data. We estimate μ and Σ from a long historical data set and denote these parameters as the true values μ_0 and Σ_0. Conditional on the true values μ_0 and Σ_0 the optimal mean-variance portfolio w_0 has utility

$$Q_0 = \frac{1}{2\gamma}\mu_0'\Sigma_0^{-1}\mu_0. \tag{27}$$

Next we repeatedly bootstrap random samples of length T from the historical data set. The bootstrap experiment accommodates the static form of the return models. Any time series properties of the data are eliminated by random sampling of N-vectors of excess returns from the full dataset. For each bootstrap sample Y_T^k ($k = 1, \ldots, K$) the experts compute their posterior means $\hat{\mu}_j^k$ and $\hat{\Sigma}_j^k$. The investor receives these estimates and constructs the optimal portfolios $\hat{w}_j^k$ conditional on each experts opinion, and also the optimal robust portfolio $\hat{w}_R^k$.

The ex-ante performance of a portfolio is summarised by averages over K bootstrap samples. For portfolios w_j ($j = 1, \ldots, J$ and $j = R$) we compute the ex-ante expected excess return and variance of the portfolio returns as

$$E_j = \frac{1}{K}\sum_k (\hat{w}_j^k)'\hat{\mu}_j^k, \tag{28a}$$

$$V_j^2 = \frac{1}{K}\sum_k (\hat{w}_j^k)'\hat{\Sigma}_j^k\hat{w}_j^k, \tag{28b}$$

and the ex-ante expected utility as

$$Q_j = E_j - \tfrac{1}{2}\gamma V_j^2. \tag{29}$$

For the robust portfolio, $j = R$, the expected return and variance are defined through the optimality conditions in theorem 1,

$$\hat{\mu}_R^k = \sum_j \lambda_j^k \hat{\mu}_j^k \tag{30a}$$

$$\hat{\Sigma}_R^k = \sum_j \lambda_j^k \hat{\Sigma}_j^k \tag{30b}$$

The corresponding ex-post performance statistics of all portfolios are

$$E_{0j} = \frac{1}{K}\sum_k (\hat{w}_j^k)'\mu_0, \tag{31a}$$

$$V_{0j}^2 = \frac{1}{K}\sum_k (\hat{w}_j^k)'\Sigma_0 \hat{w}_j^k. \tag{31b}$$

These are the out-of-sample characteristics of portfolios formed using the expert estimates. Using the expected excess return and the variance we construct the expected utility, expected loss and expected disappointment as

$$Q_{0j} = E_{0j} - \tfrac{1}{2}\gamma V_{0j}^2 \tag{32a}$$

$$L_{0j} = Q_0 - Q_{0j} \tag{32b}$$

$$D_{0j} = Q_j - Q_{0j} \tag{32c}$$

The subscript 0 in these statistics indicates evaluation under the null parameters (μ_0, Σ_0).

We are also interested in the portfolio composition itself. How many different assets does the investor hold? How much leverage is involved? How diversified is the portfolio? How stable is the portfolio? Which expert models are binding in the robust portfolio? For that purpose we examine the following properties of the optimal

portfolios:

$$\bar{w}_j = \frac{1}{K}\sum_k \iota' \hat{w}_j^k \tag{33a}$$

$$s^2(w_j) = \frac{1}{K}\sum_k (\hat{w}_j^k)' \hat{w}_j^k - \bar{w}_j^2 \tag{33b}$$

The sum of the portfolio weights tells us how much of wealth is allocated to risky assets. The cross-sectional variance signals large weights. The bigger $s^2(w_j)$ the more extreme (some of) the weights in the portfolios. We measure stability by the average turnover, i.e. change in asset investments, over the scenarios.

$$T_j = \frac{1}{K}\sum_k |\hat{w}_j^{k+1} - \hat{w}_j^k| \tag{34}$$

In the empirical analysis we also report on alternative portfolio strategies. The first alternative is to include short-sell constraints in the portfolio choice as suggested by Jagannathan and Ma (2003). Since short-sell constraints avoid the extreme weights that are typical in naive mean-variance portfolios, the sampling error in these portfolios is much reduced. Jagannathan and Ma (2003) show that short-sell constraints lead to much better small sample performance even if the constraints are false.

The second alternative is an equally weighted portfolio of all N assets. This is the benchmark advocated by DeMiguel et al. (2006). It does not require any parameter estimation in-sample. We can thus directly compute the performance of this portfolio under the true data-generating process.

The third alternative considers the effect of riskless borrowing and lending. In the mean-variance problem (1) the investor can reduce or increase risk by taking positions in the riskfree asset. In many portfolio allocation problems the construction of a portfolio of risky assets and the position in the riskless asset are separated. We therefore also compute the optimal portfolio of risky assets by imposing the constraint $\iota' w = 1$ on all portfolios.

6 Empirical Results

In the empirical analysis we consider two different datasets: the 25 Fama and French portfolios sorted on size and book-to-market and 81 European portfolios sorted both by country as well as by size and book-to-market. For the first dataset we know that the Fama and French 3-factor model provides a good fit for the expected returns, and is thus expected to perform well in the simulations. For the European data we expect that the robust portfolio will perform very well for two reasons: (*i*) standard asset pricing models do not fit the cross-section very well; and (*ii*) the large number of assets leads to erratic behavior of the naive mean-variance portfolio.

6.1 25 Fama-French Portfolios

The dataset consists of 480 monthly observations from July 1963 to December 2002 on the value weighted excess market returns of 25 value weighted portfolios in the US independently sorted on size and book-to-market quintiles. The data are available from the website of Professor Ken French.[8] These data are among the most widely analysed in the asset pricing literature.[9] In the extensive literature it has been found that the CAPM can not explain the cross-sectional differences in expected returns. Much better results are obtained for the Fama-French 3-factor, but even for this model the portfolio of small value stocks remains anomalous.

We consider two return models: the CAPM and the Fama-French three factor model (FF3). For both models we consider two priors: one with a strong belief in the asset pricing implications and another with a weak prior on expected returns. With two models and two priors for each model we have four rival models. We vary the expert's beliefs by changing σ_α: $\sigma_\alpha = 0.1\%$ (tight prior) and $\sigma_\alpha = 4\%$ (loose prior). We set the prior information on the covariance matrix at $df = 30$.

[8] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

[9] See for example Fama and French (1996), Davis, Fama and French (2000) and the textbook treatment in Cochrane (2001).

Insert table 1 about here.

For the bootstrap simulations we use the full sample moments as the true values. Table 1 provides summary statistics in terms of the maximal achievable portfolio performance for the different models. By definition the optimal portfolio w_0 conditional on the true values (μ_0, Σ_0) provides the maximal attainable performance. Expected return for the optimal portfolio is 4.6% per month with a standard deviation of 9.6%. The portfolio is leveraged with total investment in the risky assets of 164% of invested wealth. Optimal portfolio weights under the true model parameters are extreme. The best indication of the excessive weights is the cross-sectional standard deviation of the portfolio weights, which is equal to 123%.

The asset pricing models perform poorly in comparison. Both the CAPM and the Fama-French models impose restrictions on (μ_0, Σ_0). An investor with dogmatic beliefs in these models will suffer a large expected utility loss. Ex-ante the best mean-variance portfolio based on the Fama-French 3-factor model only promises an expected return of 1.1% with a standard deviation of 4.7%. The certainty equivalent utility loss is 1.7% per month. Ex-post the realised utility of the Fama-French investor will be a fraction above the promised ex-ante utility, reflecting that the investment opportunities under the true values are slightly better than anticipated by the Fama-French restrictions.

The CAPM is heavily misspecified. It only obtains expected returns of 0.2% with a standard deviation of 1.8%. The CAPM investor suffers a utility loss of 2.2% per month. On the other hand the portfolio of the CAPM investor is very well diversified over all 25 portfolios. Both the Fama-French and CAPM portfolios would not use leverage, but instead have positive investments in the riskfree asset. Under the true values the robust portfolio coincides with the CAPM, being the most restricted model.

The performance of an equally weighted portfolio of the risky assets provides a more realistic benchmark than the unattainable optimal portfolio given the true values. Since the equally weighted portfolio is not based on parameter estimates, the

investor does not have any ex-ante expectations of this portfolio strategy. Ex-post the strategy is only slightly worse than the best performance that a CAPM investor would be able to get: given the true values the CAPM has an advantage of less than 0.2% utility difference per month.

The true values thus provide an ideal setting for comparing the performance of alternative portfolios. Under the true values the odds are highly in favour of an unrestricted model, which can fully exploit the anomalous return of small value stocks in the Fama-French data. Even if uninformed mean-variance optimisation suffers from error maximisation, then the Fama-French 3-factor model still provides a formidable challenge for the robust portfolio. The 3-factor model provides much better performance than the CAPM, but since the CAPM is much more restricted, the robust portfolio will often be close to the CAPM portfolio. If the data are not informative, the robust portfolio would be close to the portfolio implied by a tight CAPM prior, which only has a narrow margin against the equally weighted benchmark. If the robust portfolio performs well in this setting, it does so against strong competition.

We compare the performance of the different portfolios if the experts only observe a sample of $T = 60$ monthly observations. Such a five year sample is typical in many portfolio management applications in practice. We bootstrap 1000 samples of T random observations y_t from the full dataset.

Insert table 2 about here.

Table 2 shows the performance of the different portfolios. The first thing to note is that the ranking of portfolios is completely reversed. The robust portfolio performs best, closely followed by the portfolios based on the tight CAPM and Fama-French models. As expected the portfolio based on the uninformative prior performs deplorably. The sampling error is overwhelming with only five years of data. The data are not informative enough to accurately estimate the individual asset outperformance (i.e. the α's).

The robust portfolio is strongly related to the portfolio based on the tight CAPM prior. In 99% of the simulations the tight CAPM prior is binding ($\lambda_j > 0$ in determining the robust portfolio). In 68% of the cases the robust portfolio is a mixture of the tight CAPM and the tight Fama-French 3-factor model. The robust portfolio features small cross-sectional variation in the 25 portfolio weights (4%) compared to the astonishing 607% for the unrestricted portfolio. Portfolio weights of the robust portfolio are well behaved: no leverage versus a 311% investment in risky assets for the unstructured portfolio. The performance of the tight CAPM portfolio is already close to its maximal performance and on par with the equally weighted benchmark.

The robust portfolio is also exceptionally stable. The absolute turnover over the K simulations is only 4.4% (1.1/25) per asset. In contrast, even the FF3 portfolios are updated more drastically with on average a 18.4% change in investment per asset. The robust portfolio is thus also robust in the sense that sampling variation in the inputs only has a modest effect in the decision variables.

How does the robust portfolio fare against short-selling constraints? From table 1 we know that under the null the optimal portfolio has many negative weights, so the constraint is incorrect. Indeed, imposing the restrictions on the truly optimal portfolio reduces the performance from 2.31% to 0.34%. But in the practical setting the portfolio performance of the unconstrained portfolio improves dramatically. The expected loss falls from 43% to 2.5% per month, which is almost the same level as achieved by the robust portfolio. The main drawback of the non-negativity constraint is that it reduces diversification. On average the investor holds only two of the 25 assets. In contrast the robust portfolio remains well diversified.

Part of the performance differences are driven by different degrees of leverage between the portfolios. From table 1 we know that the optimal investment in the riskfree asset is -64%, whereas the restricted models have positive investments in the riskfree asset and the equally weighted benchmark is in the middle. To eliminate the effect of leverage we include the extra portfolio constraint $\iota' w = 1$ and repeat the

analysis with this extra portfolio constraint. Table 3 shows the results. Again the robust portfolio performs best, even better than before. Now that both the benchmark and the robust portfolio are fully invested in risky assets, the robust portfolio also outperforms the equally weighted benchmark. The forced 100% investment brings the portfolio closer to the truly optimal portfolio.

Insert table 3 about here.

Insert figure 4 about here

With 60 observations, the outperformance of the robust portfolio is large, but still relative. Since the expected utility of the robust portfolio is negative, the investor would be better off just investing in the risk-free asset. Figure 4 shows the portfolio performance as the number of observations increases steadily. It illustrates the trade-off between estimation uncertainty and model structuring. If uncertainty is high, i.e. if only a few observations are available, structuring is desirable. But as uncertainty decreases, the bias due to model structuring is more persistent than estimation error and leads to a larger loss. Doubling the sample size to $T = 120$ supports more profitable active strategies. In that case the robust portfolio's performance becomes slightly positive. Eventually, both estimation uncertainty and the bias due to the non-dogmatic models disappears and the truly optimal portfolio results.

6.2 European dataset

The second data set consists of 81 portfolios from 9 countries: the domestic market portfolio, four value portfolios and four growth portfolios for each of the countries Belgium, France, Germany, Italy, Netherlands, Spain, Sweden, Switzerland and Great-Britain. The complete data consists of 225 monthly observations of excess returns and risk factors from January 1975 to December 2001. The US Dollar denominated

excess returns and factors are available from the website of Professor Ken French.[10] We will study the portfolio choice from the perspective of a British investor. All returns and factors are converted to British Pounds. The British riskfree rate is the LIBOR available from Datastream. The exchange rates have been obtained from the Italian Foreign Exchange Office (UIC).[11]

Four rival structured models describe return behavior: (*i*) the International CAPM (ICAPM) with a global market portfolio; (*ii*) the international Fama-French model (IFF) with the global market return and a value factor; (*iii*) the international CAPM with the excess return on the global market portfolio plus the German Mark (Euro) - Pound and US Dollar - Pound exchange rates as factors (FX); and (*iv*) a consolidated model (ALL) that contains all factors. For the prior we use $\sigma_\alpha = 0.1\%$ and $df = 82$.

Table 4 describes the European dataset in terms of optimal portfolio performance. The optimal portfolio based on the full sample moments features a monthly return of 7.3%, a standard deviation of 12.1% resulting in a mean variance performance of 3.7%. The table also shows that the CAPM is most conservative (expected performance 0.07%), followed by the ICAPM-FX, Fama-French model and the overall model (expected performance 0.32%).

Insert table 4 about here.

The European dataset contains 81 portfolios for investment. Typically a larger investment set also implies more estimation error. To obtain meaningful estimation results, we use a sample size of $T = 150$ observations for the bootstrap experiment. Table 5 reports the average performance over 1000 samples.

As before the unstructured portfolio has deplorable performance. It performs worse than in the previous setting with an investment set of 25 Fama-French portfolios. The portfolio choice problem with 81 portfolios and 150 observations is riskier

[10] See footnote 8. The data are from the files F-F International Countries with the specifier All 4 Data Items Not Reqd.

[11] `http://www.uic.it/`

than the portfolio choice for 25 Fama-French portfolios with 60 observations. Indeed, the *potential* profitability of this dataset is higher (3.7%), but the *realized* performance is considerably lower (-245%).

The structured return models are less prone to the increased estimation uncertainty. The most restrictive model, the ICAPM, performs best. Its performance (0.07%) is close to the performance of the optimal CAPM (0.20%, cf. 4). The CAPM manages to exploit the benefits of a larger investment set.

Insert table 5 about here.

The robust portfolio relies on the ICAPM and, for 58% of the simulations, also on the ICAPM-FX. The dispersion in estimates among the four models is enough to avoid extreme weights in the robust portfolio. Its performance is marginally better than the CAPM. The robust portfolio invests around 68% in the riskfree asset, much more than would be implied by each of the individual models. Even with these heavy investments in the riskfree rate, the robust portfolio still just manages to outperform the equally weighted benchmark.

Nonnegativity restrictions lead to a spectacular improvement in the performance of the unstructured portfolio. As for the 25 Fama-French portfolios, its performance is on par with the robust portfolio. Again we observe the reduction in diversification associated with the non-negativity restrictions. The robust portfolio remains best diversified. Without the constraints, the robust portfolio includes all assets.

Insert figure 5 about here

As for the US dataset, the performance of all models improves by increasing the sample size. In figure 5 all models except the uninformative prior show similar performance when the sample size doubles. The unrestricted mean-variance portfolio is still much worse than each of the other models for 500 observations.

7 Conclusion

We have studied the performance of robust mean-variance portfolios. These portfolios use an endogenously determined weighted average of the expected returns and the covariance matrices of J rival experts. A defining characteristic of robust portfolios is that they are always at least as good as the worst-case of the individual expert recommendations. In many cases they can be expected to outperform the individual expert portfolios.

Our main results pertain to empirical tests of robust portfolio choice when experts have different priors that are centered on alternative asset pricing models and deviations from these models. Priors are formulated both on expected returns as well as a particular factor structure in the covariance matrix. In practically relevant situations with limited sample size relative to the number of assets, our simulations show that the robust portfolio indeed performs better than all individual expert portfolios based on a single prior. The robust portfolio attains the highest expected utility out-of-sample, and it leads to stable and well-diversified portfolios. Even in small samples the robust portfolio is able produce stable portfolios with a performance that is on par (and sometimes even better) than an equally weighted benchmark portfolio.

The simulations are consistent with the theoretical predictions. From the theoretical analysis we concluded that robust portfolios have attractive properties. First, uninformative priors do not receive weight in the robust moments. Second, if all experts center their views on the same asset pricing model, but with more or less informative priors, then the robust investor follows the advice of the most informative prior. Third, when expert views are conflicting, the investor combines the different inputs, thus endogenously moving to a middle ground. These properties are related to the ambiguity aversion of the robust investor, leading her to avoid much of the estimation noise. On the other hand, with strong and undisputed data evidence of investment opportunities, the robust investor does exploit the opportunities, and can thus do better than a naive investor following a simple benchmark portfolio.

Appendix A Proofs

Proof of theorem 1

A formal representation of the robust portfolio problem is

$$\max Q_R, \tag{A1a}$$

$$\text{subject to} \quad Q_j(w) - Q_R \leq 0, \qquad (j = 1, \ldots, J). \tag{A1b}$$

Kuhn-Tucker conditions for the optimal portfolio are

$$1 - \sum_j \lambda_j = 0, \tag{A2a}$$

$$\sum_j \lambda_j Q_j'(w) = 0, \tag{A2b}$$

$$\lambda_j(Q_R - Q_j(w)) = 0, \qquad (j = 1, \ldots, J), \tag{A2c}$$

$$\lambda_j \geq 0, \qquad (j = 1, \ldots, J), \tag{A2d}$$

with $\lambda_j \geq 0$ a set of Lagrange multipliers and $Q_j'(w) = \mu_j - \gamma \Sigma_j w$ the first order derivative of the objective function with expert j's parameters. From (A2a) and (A2d) we deduce that the Lagrange multipliers must satisfy $0 \leq \lambda_j \leq 1$. From (A2b) and the linearity of $Q_j'(w)$ we immediately obtain the result stated in the theorem.

Proof of proposition 1

With two experts we only need to check three cases of combinations of active constraints in the Kuhn-Tucker conditions in the proof of theorem 1. We will show that these correspond to the three possible portfolios in (10). Under the assumption $\mu_1 > \mu_2$, we have $Q_1'(0) = \mu_1 > \mu_2 = Q_2'(0)$, implying that $Q_R(w) = \min(Q_1(w), Q_2(w)) = Q_2(w)$ for small w. Since both objective functions are quadratic, the difference between them is also quadratic, and they intersect at only two points: $w = 0$ and $w = w_{12}$. Therefore

$$Q_R(w) = \begin{cases} Q_2(w) & \text{if } \ 0 < w < w_{12}, \\ Q_1(w) & \text{if } \ w \geq w_{12}. \end{cases} \tag{A3}$$

The maximum of $Q_R(w)$ is either at one of the interior maxima w_1 or w_2, or at the intersection point w_{12}. The interior maxima are easily found as $w_j = \mu_j / \gamma \sigma_j^2$. The

intersection point w_{12} follows from

$$Q_1(w) = Q_2(w) \quad \Leftrightarrow \quad (\mu_1 - \mu_2)w = \frac{1}{2}\gamma(\sigma_1^2 - \sigma_2^2)w^2,$$

so that

$$w_{12} = \frac{2(\mu_1 - \mu_2)}{\gamma(\sigma_1^2 - \sigma_2^2)}. \tag{A4}$$

Since the expected returns are assumed to be positive, the optimal investment in the risky asset is positive. We therefore do not need to consider portfolios $w \leq 0$.

Let us first check if w_1 can be a valid optimum. It is only a valid optimum if $w_1 > w_{12}$, as otherwise $Q_R(w) = Q_2(w)$. Using the expressions for w_1 and w_{12} this leads to the inequality

$$\frac{\mu_1}{\gamma\sigma_1^2} > \frac{2(\mu_1 - \mu_2)}{\gamma(\sigma_1^2 - \sigma_2^2)}, \tag{A5}$$

which reduces to

$$\mu_2 > \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2}\mu_1. \tag{A6}$$

Since $\sigma_1^2 > \sigma_2^2$, we have $\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2} < 1$, and there exist pairs (μ_1, σ_1^2) and (μ_2, σ_2^2) for which the inequality holds. Moreover, if w_1 is a valid local optimum, it is also the global optimum, since the local optimum w_2 of $Q_2(w)$ can only be valid if $Q_2(w_2) < Q_1(w_2)$.

An analogous argument leads to the condition for w_2 to be a valid optimum. Solving the inequality $w_2 < w_{12}$ gives

$$\mu_2 < \frac{2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\mu_1. \tag{A7}$$

Again, since we have assumed $\sigma_1^2 > \sigma_2^2$, this optimum can occur for sufficiently small μ_2. Inequalities (A6) and (A7) can not hold simultaneously since

$$\frac{2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} < \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2}. \tag{A8}$$

Therefore there also exists an interval

$$\frac{2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\mu_1 < \mu_2 \leq \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2}\mu_1, \tag{A9}$$

for which the robust optimal portfolio is the corner solution w_{12}.

Proof of proposition 2

According to the Kuhn-Tucker optimality conditions in the proof of theorem 1 we need to check three possibilities: either one of the constraints is active or both constraints

are active. Let us start by defining the robust criterion function as

$$Q_R(w) = \min_{j=1,2} Q_j(w) = \begin{cases} Q_1(w) & \text{if } (\mu_1 - \mu_2)'w \leq 0, \\ Q_2(w) & \text{if } (\mu_1 - \mu_2)'w \geq 0. \end{cases} \tag{A10}$$

Next we check when $w_1 = \frac{1}{\gamma}\Sigma^{-1}\mu_1$ is a valid optimum. For this we need the inequality

$$(\mu_1 - \mu_2)'w_1 < 0, \tag{A11}$$

which by direct substitution for w_1 is equivalent to $\rho_{11} \leq \rho_{12}$. Analogously, the interior optimum w_2 is valid if $\rho_{22} \leq \rho_{12}$. Since the Cauchy-Schwartz inequality implies that $\rho_{11}\rho_{22} > \rho_{12}^2$, both inequalities can not hold simultaneously. Therefore if the local optimum is admissable, it is also the global optimum. On the other hand, it is possible that both constraints are active, i.e. both inequalities do not hold, in which case we obtain the solution from (A2b) as

$$w_R = \frac{1}{\gamma}\Sigma^{-1}\left(\lambda\mu_1 + (1-\lambda)\mu_2\right). \tag{A12}$$

The value of the Lagrange multiplier λ then follows from the constraint $Q_1(w) = Q_2(w)$, such that

$$(\mu_1 - \mu_2)'\, w = 0 \tag{A13}$$

Substituting (A12) into (A13) leads to a linear equation that can be solved for λ as

$$\lambda = \frac{\rho_{22} - \rho_{12}}{(\rho_{11} - \rho_{12}) + (\rho_{22} - \rho_{12})}, \tag{A14}$$

and thus proves the result of the theorem.

Proof of proposition 3

Under the assumption of the theorem the robust portfolio is a strict convex combination of w_1 and w_2, and not one of the extremes. Substituting the portfolio weight w_R in the loss function (6) with true expected returns μ_0 we obtain

$$\begin{aligned} 2\gamma L(\mu_0|w_R) &= (\mu_0 - \gamma\Sigma w_R)'\,\Sigma^{-1}\,(\mu_0 - \gamma\Sigma w_R) \\ &= (\mu_0 - \lambda\mu_1 - (1-\lambda)\mu_2)\,\Sigma^{-1}\,(\mu_0 - \lambda\mu_1 - (1-\lambda)\mu_2) \\ &= \rho_{00} + \lambda^2\rho_{11} + (1-\lambda)^2\rho_{22} \\ &\qquad - 2\lambda\rho_{01} - 2(1-\lambda)\rho_{02} + 2\lambda(1-\lambda)\rho_{12} \end{aligned} \tag{A15}$$

The analogous expression for the utility loss of expert 1 is

$$2\gamma L(\mu_0|w_1) = \rho_{00} + \rho_{11} - 2\rho_{01} \tag{A16}$$

Subtracting (A16) from (A15) gives

$$\begin{aligned} 2\gamma \left(L(\mu_0|w_1) - L(\mu_0|w_R)\right) = (1-\lambda^2)\rho_{11} - (1-\lambda)^2\rho_{22} \\ - 2\lambda(1-\lambda)\rho_{12} - 2(1-\lambda)\rho_{01} + 2(1-\lambda)\rho_{02} \end{aligned} \tag{A17}$$

For the robust investor to have lower loss, this must be positive. Divide (A17) by $1-\lambda$, require the result to be positive and solve for $\rho_{02} - \rho_{01}$,

$$\begin{aligned} 2(\rho_{02} - \rho_{01}) &> -(\lambda+1)\rho_{11} + (1-\lambda)\rho_{22} + 2\lambda\rho_{12} \\ &= -\rho_{11} + \rho_{22} - \lambda(\rho_{11} + \rho_{22} - 2\rho_{12}) \\ &= \rho_{12} - \rho_{11} \end{aligned} \tag{A18}$$

The last line follows from the definition of λ. The result establishes the first inequality in the proposition. Proof of the second inequality is analogous by comparing $L(\mu_0|w_2)$ and $L(\mu_0|w_R)$.

Proof of proposition 4

All we need in the proof is that $\lambda = 0$ in proposition 2. For this purpose we check the condition

$$\hat{\rho}_{22} < \hat{\rho}_{12}. \tag{A19}$$

For ease of notation, define $V_{\beta\beta} = \beta'\Sigma^{-1}\beta$. Substituting for $\hat{\mu}_1$ and $\hat{\mu}_2$ we have that

$$\begin{aligned} \hat{\rho}_{12} - \hat{\rho}_{22} &= (\hat{\mu}_1 - \hat{\mu}_2)'\Sigma^{-1}\hat{\mu}_2 \\ &= (1-a)\bar{y}' \left(I - \frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)' \Sigma^{-1} \left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\bar{y} \\ &= a(1-a)\bar{y}' \left(\Sigma^{-1} - \Sigma^{-1}\beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\right)\bar{y} \\ &\geq 0. \end{aligned} \tag{A20}$$

The last inequality follows, since $0 < a < 1$ and by the positive semi-definiteness of $\Sigma^{-1} - \Sigma^{-1}\beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}$.

Proof of proposition 5

For the expected loss we evaluate

$$2\gamma \mathsf{E}[L(\mu|\hat{w}_R)] = \mathsf{E}\left[(\hat{\mu}_2 - \mu)'\Sigma^{-1}(\hat{\mu}_2 - \mu)\right] \tag{A21}$$

Start by writing

$$\begin{aligned}\hat{\mu}_2 - \mu &= \left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\bar{y} - \mu \\ &= -(1-a)\left(I - \frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\mu + \left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\eta, \end{aligned} \tag{A22}$$

where we have used that $\bar{y} = \mu + \eta$ with η an error term with zero mean and covariance matrix Σ/T. Using the decomposition (A22) the expectation of the quadratic form in (A21) will have two terms. The first is the quadratic term in μ,

$$\mu'\left(I - \frac{\Sigma^{-1}\beta\beta'}{V_{\beta\beta}}\right)\Sigma^{-1}\left(I - \frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\mu = V_{\mu\mu} - \frac{V_{\mu\beta}^2}{V_{\beta\beta}}, \tag{A23}$$

where $V_{\mu\mu} = \mu'\Sigma^{-1}\mu$ and $V_{\mu\beta} = \mu'\Sigma^{-1}\beta$. Since $V_{\mu\mu}$ is also the square of the maximal Sharpe ratio given the true values of the parameters, and

$$R^2 = \frac{V_{\mu\beta}^2}{V_{\beta\beta}V_{\mu\mu}}$$

is the GLS squared correlation between β and μ, we can rewrite (A23) as

$$V_{\mu\mu} - \frac{V_{\mu\beta}^2}{V_{\beta\beta}} = \mathsf{Sh}^2(1 - R^2) \tag{A24}$$

For the second quadratic term we have

$$\begin{aligned}&\mathsf{E}\left[\eta'\left(aI + (1-a)\frac{\Sigma^{-1}\beta\beta'}{V_{\beta\beta}}\right)\Sigma^{-1}\left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\eta\right] \\ =&\frac{1}{T}\mathrm{tr}\left(\left(a\Sigma^{-1} + (1-a)\frac{\Sigma^{-1}\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\left(a\Sigma + (1-a)\frac{\beta\beta'}{V_{\beta\beta}}\right)\right) \\ =&\frac{1}{T}(a^2N + 1 - a^2)\end{aligned} \tag{A25}$$

Adding the expressions for bias and variance and substituting in (A21) we finally obtain

$$\mathsf{E}[L(\mu|\hat{w}_R)] = \frac{1}{2\gamma}\left((1-a)^2\mathsf{Sh}^2(1-R^2) + \frac{1}{T}(a^2N + 1 - a^2)\right) \tag{A26}$$

The derivation of the expected disappointment uses similar calculations,

$$\begin{aligned}\mathsf{E}[D_R(\hat{w}_R)] &= \mathsf{E}\left[Q_R(\hat{w}_R) - Q(\hat{w}_R)\right] \\ &= \frac{1}{\gamma}\mathsf{E}\left[(\hat{\mu}_2 - \mu)'\Sigma^{-1}\hat{\mu}_2)\right] \end{aligned} \tag{A27}$$

Using (A22) we have that

$$\begin{aligned}(\hat{\mu}_2 - \mu)'\Sigma^{-1}\hat{\mu}_2 = &\left(-(1-a)\left(I - \frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\mu + \left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\eta\right)' \\ &\Sigma^{-1}\left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)(\mu + \eta)\end{aligned} \tag{A28}$$

Taking expectations, using $\mathsf{E}[\eta] = 0$ and $\mathsf{E}[\eta\eta'] = \Sigma/T$, we obtain

$$\begin{aligned}\mathsf{E}\left[(\hat{\mu}_2 - \mu)'\Sigma^{-1}\hat{\mu}_2\right] =& -(1-a)\mu'\left(I - \frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\Sigma^{-1}\left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\mu \\ &+ \mathsf{E}\left[\eta'\left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\Sigma^{-1}\left(aI + (1-a)\frac{\beta\beta'\Sigma^{-1}}{V_{\beta\beta}}\right)\eta\right] \\ =& -a(1-a)\left(V_{\mu\mu} - \frac{V_{\mu\beta}^2}{V_{\beta\beta}}\right) + \frac{1}{T}\mathrm{tr}\left(a^2 I + (1-a)^2\frac{\Sigma^{-1}\beta\beta'}{V_{\beta\beta}}\right) \\ =& -a(1-a)\mathsf{Sh}^2(1-R^2) + \frac{1}{T}(a^2 N + 1 - a^2)\end{aligned} \tag{A29}$$

Dividing by γ and substituting in (A27) gives the result in the proposition.

Appendix B Factor Model Posterior

To derive the posterior estimators, define the data matrices $Y = (y_1', \ldots, y_T')$, $X = (x_1', \ldots x_t')$ and $Z = (\iota_T \; X)$ with dimensions $(T \times N)$, $(T \times k)$ and $(T \times (k+1))$, respectively. Also define the parameter matrix $A = [\alpha \; B]'$ and let $a = \mathrm{vec}(A)$.

The data information can be summarised in the regression statistics $\hat{A} = (Z'Z)^{-1}Z'Y$, $\hat{a} = \mathrm{vec}(\hat{A})$, $\hat{D} = (Y - Z\hat{A})'(Y - Z\hat{A})/T$, $\hat{\nu} = X'\iota_T/T$ and $\hat{\Psi} = (X - \iota_T\hat{\nu}')'(X - \iota_T\hat{\nu}')/T$. Pàstor and Stambaugh (2000) report the posterior

$$\begin{aligned}&a \mid D, X, Y \sim N\left((I_N \otimes F^{-1}Z'Z)\hat{a}, D \otimes F^{-1}\right), \\ &D^{-1} \mid X, Y \sim W\left(T + df - M, (H + T\hat{D} + \hat{A}'P\hat{A})^{-1}\right)\end{aligned} \tag{B1}$$

with $F = \Delta + Z'Z$ and Δ a diagonal matrix of order $(M+1)$ filled with zeros except for the (1,1)-element which is equal to s^2/σ_α^2; $P = Z'(I_T - ZF^{-1}Z')Z$. For the

parameters of the factors they report the posterior

$$\begin{aligned} \nu \mid \Psi, X, Y &\sim N\left(\hat{\nu}, \Psi/T\right), \\ \Psi^{-1} \mid X, Y &\sim W\left(T-1, \hat{\Psi}^{-1}/T\right) \end{aligned} \tag{B2}$$

Since B and ν are independent in the posterior, the posterior mean of the expected return vector is

$$\hat{\mu} = \mathsf{E}[\mu|\ X, Y] = \tilde{\alpha} + \tilde{B}\tilde{\nu}. \tag{B3}$$

where the tilde refers to the posterior expectation given the data. An analytical expression for the posterior variance

$$\Sigma = B\Psi B' + D \tag{B4}$$

is not available. We could (approximately) compute the posterior variance by Monte-Carlo simulation, although this is computationally intensive to do repeatedly within the bootstrap simulations. Instead we will use

$$\hat{\Sigma} = \tilde{B}\tilde{\Psi}\tilde{B}' + \tilde{D} \tag{B5}$$

in our computations. This underestimates the posterior covariance matrix, because it ignores the posterior variance of the factor loadings. However, since posterior (co-)variances of the factor loadings B are typically small, the underestimation is limited. For the 25 Fama-French portfolios and $T = 60$ the underestimation is of the order of 1% in our empirical study.

References

Ben-Tal, A. and Nemirovski, A.: 2001. *Lectures on modern convex optimization: analysis methods and engineering applications*. SIAM-MPS.

Black, F. and Litterman, R.: 1990. Asset allocation: Combining investor views with market equilibrium. *Technical report*. Goldman, Sachs & Co., Fixed Income Research.

Chopra, V. and Ziemba, W. 1993. The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *Journal of Portfolio Management* pp. 6–11. winter.

Cochrane, J.: 2001. *Asset Pricing*. Princeton University Press.

Davis, J., Fama, E. and French, K. 2000. Characteristics, Covariances and Average Returns: 1929 to 1997. *Journal of Finance* **55**, 389–406.

DeMiguel, V., Garlappi, L. and Uppal, R.: 2006. 1/N. Working Paper.

Fama, E. and French, K. 1996. Multifactor explanations for asset pricing anomalies. *Journal of Finance* **51**, 55–94.

Garlappi, L., Uppal, R. and Wang, T. 2007. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach.

Gilboa, I. and Schmeidler, D. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* **18**(2), pp. 141–153.

Goldfarb, D. and Iyengar, G. 2003. Robust portfolio selection problems. *Mathematics of Operations Research* **28**(1), pp. 1–38.

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

Jobson, J. and Korkie, B. 1981. Putting Markowitz theory to work. *Journal of Portfolio Management* **7**(4), pp. 70–74.

Jorion, P. 1985. International Portfolio Diversification with Estimation Risk. *Journal of Business* **58**, 259–278.

Jorion, P. 1986. Bayes–Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* **21**(3), pp. 279–292.

Kan, R. and Zhou, G. 2006. Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*. forthcoming.

Klibanoff, P., Marinacci, M. and Mukerji, S. 2005. A Smooth Model of Decision Making under Ambiguity. *Econometrica* **73**, 1849–1892.

Michaud, R. 1989. The Markowitz optimization enigma: is 'optimized' optimal?. *Financial Analyst Journal* **45**(1), pp. 31–42.

Pàstor, L. and Stambaugh, F. 2000. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics* **56**, pp. 335–381.

Rustem, B., Becker, R. and Marty, W. 2000. Robust min-max portfolio strategies for rival forecast and risk scenarios. *Journal of Economic Dynamics and Control* **24**, pp. 1591–1621.

Table 1: 25 Fama-French portfolios (full sample)

The table shows the ex-ante and ex-post monthly performance of optimal mean variance portfolios of the 25 Size and Book-to-Market portfolios with risk aversion parameter $\gamma = 5$. Parameters are estimated using the full sample (July 1963 to December 2002). The alternative models are the unrestricted model ("sample", $j = 0$) and two structured models based on dogmatic ($\sigma_\alpha = 0$) priors for the Fama-French 3-factor ("FF3", $j = 1$) and CAPM ($j = 2$). The last column refers to the equally weighted portfolio of all 25 portfolios.
The ex-ante performance is evaluated conditional on the parameters for each model ($j = 0, 1, 2$). For the equally weighted portfolio ex-ante statistics are not available, since the portfolio weights are not baesed on parameter estimates. The ex-post performance is evaluated using the true values (μ_0, Σ_0). Means and standard deviations are in units of percent per month.
The lower panel of the table reports characteristics of the optimal portfolio: the total fraction of wealth invested in risky assets and the cross-sectional standard deviation of the weights w_i.

Model	sample	FF3	CAPM	EW
Ex-ante expected performance				
Excess return ($w_j'\hat{\mu}_j$)	4.62	1.09	0.16	
Standard deviation ($(w_j'\hat{\Sigma}_j w_j)^{1/2}$)	9.61	4.66	1.82	
Utility ($Q_j(w_j)$)	2.31	0.54	0.08	
Ex-post realised performance				
Excess return ($w_j'\mu_0$)	4.62	1.11	0.18	0.64
Standard deviation ($(w_j'\Sigma_0 w_j)^{1/2}$)	9.61	4.65	1.82	5.07
Utility ($Q_0(w_j)$)	2.31	0.57	0.09	-0.01
Loss ($L(\mu_0, \Sigma_0 \mid w_j)$)		1.74	2.21	2.32
Disappointment ($D_j(w_j)$)		-0.03	-0.01	
Portfolio characteristics				
Sum weights ($\bar{w}_j$)	1.64	0.73	0.39	1
Cross-sectional stdev	1.23	0.16	0.03	0

Table 2: 25 Fama-French portfolios (60 observations)

The table shows the average ex-ante and ex-post performance of optimal mean variance portfolios with risk aversion parameter $\gamma = 5$. The alternative models are the unrestricted sample moments ("data"); a portfolio with short-sell restrictions using the unrestricted sample moments; portfolios using priors based on the Fama-French 3-factor model and the CAPM models; and the robust portfolio, where the investor is robust over the different CAPM and FF3 recommendations. The portfolio characteristics are the total fraction of wealth invested in risky assets ("sum of weights"); the cross-sectional standard deviation of the portfolio weights; the number of assets in the portfolio; and the percentage of simulations in which a prior puts a binding constraint on the robust portfolio. All results are averages over 1,000 bootstrap samples of 60 observations. The ex-post performance is evaluated under the full sample true parameters. Expected returns and standard deviations are in percent per month.

	data	FF3		CAPM		robust	short-
Prior variance (σ_α)	∞	0.1	4.0	0.1	4.0		sell
Ex-ante expected performance							
Excess return (E_j)	24.74	2.14	18.24	0.54	11.28	0.54	1.58
Standard deviation (V_j)	21.88	6.12	18.85	2.63	14.83	2.63	4.93
Utility (Q_j)	12.37	1.07	9.12	0.27	5.64	0.27	0.79
Ex-post realised performance							
Excess return (E_{0j})	9.52	1.18	7.36	0.20	5.08	0.21	0.86
Standard deviation (V_{0j})	43.17	6.57	29.99	2.70	18.82	2.70	5.35
Utility (Q_{0j})	-42.13	-0.12	-16.72	-0.10	-4.40	-0.09	-0.17
Expected loss (L_{0j})	44.44	2.42	19.03	2.41	6.70	2.40	2.47
Disappointment (D_{0j})	54.50	1.19	25.84	0.37	10.04	0.36	0.96
Portfolio characteristics							
Sum weights ($\bar{w}_j$)	3.26	0.81	2.59	0.44	1.94	0.44	1.11
Cross-sectional std. dev.	6.16	0.21	3.99	0.04	2.20	0.04	0.17
Average turnover per asset	6.4	0.18	4.0	0.04	2.2	0.04	0.07
Binding in robust		68%	43%	99%	1%		
# Assets in portfolio	25	25	25	25	25	25	2.3

Table 3: 25 Fama-French portfolios (60 observations without riskfree asset)

Table contents are as in table 2 apart from the additional portfolio restriction $\iota' w = 1$.

	sample	FF3		CAPM		robust
Prior variance (σ_α)	∞	0.1	4.0	0.1	4.0	
Ex-ante expected performance						
Excess return (E_j)	21.17	1.90	15.92	0.44	10.00	0.44
Standard deviation (V_j)	19.96	6.08	17.34	3.39	13.75	3.68
Utility (Q_j)	10.87	0.90	8.20	0.08	5.15	0.08
Ex-post realised performance						
Excess return (E_{0j})	7.17	1.35	5.79	0.68	4.23	0.69
Standard deviation (V_{0j})	38.06	6.70	26.94	4.31	17.00	4.30
Utility (Q_{0j})	-32.65	0.13	-13.63	0.20	-3.47	0.22
Expected loss (L_{0j})	34.95	2.18	15.93	2.10	5.78	2.08
Disappointment (D_{0j})	43.52	0.67	21.83	-0.13	8.62	-0.14
Portfolio characteristics						
Cross-sectional standard dev.	5.54	0.36	3.70	0.25	2.09	0.25
Binding in robust			28%	12%	99%	3%

Table 4: 81 European portfolios (full sample)

The table shows the ex-ante and ex-post monthly performance of optimal mean variance portfolios of the 81 European portfolios with risk aversion parameter $\gamma = 5$. Parameters are estimated using the full sample (January 1975 to December 2001). The alternative models are the unrestricted model ("sample", $j = 0$) and four structured models based on dogmatic ($\sigma_\alpha = 0$) priors for the International CAPM (CAPM), Fama-French 3-factor (IFF), International CAPM with exchange rate premia (FX) and a model integrating all risk factors in the previous models (ALL). The last column refers to the equally weighted portfolio of all 81 portfolios.
The ex-ante performance is evaluated conditional on the parameters for each model ($j = 0, 1, 2$). For the equally weighted portfolio ex-ante statistics are not available, since the portfolio weights are not baesed on parameter estimates. The ex-post performance is evaluated using the true values (μ_0, Σ_0). Means and standard deviations are in units of percent per month.
The lower panel of the table reports characteristics of the optimal portfolio: the total fraction of wealth invested in risky assets and the cross-sectional standard deviation of the weights w_i.

Model	sample	CAPM	IFF	FX	ALL	EW
Ex-ante expected performance						
Excess return $(w_j'\mu_j)$	7.31	0.13	0.58	0.17	0.63	
Standard deviation $((w_j'\Sigma_j w_j)^{1/2})$	12.09	1.63	3.40	1.84	3.55	
Utility $(Q_j(w_j))$	3.66	0.07	0.29	0.08	0.32	
Ex-post realised performance						
Excess return $(w_j'\mu_0)$	7.31	0.28	0.67	0.23	0.63	0.64
Standard deviation $((w_j'\Sigma_0 w_j)^{1/2})$	12.09	1.61	3.37	1.83	3.53	4.78
Utility $(Q_0(w_j))$	3.66	0.21	0.39	0.14	0.32	0.07
Loss $(L(\mu_0, \Sigma_0 \mid w_j))$		3.44	3.27	3.51	3.34	3.59
Disappointment $(D_j(w_j))$		-0.14	-0.10	-0.06	0.00	
Portfolio characteristics						
Sum weights $(\bar{w}_j)$	1.08	0.32	0.36	0.28	0.31	1
Cross-sectional stdev	2.17	0.08	0.19	0.17	0.26	0

Table 5: 81 European portfolios (150 observations)

The table shows the ex-ante and ex-post monthly performance of optimal mean variance portfolios of the 81 European portfolios with parameters estimated from 150 observations. See tables 2 and 4 for additional notes.

		Asset Pricing Models					short-
	data	CAPM	IFF	FX	ALL	robust	sell
Ex-ante expected performance							
Excess return (E_j)	51.06	0.28	1.04	0.61	1.43	0.28	1.88
Standard deviation (V_j)	31.60	1.96	4.33	3.27	5.17	1.96	5.89
Utility (Q_j)	25.53	0.13	0.51	0.31	0.71	0.14	0.94
Expost realised performance							
Excess return (E_{0j})	19.50	0.27	0.74	0.25	0.72	0.27	1.07
Standard deviation (V_{0j})	97.09	2.32	5.66	4.62	7.27	2.32	6.19
Utility (Q_{0j})	-244.5	0.07	-0.18	-0.38	-0.73	0.08	-0.00
Expected loss (L_{0j})	248.1	3.59	3.84	4.04	4.39	3.58	3.66
Disappointment (D_{0j})	270.1	0.06	0.69	0.69	1.44	0.06	0.94
Portfolio characteristics							
Sum of weights ($\bar{w}_j$)	2.23	0.32	0.35	0.27	0.28	0.32	1.18
Cross-sectional std.dev.	16.62	0.25	0.73	0.69	1.06	0.25	0.11
Average turnover per asset	85.50	0.03	0.18	0.17	0.38	0.03	0.01
Binding in robust		99%	13%	58%	5%		
# Assets in portfolio	81	81	81	81	81	81	6.1

Figure 1: Robust Portfolio Choice

The figure shows the mean-variance objective functions for the case with two experts and a single risky asset. Expert one (dashed line) proposes both higher expected returns and larger variance than expert two (dotted line). The solid line is the minimum of the two objective functions and presents the robust objective function. The horizontal axis shows the portfolio weight of the risky asset, the vertical axis shows the value of the mean-variance objective. Panel A depicts the situation where the robust optimum coincides with the optimum of expert one, in panel C with the optimum of expert two, and in panel B with the portfolio for which the objective values for both experts are equal.

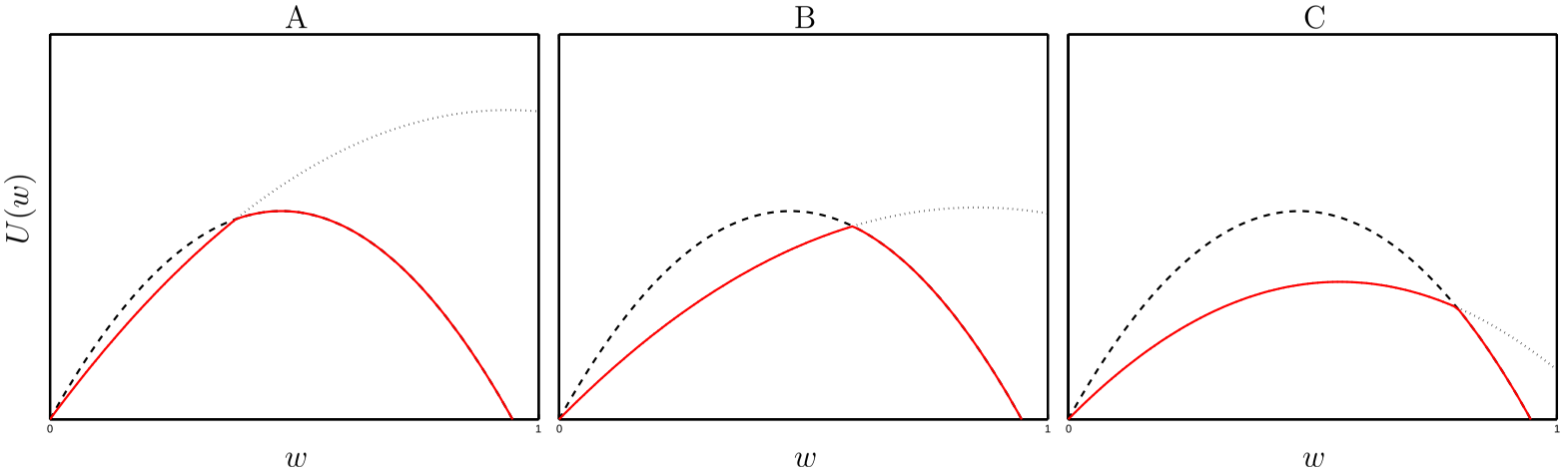


Figure 2: Expected Loss

The figure shows the relative loss $\tilde{L}(\mu/\sigma^2)$ (vertical axis) as a function of μ/σ^2 (horizontal axis) for three portfolios: the optimal portfolios according to the advice of experts 1 and 2 (dashed lines) for which the the expected return are depicted in figure 1 and the robust portfolio $w_R = w_{12}$ (solid line). The relative utility loss is defined by

$$\tilde{L}(\mu/\sigma^2|w) = \frac{L(\mu,\sigma^2|w)}{Q(w_0|\mu,\sigma^2)} = \left(1 - \frac{\gamma w}{\mu/\sigma^2}\right)^2$$

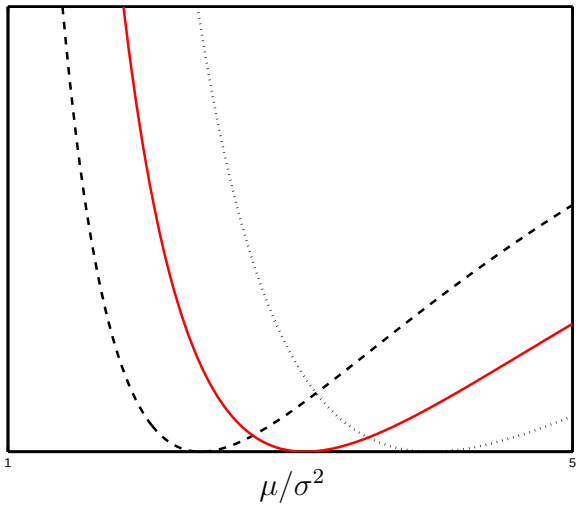


Figure 3: Relative Performance

The figure indicates the areas where each of the expert recommendations has the lowest loss. The figure refers to a situation with two assets and two experts who agree on the covariance matrix, but disagree on the expected returns for assets 1 and 2.

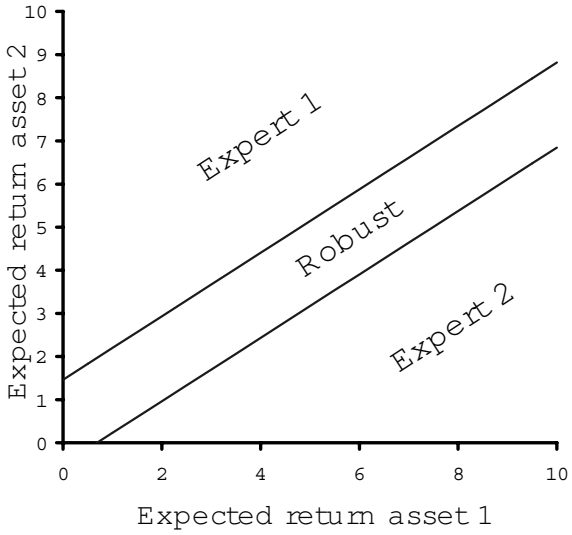


Figure 4: Expected loss as a function of T (FF25 data)

Notes: The figure shows the expected loss as a function of the number of observations (horizontal axis). The lines correspond to the unstructured portfolio (solid); the portfolio based on the tight CAPM prior (dashed gray); the portfolio based on the loose CAPM prior (dashed); the portfolio based on the tight FF3 prior (dashed gray); the portfolio based on the loose FF3 prior (dashed); and the model robust portfolio (also dashed gray).

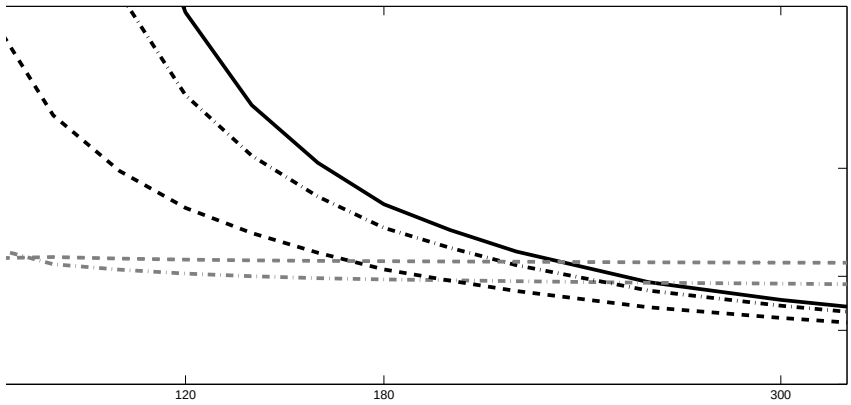
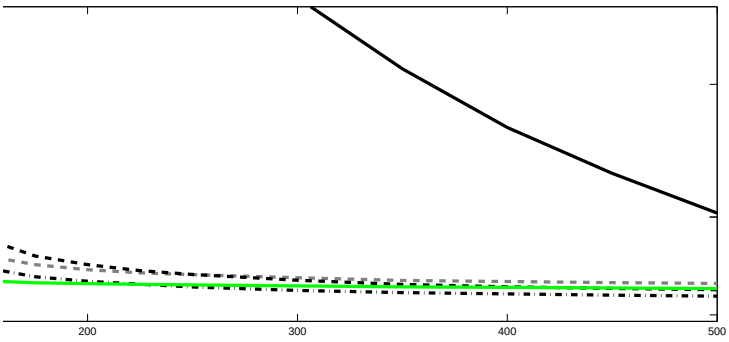


Figure 5: Expected loss as a function of T (European data)

Notes: The figure shows the expected loss as a function of the number of observations (horizontal axis). The lines correspond to the unstructured portfolio (solid); the portfolio based on the ICAPM-FX (dashed); the portfolio based on the ICAPM prior (dashed gray); the portfolio based on the FF 3 factor prior (dash-dot); and the CAPM and model robust portfolio (solid gray).