Robust Portfolio Optimisation with Multiple Experts

Review of Finance (2010) 14: 343–383
doi: 10.1093/rof/rfn028
Advance Access publication: 25 December 2008

Robust Portfolio Optimisation with Multiple Experts*

FRANK LUTGENS[1] and PETER C. SCHOTMAN[2]
[1]*Maastricht University;* [2]*NETSPAR*

Abstract. We consider mean-variance portfolio choice of a robust investor. The investor receives advice from *J* experts, each with a different prior for expected returns and risk, and follows a min-max portfolio strategy. The robust investor endogenously combines the experts' estimates. When experts agree on the main return generating factors, the investor relies on the advice of the expert with the strongest prior. Dispersed advice leads to averaging of the alternative estimates. The robust investor is likely to outperform alternative strategies. The theoretical analysis is supported by numerical simulations for the 25 Fama-French portfolios and for 81 European country and value portfolios.

JEL Classification: C11, C44, D80

1. Introduction

Quantitative portfolio management requires estimates of expected return and risk. Unfortunately analysts often disagree about the general outlook of the stock market and the relative risks and returns of individual stocks. The decision maker is thus confronted with uncertain inputs for the portfolio optimisation.

We analyse the decisions and performance of a robust decision maker. The robust portfolio is defined as the portfolio that performs best under the worst case scenario formed by the recommendations of *J* expert advisors. They employ different models to estimate expected return and risk. 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 behavioural viewpoint the decision maker can be ambiguity averse. Gilboa and Schmeidler (1989) and Klibanoff et al (2005) provide an axiomatic foundation for

* 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. An anonymous referee provided valuable suggestions on the presentation the results. All errors are of course our own.

the max-min utility optimisation of the decision maker. The max-min approach does not require the investor to attach probabilities to the recommendations of the experts.

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. The result is what Michaud (1989) has called 'error maximisation'.

The performance of the standard mean-variance portfolio based on sample estimates is generally disappointing. The portfolio is not only suboptimal, but the estimated means and covariances also 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. Accordingly we conjecture that the robust portfolio will exhibit good out-of-sample performance.

The robust decision rule advocates a portfolio that is not necessarily optimal for a specific set of inputs but performs well over a range of inputs. 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. The robust portfolio is less subject to sampling variation.

Analytically the robust portfolio resembles the optimal portfolio based on a convex combination of alternative parameters estimates (section 3). The robust optimisation endogenously determines the weights given to the different estimates. Two-fund separation applies: the optimal portfolio is a combination of the riskfree asset and a portfolio of risky assets, while the composition of the risky portfolio is independent of the investor's risk aversion.

Full analytical results are available 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 pessimistic recommendations would. This result is interesting, since robust decision rules (in particular the max-min approach) are often criticised for being extremely conservative. The second case delineates 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.

For further study of the robust portfolio structure we consider parameter estimates based on specific Bayesian model priors. In section 4.2 we derive analytical results for an investor confronted with different, nested models. The central analytical result in this section states that the robust portfolio is determined by the most informative prior on an asset pricing model. 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.

When both expected returns and the covariance matrix are unknown, we need numerical methods to construct the optimal portfolios and evaluate the performance. This is the setting for section 5. To obtain relevant simulation results, the simulation is based on two commonly used datasets. 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 use Pastor and Stambaugh (2000) type priors to obtain their estimates of expected returns and covariances. Experts differ in their prior views.

For the 25 Fama-French portfolios we consider a set of priors with different degrees of precision that are centered on either the CAPM or the Fama-French 3-factor model. Although the models and priors are not fully nested, an expert using a tight prior on the CAPM presents the most structured advice to the investor. Our simulations confirm that the robust investor will almost always follow this advice. For sample sizes up to about ten years of monthly data for estimation, this robust strategy also outperforms all other portfolios.

For the European data we consider models that offer conflicting views on the cross section of expected returns. One expert uses the International CAPM including exchange rate risk factors. Another expert uses an international Fama-French model with a value factor. In this case the robust investor will always use a weighted average of the estimates of the two experts. This portfolio is shown to outperform all other strategies for sample sizes up to 15 years of monthly data.

Important in both cases is the close correspondence between the ex-ante anticipated performance and the ex-post realised performance of the robust investor. The robust investor will on average not be disappointed by the realised risk-return trade-off. For most other strategies the ex-ante performance promised by the parameter estimates strongly overestimates the realised ex-post performance of the portfolio.

2. Literature

Mean-variance portfolios are known to perform poorly out-of-sample due to estimation error and model misspecification. Early references include Jobson and Korkie (1981) and Jorion (1985). More recently DeMiguel et al (2007) find that optimised portfolios rarely outperform a naive equally weighted benchmark. Kan and Zhou (2006) provide further analytical treatment and more references.

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 (1992) the prior means are the expected returns for which the observed market portfolio would be optimal. In Pastor and Stambuagh (2000) the prior is centered around the expected returns implied by the CAPM. Their prior also involves shrinkage on the covariance matrix. Ledoit and Wolf (2003) provide further evidence on the benefits of shrinking the covariance matrix towards a factor structure. 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 (1992). Like their Bayesian investor, our robust investor combines the distinct views of a finite set of experts. The fundamental difference is that the robust investor dislikes the uncertainty generated by the dispersion in opinions. The operational difference is that the Bayesian investor must attach prior probabilities to each of the views, either exogenously or empirically determined. The robust investor is not able to attach weights to different models a priori, but simultaneously determines the worst case and the robust portfolio. The weights that constitute the worst case are endogenous and depend on the investor's objective function.

Several other studies have looked at robust solutions to the static mean-variance portfolio problem. Garlappi et al. (2007) and Goldfarb and Iyengar (2003) address estimation uncertainty by considering classical confidence intervals. Both studies consider confidence intervals for a single model and define the robust portfolio return as the worst case expected return within the joint confidence set of all expected returns.[1] The difference with our approach is that we consider point estimates from different models.

Like us, Rustem et al (2000) consider a finite set of rival models of expected return and risk, where the investor is confronted with alternative models for the returns. We extend Rustem et al (2000) by studying the robust portfolio structure for the main asset pricing models and its related stochastic and empirical performance. Another difference with Rustem et al (2007) is our assumption that a one-period

[1] As for terminology, Garlappi et al (2007) use the phrase *model uncertainty* for the case where they estimate confidence intervals for asset pricing alpha's. We use the term *model uncertainty* in the Bayesian sense as referring to different priors, similar as in, e.g., Avramov (2002) and Wang (2005).

riskfree asset is available. This assumption leads to two-fund separation and reduces the role of the risk-aversion parameter to a scaling factor.

The inclusion of a riskfree rate is also an important difference with Wang (2005). Wang (2005) considers portfolio choice for a family of priors indexed by a single prior precision hyperparameter that relates to both the means and the covariances of the returns. In Wang's (2005) model opinions only differ along one dimension: either the prior is centered on the CAPM, or it is centered on the Fama-French 3-factor model. Our setup is more general, since we simultaneously allow for different degrees of prior precision and for different asset pricing models. We further extend Wang (2005) by studying the ex-post performance of the robust portfolio.

3. Robust mean-variance portfolios

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 maximises 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.

3.1 ROBUST PORTFOLIO CHOICE

When true values (μ_0, Σ_0) for the parameters (μ, Σ) are known, the optimal investment in the risky assets is

$$w_0 = \frac{1}{\gamma}\Sigma_0^{-1}\mu_0. \tag{2}$$

In practice, the investor does not know the true values. She obtains advice from J experts who recommend their personal estimate for the expected returns and covariance matrices (μ_j, Σ_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 a 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 maximises 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 et al (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

$$S = \sum_{j=1}^{J} \lambda_j \Sigma_j,$$
$$m = \sum_{j=1}^{J} \lambda_j \mu_j, \tag{6}$$

and λ_j *are constants satisfying* $0 \leq \lambda_j \leq 1$ *and* $\sum \lambda_j = 1$. *Moreover, the* λ_j *are independent of the risk aversion* γ.

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. The last part of the theorem implies two-fund separation. Each optimal portfolio is a convex combination of an investment in the riskfree asset and the risky portfolio $S^{-1}m$. Two-fund separation critically depends on the existence of a riskfree rate. It does not obtain in alternative settings like the models considered in Wang (2005) and Rustem et al (2000). With a riskfree rate the $w = 0$ portfolio has a certain return. The worst-case utility is thus always non-negative.

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 be determined numerically. Efficient numerical solutions are easily obtained using standard optimisation software.

Since the Lagrange multipliers λ_j have the properties of probabilities, the solution resembles a Bayesian approach with λ_j as posterior probabilities on the reliability of the experts. In that sense it is similar to portfolio choice based on

Bayesian Model Averaging techniques considered by Avramov (2002). But in contrast to the Bayesian prior probabilities specified upfront, the Lagrange multipliers are endogenously determined in the robust optimisation simultaneously with the optimal portfolio. Unlike their Bayesian counterpart, the worst case depends on the investor's objective.[2] Positive Lagrange multipliers indicate the most pessimistic estimates in view of the optimal portfolio.

3.2 LOSS AND DISAPPOINTMENT

For general true values (μ_0, Σ_0) the *loss function* expresses the utility loss that results from selecting an arbitrary portfolio w which differs from the optimal portfolio w_0 based on the true parameters,

$$\begin{aligned} L(\mu_0, \Sigma_0 | w) &\equiv Q_0(w_0) - Q_0(w) \\ &= \frac{\gamma}{2}(w - w_0)' \Sigma_0 (w - w_0) \end{aligned} \tag{7}$$

The second line in (7) follows from the utility function (1), in which (2) is used to substitute $\mu_0 = \gamma \Sigma_0 w_0$ to express the loss in terms of the difference between the portfolios. The loss is large if the portfolio w is far from the optimal portfolio w_0. The loss depends on μ_0 through w_0.

Since $Q(w)$ is the risk adjusted expected excess return, or certainty equivalent, the loss function measures the risk-adjusted expected return difference between the optimal portfolio and a sub-optimal portfolio. For the interpretation of the loss function we view it as a function of the true parameters (μ_0, Σ_0) and compare the loss of each individual portfolio $w_j = \frac{1}{\gamma}\Sigma_j^{-1}\mu_j$ with the loss of the robust portfolio w_R.

If the investor strongly believes in the estimates of expert *j*, she would evaluate her portfolio decisions using the utility function $Q_j(w)$. Ex-ante, expert *j* expects that a portfolio w obtains a performance $Q_j(w)$. Ex-post, when true parameters have become known, the actual utility of portfolio w is $Q_0(w)$. We call the difference between promised and realised utility the investor's disappointment caused by following expert *j*'s advice,

$$D_j(w) = Q_j(w) - Q_0(w). \tag{8}$$

The disappointment can be both positive and negative. If expert *j* has been very pessimistic, while the actual parameters support much better investment opportunities,

[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 $\tilde{S} = \sum \lambda_j(\Sigma_j + (\mu_j - m)(\mu_j - m)')$ 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.

the investor is pleasantly surprised by the performance. For the robust investor the analogue of disappointment is

$$D_R(w) = Q_R(w) - Q_0(w) \tag{9}$$

If the true parameters (μ_0, Σ_0) are in the set of expert estimates, 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.

3.3 EXAMPLE: ONE RISKY ASSET, TWO EXPERTS

An example shows that a robust portfolio is not necessarily the most conservative portfolio. Consider an investor whose choice is limited to one risky asset and the risk free asset. 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 maximise the minimum of the two. The optimal portfolio is located at the maximum of the robust objective function.

The figure shows the three situations described in proposition 1. The parameter values (expected returns and variances) recommended by the experts determine which situation applies. In the figure the parameters of expert 1 (μ_1, σ_1^2) are held fixed and the parameters of expert 2 varied to obtain the three possible solutions. The robust portfolio is not necessarily the most conservative portfolio. In the middle

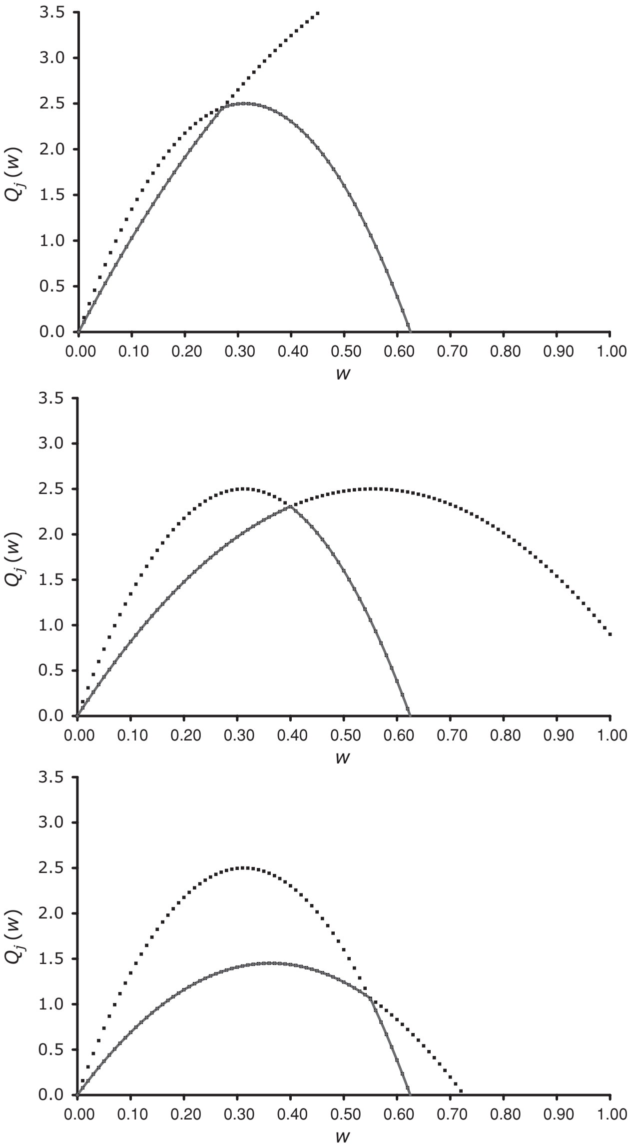

Figure 1. Robust Portfolio Choice with 2 Assets. The figure shows the mean-variance objective functions for the case with two experts and a single risky asset. The dotted lines are the objective functions using the parameters (μ_j, σ_j^2) of experts $j = 1, 2$. The solid line is the minimum of the two 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. The top panel depicts the situation where the robust optimum coincides with the optimum of expert 1, in the lower panel with the optimum of expert 2, and in middle panel with the portfolio for which the objective values for both experts are equal.

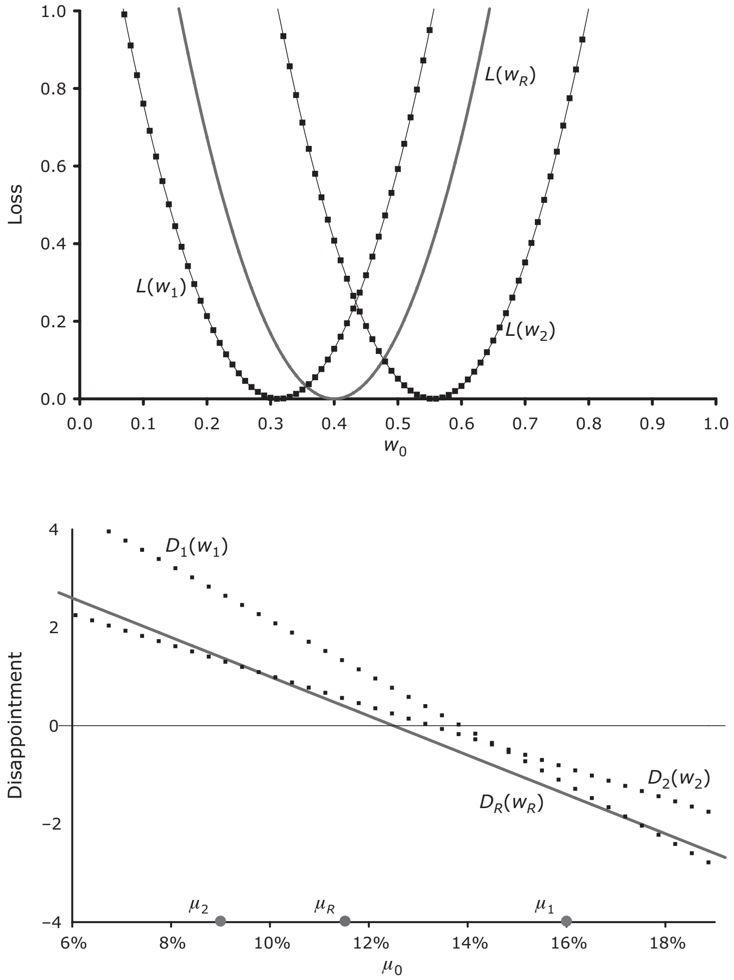

Figure 2. Loss and disappointment with 2 assets. The figure shows the loss $L(w_0|w_j)$ (vertical axis) as a function of the optimal weight w_0 (horizontal axis) under the true parameters (μ_0, σ_0^2) for three portfolios: the optimal portfolios according to the advice of experts 1 and 2 (dotted lines) for which the expected return are depicted in the middle panel of figure 1 and the robust portfolio $w_R = w_{12}$ (solid line). The lower panel shows the disappointment $D_j(\mu_0|w_j)$ (vertical axis) as a function of the true expected return μ_0 (horizontal axis) for given constant σ_0^2 for the same three portfolios. Expected returns μ_j for $j = 1, 2, R$ are marked on the horizontal axis.

panel, where the optimum is at the intersection of the two utility functions, the robust solution allocates a larger share to the risky asset than expert 1 would.

Figure 2 depicts the loss function of the alternative portfolios for the parameters in the middle panel of figure 1. None of the three portfolios dominates the others for all parameter values. For parameters that support a low optimal investment w_0, *i.e.*

small expected return and high variance, 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 (7) is strictly convex. This implies that the performance of the robust portfolio is strictly better than the worst performance alternative.

For given variance σ_0^2 the disappointment is linear in the ex-post optimal portfolio weight w_0 and thus linear in μ_0. In figure 2 the robust investor is never the most disappointed and for a wide range of true values even the least disappointed.

4. Known Covariance Matrix

To gain further insight in the structure and performance of robust portfolios we need to be more specific on the combinations (μ_j, Σ_j). We will assume that experts report Bayesian posterior moments based on different prior views. To obtain analytical results we also need a simplification. In this section the expert opinions only differ in their expected returns μ_j, but not in the covariance matrix Σ. As shown by Chopra and Ziemba (1993) uncertainty about the mean is usually more important than uncertainty about the covariance matrix. Furthermore, 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.

4.1 THE ROBUST PORTFOLIO

With a known covariance matrix 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 summarised 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 the same 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 characterised by the quadratic forms ρ_{ij}. The diagonal elements, ρ_{11} and ρ_{22}, are the squared Sharpe ratios of the optimal mean-variance portfolios anticipated by the experts. The scaled cross term $r_{12} = \rho_{12}/\sqrt{\rho_{11}\rho_{22}}$ is the correlation between the expert's recommendations. The squared correlation r_{12}^2 can be computed as the R^2 of the cross-sectional GLS regression of μ_1 on μ_2 (or vice versa) using Σ as the GLS covariance matrix.

To gain more insight in the structure of the robust portfolio we write the expression for λ in an equivalent form. Define $\kappa = \sqrt{\rho_{11}/\rho_{22}}$ as the optimal Sharpe ratio of expert 1 relative to expert 2. Then

$$\lambda = \begin{cases} 0 & \text{if} \quad \frac{1}{\kappa} \leq r_{12} \\ \frac{1}{2}\left(1 + \frac{1/\kappa - \kappa}{\kappa + 1/\kappa - 2r_{12}}\right) & \text{if} \quad \min\left(\kappa, \frac{1}{\kappa}\right) > r_{12} \\ 1 & \text{if} \quad \kappa \leq r_{12}. \end{cases} \tag{13}$$

only depends on relative quantities. As λ is decreasing in κ, the robust investor downweights the advice of an expert who promises a relatively high Sharpe ratio. The robust portfolio is tilted towards w_1 if expert 1 is more pessimistic ($\kappa < 1$) and tilted towards w_2 if expert 2 is more pessimistic ($\kappa > 1$). If both experts give advice with the same implied optimal Sharpe ratio ($\kappa = 1$), equation (13) simply states that the robust investor would give a weight of one half to both experts. A lower correlation implies more disparate expert recommendations and more incentive for the robust investor to combine the recommendations, *i.e.* it is more likely that the condition $r_{12} < \min(\kappa, \frac{1}{\kappa})$ is satisfied.

For the performance we compute the loss function (7) 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*

$$\frac{1}{2}(\rho_{12} - \rho_{11}) < \rho_{02} - \rho_{01} < \frac{1}{2}(\rho_{22} - \rho_{12}), \tag{14}$$

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

The proposition delineates the range of parameter values for which the robust portfolio will outperform the portfolio choice based on each of the individual experts. The robust portfolio has lower loss than either of the two expert portfolios for a set of expected returns delimited by the expert's estimates for the expected returns (geometrically these are parallel planes). Outside this set, the robust portfolio still features a better performance than the worst of the alternatives. Condition (14) 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.

4.2 PRIORS AND SAMPLING VARIATION

So far we have not been specific about how the experts arrive at their estimates. From here on we will assume that experts have different prior views on the parameters, but use the same sample data. The experts observe a sample of T observations, which they combine with their individual prior views to obtain posterior expectations. Each expert reports his posterior expectation $\hat{\mu}_j$ 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.

$$\hat{\mu}_j = \mu_j(Y) = \int_\mu \mu p_j(\mu|Y)\, \mathrm{d}\mu, \tag{15}$$

where $p_j(\mu|Y)$ is the posterior of μ using the prior of expert j. The investor does not know how the experts arrived at their estimates. She only obtains the posterior means, but neither knows the prior nor the sample data. As a result she can not assess the precision of the estimates.[3]

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 $p(Y|\mu_0)$ denote the true density, and let $\hat{w} = w(Y)$ be a portfolio rule depending on Y. The expected loss is defined as

$$\mathsf{E}\,[L(\mu_0|\hat{w})] = \int_Y L(\mu_0|w(Y))\; p(Y|\mu_0)\, \mathrm{d}Y \tag{16}$$

Analogously, the expected disappointment is

$$\mathsf{E}[D_j(\hat{w})] = \int_Y D_j(w(Y)|\hat{\mu}_j)\; p(Y|\mu_0)\, \mathrm{d}Y \tag{17}$$

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.

Suppose the experts disagree on expected returns, because they have different degrees of beliefs in an asset pricing model based on linear factor pricing. We

[3] Similar expressions exist for the posterior of the covariance matrix Σ. Since in this section it is assumed that Σ is known we suppress references to estimates of Σ.

assume that expert j uses the model

$$\mu_j = \alpha_j + B_j \nu_j, \tag{18}$$

where B_j is a $(N \times K)$ matrix of factor loadings, ν_j a vector of risk premia corresponding to the factors used by expert j, and α_j the vector of expert j's deviations from the factor pricing model. The different degrees of belief in the asset pricing model are modelled through the prior precision on the abnormal returns α_j,

$$\alpha_j \sim \mathsf{N}(0, \Sigma / k_j), \tag{19}$$

where k_j is the prior precision of expert j. The bigger k_j, the stronger the belief in the asset pricing restrictions. The experts do not have information on the K-vector of risk premia ν_j. Expert j therefore uses a flat prior on ν_j,

$$p(\nu_j) \propto 1 \tag{20}$$

In this section we make the simplification that expert j takes the factor loadings B_j as given.

All experts observe a sample of T historic returns with sample average $\bar{y}$, an N-vector. They combine their prior beliefs with the data evidence to form their posterior estimates of expected returns. The posterior mean $\hat{\mu}_j$ shrinks the sample mean of each asset towards the implied expected returns, leading the expert to estimate the quantities α_j and ν_j in (18) by

$$\hat{\nu}_j = (B_j' \Sigma^{-1} B_j)^{-1} B_j' \Sigma^{-1} \bar{y} \tag{21}$$

$$\hat{\alpha}_j = a_j (\bar{y} - B_j \hat{\nu}_j) \tag{22}$$

where

$$a_j = \frac{1}{1 + k_j / T} \tag{23}$$

is the shrinkage factor.[4] If $k_j = 0$, the expert does not have any belief in the asset pricing restrictions and the posterior mean $\hat{\mu}_j$ reduces to the unrestricted sample average $\bar{y}$.

The robust investor consults the experts and chooses the optimal robust portfolio. He receives the estimates $(\hat{\mu}_j, \Sigma)$, without knowing how the experts obtained these estimates.

[4] Equations (21)–(22) follow directly from standard Bayesian inference. In the portfolio choice literature, for example, a derivation of this result can be found in the appendix of Jorion (1986). Instead of B_j Jorion (1986) uses the vector of ones ι and shrinks towards a common mean. The shrinkage formula also appears in Wang (2005).

4.3 NESTED MODELS

We analyse the case of two experts, where expert 2 has a stricter prior than expert 1. We define a stricter prior by two conditions. First, expert 2 has more confidence in the asset pricing model,

$$k_1 \leq k_2 \tag{24}$$

Second, expert 1 considers additional factors to explain the cross section of expected returns. Technically we assume that the additional factors are orthogonal to the factors that they have in common,

$$B_1 = (B_2 \;\; \tilde{B}_1) \quad \text{with} \quad B_2' \Sigma^{-1} \tilde{B}_1 = 0 \tag{25}$$

By rotating the factors the orthogonality condition is not restrictive. What is required is that the factor loadings B_2 are in the space spanned by the loadings B_1. We refer to the case defined by the conditions (24) and (25) as nested models. Proposition 4 provides the robust portfolio for nested models.

Proposition 4. *Consider a robust investor who receives advice from two experts. The experts agree on the covariance matrix* Σ. *To the investor they report the expected returns* $\hat{\mu}_1$ *and* $\hat{\mu}_2$*, respectively. If priors are nested by satisfying conditions (24) and (25) then the robust investor follows the advice of expert 2 and* $\hat{w}_R = \frac{1}{\gamma}\Sigma^{-1}\hat{\mu}_2$.

While the proof in the appendix builds on proposition 2, proposition 4 is best understood by presenting the problem as a game in which the investor chooses a portfolio and the opponent subsequently selects the worst case objective function. The expected pay-offs, scaled by $\frac{1}{2\gamma}$, are:

	Q_1	Q_2
$\hat{w}_1$	$\hat{\rho}_{11} > \hat{\rho}_{22}$	$< \hat{\rho}_{22}$
$\hat{w}_2$	$\geq \hat{\rho}_{22}$	$\hat{\rho}_{22}$,

where $\hat{\rho}_{ii} = \hat{\mu}_i' \Sigma^{-1} \hat{\mu}_i$ are the estimated optimal squared Sharpe ratios. According to proposition 4 the robust investor chooses $\hat{w}_2$ and the opponent selects objective function Q_2. For the investor portfolio $\hat{w}_2$ is optimal under Q_2, so moving away from $\hat{w}_2$ will reduce her expected utility. The opponent has no incentive to move away from Q_2, since the orthogonality conditions imply that under Q_1 portfolio $\hat{w}_2$ attains at least the same utility.

With known covariance matrix the only difference between the objectives is in the expected returns. The objective function Q_1 uses expected returns $\hat{\mu}_1$, but the additional return components are either orthogonal to $\hat{\mu}_2$ (the additional factors) or imply even better performance than under Q_2 (less shrinkage). The objective function Q_2 will thus be the worst case scenario. Although the portfolio $\hat{w}_1$ may significantly outperform w_2 under Q_1, the robust investor is not interested. Under the orthogonality conditions the additional expected return components that $\hat{w}_1$ seeks to exploit, have no value under Q_2. Under Q_2 the portfolio $\hat{w}_1$ just adds risk. The expected performance of $\hat{w}_1$ is much more sensitive to the objective function under which it is evaluated. The same argument holds for every convex combination of $\hat{w}_1$ and $\hat{w}_2$.

The result does not depend on the covariance matrix, as Σ can be any covariance matrix and does not necessarily adhere to a factor structure. Proposition 4 can easily be extended to multiple experts. As long as all priors are nested, the robust investor will follow the advice from the expert with the most parsimonious and tightest prior.

The prior precision also affects the expected loss. Given the estimate $\hat{\mu}$ of the expert with the highest precision, the loss for the robust portfolio is:[5]

$$L(\mu_0|\hat{w}_R) = \frac{1}{2\gamma}(\mu_0 - \hat{\mu})'\Sigma^{-1}(\mu_0 - \hat{\mu}). \tag{26}$$

The expected loss of this portfolio is obtained by integrating over the sampling variation using $\bar{y} \sim \mathsf{N}(\mu_0, \Sigma/T)$. In the appendix we find:

Proposition 5. *The robust portfolio based on the prior (19) has expected loss*

$$\mathsf{E}\left[L(\mu_0|\hat{w}_R)\right] = \frac{1}{2\gamma}\left((1-a)^2\alpha'\Sigma^{-1}\alpha + \frac{1}{T}(a^2N + (1-a^2)K)\right) \tag{27}$$

where a is the shrinkage factor corresponding to the highest precision k_j among the experts, and α are the residual expected returns from the cross-sectional GLS regression of μ_0 on the factor loadings B_j.

The expected loss has two components: bias and variance. Since $a < 1$ and $K < N$, the variance term $\frac{1}{T}(a^2N + (1-a^2)K)$ is strictly less than for the unconstrained mean-variance portfolio. The bias term depends on the alpha's relative to the expert's factor model. If the pricing model is well specified, the expected returns can be perfectly explained by the factor loadings and $\alpha = 0$. 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 (27). As the confidence of the expert in the model increases, meaning larger k, the variance in (27) decreases, whereas the bias due to misspecification increases. Expected loss increases with the investment universe N and decreases with the number of observations T. When $T \to \infty$, the shrinkage factor a converges to 1, $\hat{\mu}$ goes to the true mean μ and the expected loss will approach zero.

[5] With a known covariance matrix we omit Σ as an argument in the loss function.

For the uninformative prior ($a = 1$) the expected loss (27) reduces to the well known result:[6]

$$\mathsf{E}[L(\mu_0|\hat{\mu} = \bar{y})] = \frac{1}{2\gamma}\frac{N}{T} \tag{28}$$

If only a small number of observations is available, the magnitude of the loss is tremendous. In empirically relevant situations the expected loss of the shrinkage estimators will be much lower than the expected loss of the uninformative prior. For a typical example, suppose we construct a mean-variance portfolio of $N = 25$ assets with expected returns estimated from $T = 60$ monthly observations and risk aversion $\gamma = 5$. This implies that the uninformative prior leads to a certainty equivalent expected loss of $\frac{25}{600} = 4.2\%$. Now further assume that the true Sharpe ratio is 0.6, that the fit of the cross-sectional GLS regression of sample means on the factor loadings B is as low as $R^2 = 0.3$ with $K = 3$ factors (all typical numbers for the 25 Fama-French portfolios sorted on size and book-to-market), and that the most precise prior has $k = 60$, such that $a = 0.5$. In this case the robust portfolio achieves a loss of 2.1%.[7]

The loss 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 robust portfolio choice, we can only assess this additional effect in section 5.2.

Disappointment exhibits a different trade-off between bias and variance. The disappointment of the uninformed mean-variance portfolio $\hat{w}_1$ is 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. Proposition 6 states the expected disappointment for the robust portfolio.

Proposition 6. *The robust portfolio based on the asset pricing prior (19) has disappointment*

$$\mathsf{E}\left[D_R(\hat{w}_R)\right] = \frac{1}{\gamma}\left(-a(1-a)\alpha'\Sigma^{-1}\alpha + \frac{1}{T}(a^2 N + (1-a^2)K)\right) \tag{29}$$

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 (29). Ex-post the true value of μ will be more likely

[6] See for example Kan and Zhou (2006, eq. (17)).

[7] To do the calculation note that $\alpha'\Sigma^{-1}\alpha = \mathsf{Sh}^2(1 - R^2) = (0.6)^2 \times (1 - 0.3)$.

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. For the numerical values above we find that the uninformative prior leads to an expected disappointment of 8.3%, whereas the robust investor has the much smaller out-of-sample disappointment of 1.4%.

4.4 CONFLICTING VIEWS

Next assume that two experts use conflicting, *i.e.* non-nested, models. An example we explore in more detail later is the case of international diversification, when one expert estimates expected returns using an international CAPM with beta's relative to a global market index and exchange rate risk factor, while the second expert uses a international version of the Fama-French model.

Formally, expert j ($j = 1, 2$) assumes that expected returns are proportional to B_j. As before they agree on the covariance matrix Σ. Even with two experts the general solution for the active constraints in the robust portfolio is messy. A more tractable solution emerges when we make the further simplification that the experts have the same prior confidence k in their respective asset pricing models.

Proposition 7. *Consider a robust investor who receives advice from two experts, who agree on the covariance matrix Σ, but estimate expected returns using models with different factor loadings B_1 and B_2, respectively. Both experts have the same prior confidence k in their model. Let $\bar{\rho}_{ij} = \hat{\nu}_i' B_i' \Sigma^{-1} B_j \hat{\nu}_j$ and define*

$$\bar{r}_{12} = \bar{\rho}_{12}/\sqrt{\bar{\rho}_{11}\bar{\rho}_{22}}$$
$$\bar{\kappa} = \sqrt{\bar{\rho}_{11}/\bar{\rho}_{22}}$$

If

$$\bar{r}_{12} < \frac{1}{1-a} \min\left(\frac{1}{\bar{\kappa}} - a\bar{\kappa}, \bar{\kappa} - a\frac{1}{\bar{\kappa}}\right) \tag{30}$$

the robust investor selects a strictly convex combination $\hat{w}_R = \lambda\hat{w}_1 + (1-\lambda)$ of the two portfolios with

$$\lambda = \tfrac{1}{2} + \tfrac{1}{2}\,\frac{1+a}{1-a}\,\frac{\frac{1}{\bar{\kappa}} - \bar{\kappa}}{\frac{1}{\bar{\kappa}} + \bar{\kappa} - 2\bar{r}_{12}} \tag{31}$$

Proposition 7 is a direct application of proposition 2 to the posterior means $\hat{\mu}_j$. A strictly convex combination of different portfolios $\hat{w}_1$ and $\hat{w}_2$ is more likely if either the views are very different (small or even negative $\bar{r}_{12}$), or if the prior views do not imply too different optimal Sharpe ratios ($\bar{\kappa} \approx 1$). If the first expert has a more optimistic prior than the second expert, $\bar{\kappa} > 1$, then (31) immediately implies that the investor will downweight the advice from the first expert. As before, in the special case that the priors imply the same optimal Sharpe ratio, $\bar{\kappa} = 1$, the robust portfolio will be exactly the average $\frac{1}{2}(\hat{w}_1 + \hat{w}_2)$ of the two portfolios. This is a second form of shrinkage in the robust portfolios independent of the Bayesian shrinkage factor a.

For a precise quantitative evaluation of the performance we need simulation methods, which will be introduced in the next section, where we also introduce a more general class of priors.

5. Unknown covariance matrix

In this section we study robust portfolio choice when experts have different prior opinions on both expected returns as well as the return covariance matrix. For this case we need to rely on numerical simulations. Below we first discuss the model and the multiple priors in section 5.1. Next, in section 5.2 we detail the bootstrap procedure for the numerical evaluation. To obtain relevant results, we base the simulation on two commonly used datasets. The numerical calculations are chosen to represent a case with almost nested priors (section 5.3) and another case where conflicting priors are more natural (section 5.4).

5.1 FACTOR MODEL PRIORS

All priors are based on the linear factor model

$$y_t = \alpha + Bx_t + u_t, \tag{32}$$

where y_t is the N-vector of excess returns; x_t is an K-vector of observed factors; and u_t is an error term with mean zero, $(N \times N)$ covariance matrix D and uncorrelated with x_t. Both the vector of abnormal returns α, the factor loadings B and the error covariance matrix D are unknown parameters. The factors themselves have mean ν and covariance matrix Ψ.

For the prior on the set of parameters $\Theta = (\alpha, B, D, \nu, \Psi)$ we assume the conjugate structure suggested by Pastor and Stambaugh (2000). Priors for α, B and D differ across experts. Expert j has the prior

$$\begin{aligned} D^{-1} &\sim W\left(H_j^{-1}, df_j\right) \\ \alpha \mid D &\sim N(0, D/k_j) \end{aligned} \tag{33}$$

where $W(H^{-1}, df)$ denotes the Wishart distribution with df degrees of freedom and expected covariance matrix $\mathsf{E}[D] = H/(df - K - 1)$. As in Pastor and Stambaugh (2000) the covariance matrix prior is calibrated by setting $H_j = s^2(df_j - N - 1)I$. Following an empirical Bayes approach the scaling parameter s^2 is set equal to the average of the N diagonal elements of the sample estimate of D. The prior for D approximately contains the equivalent information of df_j observations. The larger df_j 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 k_j. We assume that all experts have non-dogmatic priors, i.e. k_j is positive but finite. To elicit the prior precision k_j we work with the prior standard deviation $\sigma_{\alpha j} = s/\sqrt{k_j}$.

Expert j assumes a model with K_j factors. Given the choice of factors, the expert uses an uninformative uniform prior for B and ν, and the uninformative Wishart prior for Ψ.

All experts use the same sample period for the likelihood function. Posterior expectations for all parameters are derived in Pastor and Stambaugh (2000). Since $\hat{B}$ and $\hat{\nu}$ are independent in the posterior, the posterior mean of the expected return vector is

$$\hat{\mu}_j = \hat{\alpha}_j + \hat{B}_j \hat{\nu}_j \tag{34}$$

An analytical expression for the posterior expectation of the variance Σ is not available. We could compute it by Monte-Carlo simulation, although this is computationally intensive to do repeatedly within simulations. Instead we will use

$$\hat{\Sigma}_j = \hat{B}_j \hat{\Psi}_j \hat{B}_j' + \hat{D}_j \tag{35}$$

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.[8] Each expert j reports his posterior $(\hat{\mu}_j, \hat{\Sigma}_j)$ to the investor.

5.2 BOOTSTRAP

In this general setting a full analytical solution for the robust portfolio beyond the characterisation in theorem 1 is not available. A numerical solution for a set of expert advice can, however, be computed quickly and reliably. We perform bootstrap simulations to numerically evaluate the sampling performance of different portfolio choices $\hat{w}_j$ for the portfolios based on an uninformed prior, on one expert recommendation and the robust portfolio that takes all expert recommendations into account. For comparison purposes we also report the performance of the portfolio with short selling restrictions, cf. Jagannathan and Ma (2003), and the equally weighted portfolio of all N assets. Since short-sell constraints avoid the extreme weights that are typical in naive mean-variance portfolios, the sampling error in

[8] We verified this for one case using the 25 Fama-French portfolios sorted on Size and Book-to-Market. When $T = 60$ months the underestimation is of the order of 1%.

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 equally weighted portfolio is the benchmark advocated by De Miguel et al (2007). 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.

For the numerical experiments we first fix the true values μ_0 and Σ_0. Next we repeatedly draw random samples of length T from a much larger sample of historical returns with expected return μ_0 and variance Σ_0. Observations are drawn independently one by one with replacement. By drawing from a set of historical returns, the distribution of the returns reflects the historical distribution and is not necessarily normal.

We measure the sampling performance in terms of expected loss (16), the expected disappointment (17). Hereto we need ex-ante and ex-post evaluations of the portfolios. Ex-ante measures take the estimates from the different models at face value. Ex-post evaluations are based on the true values (μ_0, Σ_0). Appendix B details the computation of the bootstrap statistics.

5.3 NESTED MODELS

As the first set of true values we take the sample means and covariance matrix from 40 years of monthly returns of the 25 Fama-French portfolios of US firms sorted on size and book-to-market. These portfolios are among the most widely analysed in the asset pricing literature.[9] For the bootstrap we independently sample T observations from this data set, construct alternative portfolios and evaluate their performance.

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 restrictions and another with a weak prior on the restrictions. We vary the expert's beliefs by changing σ_α: $\sigma_\alpha = 0.15\%$ per month (tight prior) and $\sigma_\alpha = 0.6\%$ (loose prior). We set the prior information on the covariance matrix at $df = 30$. With two models and two priors for each model we have four rival models.

Table I provides summary statistics of the true values. 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%. Optimal portfolio weights under the true model parameters are extreme. The portfolio is also leveraged with total investment in the risky assets of 164% of invested wealth.

[9] See for example Fama and French (1996), Davis et al (2000) and the textbook treatment in Cochrane (2005). The data are available on the website of Professor Ken French, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Table I. 25 Fama-French portfolios: population statistics

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 $\gamma = 5$ given the true value of the parameters. Population parameters (μ_0, Σ_0) are equal to the sample moments for the period July 1963 to December 2002. The alternative models are the Fama-French 3-factor "FF " $(j = 1)$, and the CAPM $(j = 2)$. The last columns refer to the short selling restricted (ShSell) and equally weighted portfolios (EW).

The ex-ante performance is evaluated conditional on the parameters (μ_j, Σ_j) for each model $(j = 0, 1, 2)$. The ex-ante parameters are obtained from projecting the model parameters on the true values. For the equally weighted portfolio ex-ante statistics are not available, since the portfolio weights are not based on parameter estimates. 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	True	FF3	CAPM	ShSell	EW
Ex-ante expected performance					
Excess return $(w_j'\mu_j)$	4.62	1.09	0.16	0.68	
Standard deviation $((w_j'\Sigma_j w_j)^{1/2})$	9.61	4.66	1.82	3.68	
Utility $(Q_j(w_j))$	2.31	0.54	0.08	0.34	
Ex-post realised performance					
Excess return $(w_j'\mu_0)$	4.62	1.11	0.18	0.68	0.64
Standard deviation $((w_j'\Sigma_0 w_j)^{1/2})$	9.61	4.65	1.82	3.68	5.07
Utility $(Q_0(w_j))$	2.31	0.57	0.09	0.34	−0.01
Loss $(L(\mu_0, \Sigma_0 \vert w_j))$		1.74	2.21	1.97	2.32
Disappointment $(D_j(w_j))$		−0.03	−0.01		
Portfolio characteristics					
Sum weights $(\bar{w}_j)$	1.64	0.73	0.39	0.68	1
Cross-sectional stdev	1.23	0.16	0.03	.	0

Both the CAPM and the Fama-French models impose restrictions on μ_0. The implied portfolios based on the asset pricing restrictions would not use leverage, but instead have positive investments in the riskfree asset. The asset pricing models are sufficiently misspecified for μ_0 to imply a substantial bias component in the expected loss. An investor with dogmatic beliefs in these models will suffer a large utility loss if expected returns could be estimated with sufficient precision.

Since under the true values the CAPM nests the Fama-French model, we expect from our proposition 4 that the robust portfolio will very often coincide with the tight CAPM prior. However, since the CAPM fits the expected returns much worse than the Fama-French 3-factor model, the robust portfolio is likely to perform worse than the Fama-French model.

The true values 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

Table II. 25 Fama-French portfolios: simulations ($T = 60$)

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 "naive"); a portfolio with short-sell restrictions using the unrestricted sample moments (ShSell); 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. For the ex-post performance the bias denotes the percentage of expected loss due to the bias defined in (B6). 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 true population parameters. Expected returns, standard deviations, expected utility, loss and disappointment are all in percent per month.

	naive	FF3		CAPM			
Prior variance (σ_α)	∞	0.15	0.60	0.15	0.60	robust	ShSell
Ex-ante expected performance							
Excess return (E_j)	24.55	3.37	14.83	0.68	6.49	0.68	1.63
Standard deviation (V_j)	21.79	8.04	17.02	3.32	11.21	3.32	5.04
Utility (Q_j)	12.28	1.69	7.42	0.34	3.24	0.34	0.81
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})	−42.79	−0.02	−11.92	0.29	−1.37	0.29	−0.17
Expected loss (L_{0j})	45.10	2.33	14.23	2.01	3.67	2.01	2.47
Bias (% of L_{0j})	6%	32%	4%	87%	7%	87%	85%
Disappointment (D_{0j})	55.07	1.71	19.34	0.05	4.61	0.05	0.98
Portfolio characteristics							
Sum weights ($\bar{w}_j$)	3.10	1.06	2.20	0.54	1.40	0.54	1.09
Cross-sectional std. dev.	6.18	0.97	3.47	0.23	1.56	0.23	0.16
Binding in robust				100%			
# Assets in portfolio	25	25	25	25	25	25	3

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. 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.

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 from the full dataset.

Table II shows the performance of the different portfolios. The first thing to note is that the robust portfolio fully coincides with the portfolio implied by the tight

CAPM prior. Proposition 4 thus also provides an accurate description of robust portfolio choice in this empirical setting with unknown covariance matrix.

Adding the sampling variation reverses the ranking of portfolios compared to the performance under the true values. The robust portfolio performs best, closely followed by the portfolios based on the tight Fama-French prior and portfolios based on equal weighting or short-sell constraints. 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. The CAPM (robust) portfolio is the only portfolio strategy that achieves positive utility. For all others it would have been better ex-post to invest exclusively in the riskfree asset.

With only five years of data estimation error is overwhelming. With loose priors the expected loss is dominated by sampling variance.[10] The tight CAPM (robust) portfolio and the short-sell restricted portfolio are the only ones where expected loss is mostly due to the bias component. Even though the true values (μ_0, Σ_0) allow substantial gains, the data are not informative enough to accurately estimate the individual asset outperformance (i.e. the α's) without prior information. Interestingly, the tight CAPM prior is able to pick up some data information about alpha's from only 60 monthly observations. The ex-post expected utility of the tight CAPM (robust) portfolio in table II is larger than the expected utility of the exact CAPM portfolio in table I.

Part of the performance differences are driven by different degrees of leverage between the portfolios. From table I 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.

Figure 3 shows the portfolio performance for different numbers of observations. It illustrates the trade-off between estimation uncertainty and model structuring. With more observations for estimation all portfolios construction methods, except the $1/N$ rule, perform better. Eventually, for the four different priors estimation variance and bias disappear both: as $T \to \infty$ expected loss converges to zero. For the short-sell constrained model a positive bias remains.

For all sample sizes the robust strategy coincides with the most restrictive of the four priors, as predicted by proposition 4. It has the lowest expected loss of all strategies until $T = 100$. Since the prior is tight, but not dogmatic, the performance improves slowly with the sample size. All other models, except the strategy with short-sell constraints imposed, show faster convergence. Since the ex-post optimal portfolio w_0 has some negative weights, the short-sell restrictions are seriously

[10] For the naive portfolio the percentage of expected loss due to bias is not zero because returns are not normally distributed. Even though the naive model does not impose restrictions or prior views on the parameters, the estimator of portfolio weights is not unbiased for w_0.

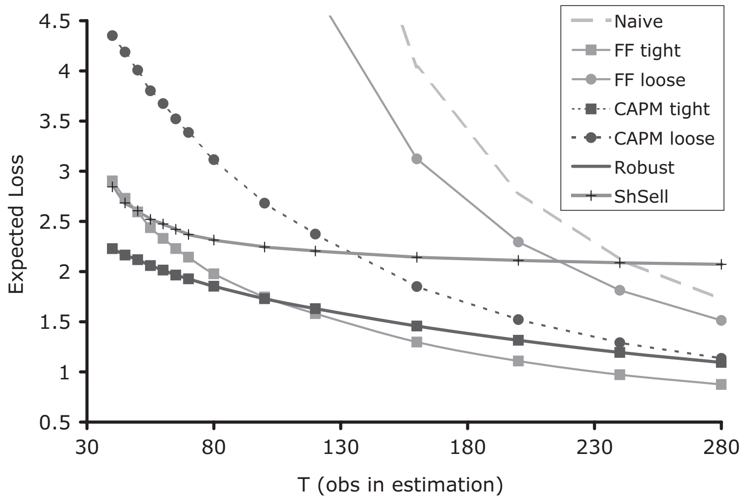

Figure 3. Expected loss as a function of T: 25 Fama-French portfolios. The figure shows the expected loss $Q_j(w_j) - Q_0(w_0)$ as a function of the number of observations for data generated under the null of the long-term historical moments of the 25 Fama-French portfolios sorted on Size and Book-to-Market. The lines correspond to portfolios based on the unrestricted sample moments (naive and with Short-Sell restrictions), the tight and loose CAPM prior, the tight and loose Fama-French prior, and the robust portfolio. Expected loss is in units of certainty equivalent returns in percent per month.

binding for large T. Not surprisingly, given the design of the experiment, the Fama-French model dominates for larger T. The prior for this model is closely related to the design of the 25 portfolios, meaning that even the tight prior has little bias. The naive strategy has almost zero bias, but a huge variance, and needs about 25 years of data for estimation to come close to the structured models. The loose priors for the CAPM and Fama-French specifications have more variance and therefore benefit more from an increasing sample size. The robust portfolio, based on a tight prior, has little variance for low T, but converges slower due to its larger bias.

Figure 4 shows the investor's disappointment for different sample sizes. For all sample sizes the robust investor is the least disappointed. Her ex-post performance is mostly very close to, and often better than, she anticipated ex-ante. Most other strategies lead to large disappointment. Given the scaling of the figure only the three best strategies are visible. The others are so much worse that even for large T they do not fit within the scale of the figure.

5.4 CONFLICTING VIEWS

For the second experiment we choose a setting where non-nested priors are more natural. True values are derived from the sample means and covariance matrix of returns on 81 portfolios from 9 countries: the domestic market portfolio, four value

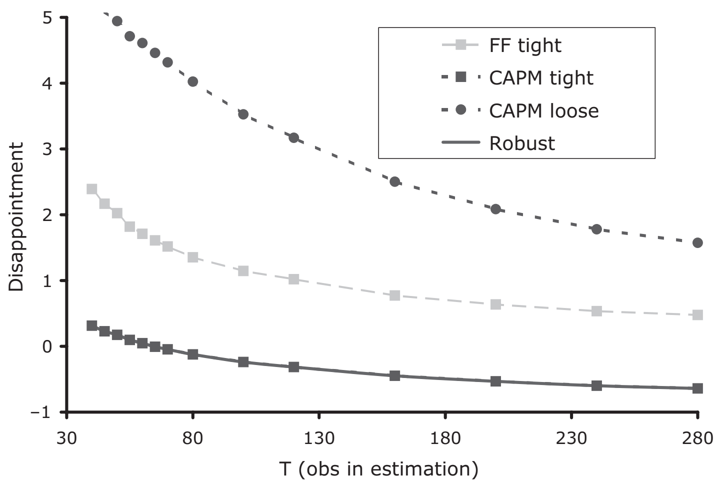

Figure 4. Disappointment as a function of *T*: 25 Fama-French portfolios. The figure shows the ex-post disappointment $Q_j(w_j) - Q_0(w_j)$ for an investor using portfolio w_j as a function of the number of observations for data generated under the null of the long-term historical moments of the 25 Fama-French portfolios sorted on Size and Book-to-Market. The lines correspond to portfolios based on the tight CAPM prior, the tight and loose Fama-French prior, and the robust portfolio. Other strategies are not visible, since their disapointment is outside the vertical scale of the figure. Disappointment is in units of certainty equivalent returns in percent per month.

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.[11] We will study the portfolio choice from the perspective of a British investor. All returns and factors have been converted to British Pounds. For the British riskfree rate we use the LIBOR rate available from Datastream. Exchange rates are also available from Datastream.

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

For this data we expect that the robust portfolio will perform very well for two reasons: (*i*) the standard asset pricing models fit the cross-section very poorly;

[11] See footnote 9. The data are from the files F-F International Countries with the specifier All 4 Data Items Not Reqd.

Table III. 81 European portfolios: population statistics

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 $\gamma = 5$ given the true value of the parameters. Population parameters (μ_0, Σ_0) are equal to the sample moments for the period January 1975 to December 2001. The alternative models are the International CAPM with exchange rate premia (CAPM), the Fama-French international 2-factor model (IFF), and a model integrating all risk factors in the previous models (ALL). See Table I for further explanatory notes.

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

(*ii*) the large number of assets leads to erratic behaviour of the naive mean-variance portfolio; (*iii*) because priors are conflicting the robust portfolio will tend to a weighted average of the simple models. The latter may predict better than the less parsimonious overall model with all factors included simultaneously.

Table III describes the European dataset in terms of population statistics. The optimal portfolio based on the true values (μ_0, Σ_0) 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 ICAPM is most conservative followed by the Fama-French model and the overall model.

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 IV 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 requires structured models for the cross section of expected returns.

The robust portfolio is indeed a combination of two conflicting recommendations. In 89% of the simulations the robust portfolio uses a strict convex combination of the ICAPM and the Fama-French model. The more general overall model is never

Table IV. 81 European portfolios: simulations ($T = 150$)

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 II and III for additional notes.

		Asset Pricing Models				
	naive	IFF	CAPM	ALL	robust	ShSell
Ex-ante expected performance						
Excess return (E_j)	51.81	1.08	0.66	1.49	0.50	1.83
Standard deviation (V_j)	31.79	4.43	3.39	5.28	2.95	5.80
Utility (Q_j)	25.90	0.54	0.33	0.74	0.24	0.92
Ex-post realised performance						
Excess return (E_{0j})	19.65	0.94	0.47	0.92	0.59	1.06
Standard deviation (V_{0j})	98.08	5.84	4.86	7.51	3.95	6.16
Utility (Q_{0j})	−248.6	−0.03	−0.21	−0.62	0.14	−0.00
Expected loss (L_{0j})	252.3	3.69	3.87	4.28	3.51	3.66
Bias (% of L_{0j})	6%	82%	85%	72%	90%	91%
Disappointment (D_{0j})	274.5	0.57	0.54	1.37	0.10	0.92
Portfolio characteristics						
Sum of weights ($\bar{w}_j$)	2.33	0.37	0.29	0.31	0.32	1.18
Cross-sectional std.dev.	9.31	0.44	0.41	0.62	0.31	0.07
Binding in robust		89%	99%	0%		
# Assets in portfolio	81	81	81	81	81	7

part of the robust portfolio. The robust portfolio also performs best. It has the lowest expected loss of all strategies, including equal weighting and short-sell constrained portfolios. The endogenous averaging of the ICAPM and IFF produces better performance than each of the model on its own or the overall model with all factors. Again, as before, the disappointment of the robust investor is minimal and lowest among all strategies.

As for the 25 Fama-French portfolios, the performance of all models improves by increasing the sample size. In figure 5 the robust strategy has the lowest loss for sample sizes up to $T = 200$, after which the International Fama-French performs slightly better. For every T it remains on average a combination of the IFF and ICAPM models. It is never, for any T, the worst performing strategy. The comprehensive model with all factors has a larger expected loss and requires a very large sample size to catch up with the robust strategy. As with the $T = 150$ case discussed above the robust investor never considers the advice based on the comprehensive general model or the naive sample moments. The naive mean-variance strategy is not visible in the scales of the figure. For small T short-sell constraints provide a valuable reduction of variance in the portfolio risk, but incur a large bias risk for larger sample sizes.

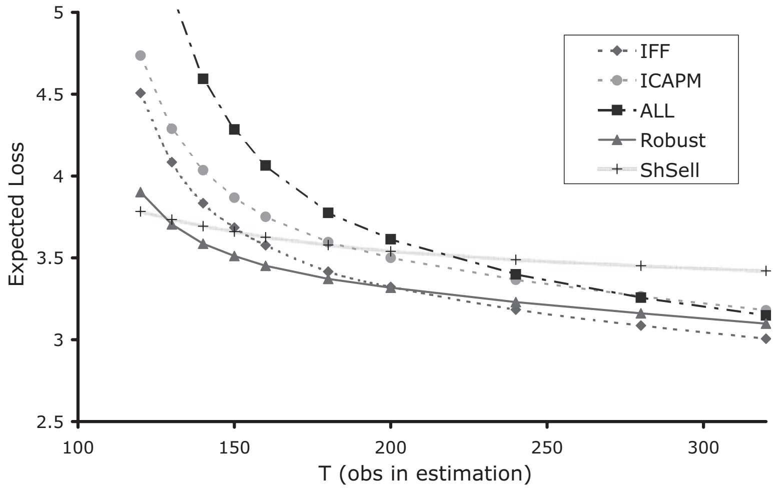

Figure 5. Expected loss as a function of T: European countries. The figure shows the expected loss $Q_j(w_j) - Q_0(w_0)$ as a function of the number of observations for data generated under the null of the long-term historical moments of 81 international equity portfolios. The lines correspond to the portfolios based on the unrestricted sample moments with Short-Sell restrictions, the International CAPM, the international Fama-French model (IFF), the combined model (ALL), and the robust portfolio. Expected loss is in units of certainty equivalent returns in percent per month.

The disappointment shown in figure 6 emphasises that the disappointment of the robust strategy is the lowest for every T. For large T is becomes negative, indicating that the ex-post performance of the portfolio is better than the ex-ante expectations. Its constituent models, IFF and ICAPM, both promise more and accordingly incur larger disappointment.

6. Conclusion

Robust mean-variance 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.

The robust strategy is most valuable when uncertainty about expected returns is large. With small samples the robust portfolio outperforms alternative rules. If sample size increases and uncertainty reduces, all strategies improve and converge, including the robust strategy. The benefits of the robust strategy are most pronounced when experts offer dispersed advice. In this case robust portfolios combine Bayesian shrinkage with multiple model sourcing.

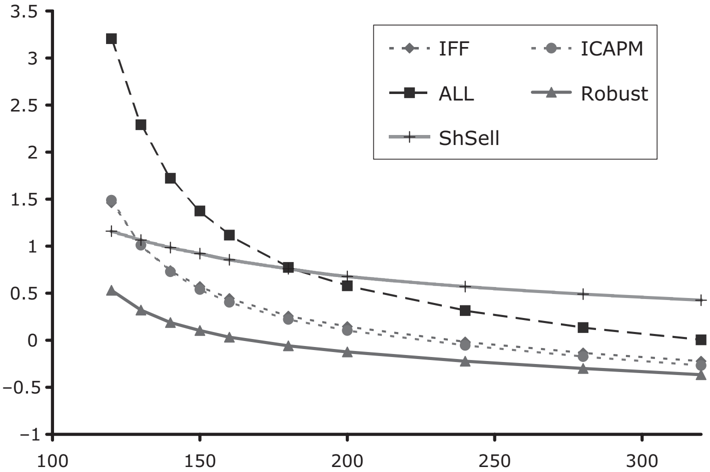

Figure 6. Disappointment as a function of *T*: European countries. The figure shows the ex-post disappointment $Q_j(w_j) - Q_0(w_j)$ for an investor using portfolio w_j as a function of the number of observations for data generated under the null of the long-term historical moments of 81 international equity portfolios. The lines correspond to the portfolios based on the unrestricted sample moments with Short-Sell restrictions, the International CAPM, the international Fama-French model (IFF), the combined model (ALL), and the robust portfolio. Disappointment is in units of certainty equivalent returns in percent per month.

From our analytical results, under the simplifying assumption of a known covariance matrix, we concluded that robust portfolios have attractive properties. First, uninformative priors do not receive weight in the robust moments. Second, if experts use nested models, then the robust investor follows the advice of the most informative prior in the most restrictive model. 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. Results from numerical simulations for 25 Fama-French portfolios and for 81 European country and value portfolios are consistent with the theoretical predictions. From the simulations we also learn that the robust portfolios are stable and well-diversified.

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. 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.

Appendix A: Proofs

PROOF OF THEOREM 1

A formal representation of the robust portfolio problem is

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

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

Kuhn-Tucker conditions for the optimal portfolio are

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

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

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

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

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.

Substituting the optimal portfolio in the objective function we find

$$Q_j(w_R) = \frac{1}{\gamma} \left(m' S^{-1} \mu_j - \frac{1}{2} m' S^{-1} \Sigma_j S^{-1} m \right) \tag{A3}$$

which shows that γ is purely a scaling factor. This implies that for the active constraints with $\lambda_j > 0$ the conditions that determine λ_j take the form

$$m' S^{-1} \mu_j - \frac{1}{2} m' S^{-1} \Sigma_j S^{-1} m = \min_i \left(m' S^{-1} \mu_i - \frac{1}{2} m' S^{-1} \Sigma_i S^{-1} m \right) \tag{A4}$$

and do not depend on γ. Hence the solution for λ_j is independent of γ and two-fund separation holds.

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

$$Q_R(w) = \begin{cases} Q_2(w) & \text{if} \quad 0 < w < w_{12}, \\ Q_1(w) & \text{if} \quad w \geq w_{12}. \end{cases} \tag{A5}$$

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

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

so that

$$w_{12} = \frac{2(\mu_1 - \mu_2)}{\gamma\left(\sigma_1^2 - \sigma_2^2\right)}. \tag{A6}$$

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

$$\frac{\mu_1}{\gamma\sigma_1^2} > \frac{2(\mu_1 - \mu_2)}{\gamma\left(\sigma_1^2 - \sigma_2^2\right)}, \tag{A7}$$

which reduces to

$$\mu_2 > \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2}\mu_1. \tag{A8}$$

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

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

Again, since we have assumed $\sigma_1^2 > \sigma_2^2$, this optimum can occur for sufficiently small μ_2. Inequalities (A8) and (A9) can not hold simultaneously since

$$\frac{2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} < \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2}. \tag{A10}$$

Therefore there also exists an interval

$$\frac{2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\mu_1 < \mu_2 \leq \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2}\mu_1, \tag{A11}$$

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

$$Q_R(w) = \min_{j=1,2} Q_j(w) = \begin{cases} Q_1(w) & \text{if } (\mu_1-\mu_2)\,w \leq 0, \\ Q_2(w) & \text{if } (\mu_1-\mu_2)\,w \geq 0. \end{cases} \tag{A12}$$

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

$$(\mu_1-\mu_2)' w_1 < 0, \tag{A13}$$

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 admissible, 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

$$w_R = \frac{1}{\gamma}\Sigma^{-1}\left(\lambda\mu_1 + (1-\lambda)\mu_2\right). \tag{A14}$$

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

$$(\mu_1-\mu_2)'\, w = 0 \tag{A15}$$

Substituting (A14) into (A15) leads to

$$\begin{aligned} 0 &= \lambda(\mu_1-\mu_2)'\Sigma^{-1}\mu_1 + (1-\lambda)(\mu_1-\mu_2)'\Sigma^{-1}\mu_2 \\ &= \lambda(\rho_{11}-\rho_{12}) + (1-\lambda)(\rho_{12}-\rho_{22}) \end{aligned} \tag{A16}$$

which is linear in λ and has the solution stated in (12) in the proposition.

PROOF OF PROPOSITION 3

Under the condition 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 (16) with true expected returns μ_0 we obtain

$$\begin{aligned}2\gamma L(\mu_0|w_R) &= (\mu_0 - \gamma\Sigma w_R)'\,\Sigma^{-1}\,(\mu_0 - \gamma\Sigma w_R)\\ &= (\mu_0 - \lambda\mu_1 - (1-\lambda)\mu_2)\,\Sigma^{-1}\,(\mu_0 - \lambda\mu_1 - (1-\lambda)\mu_2)\\ &= \rho_{00} + \lambda^2\rho_{11} + (1-\lambda)^2\rho_{22}\\ &\quad - 2\lambda\rho_{01} - 2(1-\lambda)\rho_{02} + 2\lambda(1-\lambda)\rho_{12} \end{aligned} \tag{A17}$$

The analogous expression for the utility loss of expert 1 is

$$2\gamma L(\mu_0|w_1) = \rho_{00} + \rho_{11} - 2\rho_{01} \tag{A18}$$

Subtracting (A18) from (A17) gives

$$\begin{aligned}2\gamma\,(L(\mu_0|w_1) - L(\mu_0|w_R)) &= (1-\lambda^2)\rho_{11} - (1-\lambda)^2\rho_{22} - 2\lambda(1-\lambda)\rho_{12}\\ &\quad - 2(1-\lambda)\rho_{01} + 2(1-\lambda)\rho_{02}\end{aligned} \tag{A19}$$

For the robust investor to have lower loss, this must be positive. Divide (A19) by $1-\lambda$, require the result to be positive and solve for $\rho_{02} - \rho_{01}$,

$$\begin{aligned}2(\rho_{02} - \rho_{01}) &> -(\lambda+1)\rho_{11} + (1-\lambda)\rho_{22} + 2\lambda\rho_{12}\\ &= -\rho_{11} + \rho_{22} - \lambda(\rho_{11} + \rho_{22} - 2\rho_{12})\\ &= \rho_{12} - \rho_{11}\end{aligned} \tag{A20}$$

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

By proposition 2 it is sufficient to show that

$$\hat{\mu}_2'\Sigma^{-1}\hat{\mu}_2 \leq \hat{\mu}_2'\Sigma^{-1}\hat{\mu}_1 \tag{A21}$$

We will prove this inequality in two steps. Define

$$\hat{\mu}_2^* = \frac{a_1}{a_2}\hat{\alpha}_2 + B_2\hat{\nu}_2, \tag{A22}$$

which is the estimator of the means that uses the factor loadings B_2 of expert 2 but the prior precision k_1 of expert 1. To prove (A21) we will prove the two inequalities

$$\hat{\mu}_2\Sigma^{-1}\hat{\mu}_2 \leq \hat{\mu}_2^*\Sigma^{-1}\hat{\mu}_2 \leq \hat{\mu}_1\Sigma^{-1}\hat{\mu}_2 \tag{A23}$$

For the proof of this and the following propositions it will be convenient to introduce some notation related to GLS regressions. If B is a matrix of factor loadings the associate projection matrix is defined as

$$P = B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1} \tag{A24}$$

with the property $P'\Sigma^{-1}P = \Sigma^{-1}P$. We also define the residual matrix

$$M = \Sigma^{-1} - \Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1} \tag{A25}$$

which is positive semidefinite and has the useful property $M\Sigma M = M$. When factor loadings are denoted B_i, we will write the associated projection and residual matrices as P_i and M_i, respectively. In projection matrix notation expected returns are

$$\begin{aligned}\hat{\mu}_i &= \hat{\alpha}_i + B_i\hat{\nu}_i \\ &= a_i(I - P_i)\bar{y} + P_i\bar{y} \\ &= (a_i I + (1 - a_i)P_i)\,\bar{y}.\end{aligned} \tag{A26}$$

Similarly we find

$$\hat{\mu}_2^* = (a_1 I + (1 - a_1)P_2)\bar{y} \tag{A27}$$

Starting with the first inequality in (A23), note that

$$\hat{\mu}_2^* - \hat{\mu}_2 = (a_1 - a_2)\,(I - P_2)\,\bar{y} \tag{A28}$$

and thus

$$\begin{aligned}(\hat{\mu}_2^* - \hat{\mu}_2)'\Sigma^{-1}\hat{\mu}_2 &= (a_1 - a_2)\bar{y}'(I - P_2)'\Sigma^{-1}\,(a_2\bar{y} + (1 - a_2)P_2\bar{y})\,, \\ &= (a_1 - a_2)a_2\bar{y}'M_2\bar{y}\end{aligned} \tag{A29}$$

As $a_1 \geq a_2$ and M_2 is positive semi-definite, the right-hand side of (A29) is non-negative. This proves the first inequality in (A23).

For the second inequality we use the partitioning $B_1 = (B_2\tilde{B}_1)$. Then, due to the assumed orthogonality $B_2'\Sigma^{-1}\tilde{B}_1 = 0$, the projection on B_1 is simply the sum of two separate projections,

$$P_1 = P_2 + \tilde{P}_1 \tag{A30}$$

and

$$\hat{\mu}_1 = (a_1 I + (1 - a_1)(P_2 + \tilde{P}_1))\bar{y} \tag{A31}$$

Hence

$$\hat{\mu}_1 - \hat{\mu}_2^* = (1 - a_1)\tilde{P}_1\bar{y}. \tag{A32}$$

This implies

$$\begin{aligned}(\hat{\mu}_1 - \hat{\mu}_2^*)'\Sigma^{-1}\hat{\mu}_2 &= (1 - a_1)\bar{y}'\tilde{P}_1'\Sigma^{-1}\,(a_2 I + (1 - a_2)P_2)\,\bar{y}, \\ &= (1 - a_1)a_2\bar{y}'\tilde{P}_1'\Sigma^{-1}\bar{y} \\ &= (1 - a_1)a_2\bar{y}'\tilde{P}_1'\Sigma^{-1}\tilde{P}_1\bar{y}\end{aligned} \tag{A33}$$

The last expression is non-negative because the quadratic form in $\bar{y}$ is non-negative and $a_1, a_2 \in [0, 1]$. This proves the second inequality in (A23).

PROOF OF PROPOSITION 5

This proof and the following ones distinguish between bias and variance and rely on the decomposition

$$\bar{y} = \mu_0 + \eta, \tag{A34}$$

with η an error term with zero mean and covariance matrix Σ/T.

For the expected loss we evaluate

$$2\gamma \mathsf{E}[L(\mu_0|\hat{w}_R)] = \mathsf{E}[(\hat{\mu} - \mu_0)'\Sigma^{-1}(\hat{\mu} - \mu_0)] \tag{A35}$$

Start by writing

$$\begin{aligned} \hat{\mu} - \mu_0 &= (aI + (1-a)P)\,\bar{y} - \mu_0 \\ &= -(1-a)(I - P)\,\mu + (aI + (1-a)P)\,\eta \\ &= -(1-a)\alpha + (aI + (1-a)P)\,\eta, \end{aligned} \tag{A36}$$

Using the decomposition (A36) the expectation of the quadratic form in (A35) will have two terms. The first is the quadratic term in α,

$$(1-a)^2\alpha'\Sigma^{-1}\alpha, \tag{A37}$$

which defines the bias. The second quadratic term provides the variance as

$$\begin{aligned} &\mathsf{E}[\eta'(aI + (1-a)P)'\,\Sigma^{-1}(aI + (1-a)P)\,\eta] \\ &= \frac{1}{T}\mathrm{tr}((aI + (1-a)P)'\Sigma^{-1}(aI + (1-a)P)\Sigma) \\ &= \frac{1}{T}\mathrm{tr}((aI + (1-a)P)'(aI + (1-a)\Sigma^{-1}P\Sigma)) \\ &= \frac{1}{T}(\mathrm{tr}(a^2 I) + a(1-a)\mathrm{tr}(P') + (1-a)\mathrm{tr}(P)) \\ &= \frac{1}{T}(a^2 N + (1-a^2)K) \end{aligned} \tag{A38}$$

Adding the expressions for bias and variance and substituting in (A35) immediately leads to (27) in the proposition.

PROOF OF PROPOSITION 6

The derivation of the expected disappointment uses similar calculations as for proposition 5,

$$\mathsf{E}[D_R(\hat{w}_R)] = \mathsf{E}\,[Q_R(\hat{w}_R) - Q_0(\hat{w}_R)] = \frac{1}{\gamma}\mathsf{E}[(\hat{\mu} - \mu_0)'\Sigma^{-1}\hat{\mu}_0)] \tag{A39}$$

Using (A36) we have that

$$(\hat{\mu}-\mu_0)'\Sigma^{-1}\hat{\mu} = (-(1-a)\alpha + (aI + (1-a)P)\eta)'\,\Sigma^{-1} \times\ (a\alpha + P\mu_0 + (aI + (1-a)P)\eta) \tag{A40}$$

Taking expectations we obtain

$$\begin{aligned}\mathsf{E}[(\hat{\mu}-\mu)'\Sigma^{-1}\hat{\mu}] &= -(1-a)\alpha'\Sigma^{-1}(a\alpha + P\mu)\\ &\quad + \mathsf{E}[\eta'(aI + (1-a)P)'\Sigma^{-1}(aI + (1-a)P)\eta]\\ &= -a(1-a)\alpha'\Sigma^{-1}\alpha + \frac{1}{T}\mathrm{tr}(a^2 I + (1-a^2)P)\\ &= -a(1-a)\alpha'\Sigma^{-1}\alpha + \frac{1}{T}(a^2 N + (1-a^2)K)\end{aligned} \tag{A41}$$

Dividing by γ and substituting in (A39) gives the result in the proposition.

PROOF OF PROPOSITION 7

According to proposition 2 we will have a strictly convex combination of the two portfolios if

$$\hat{\rho}_{12} < \min(\hat{\rho}_{11}, \hat{\rho}_{22}) \tag{A42}$$

Using the expression for $\hat{\mu}_j$ in (A26) we find

$$\begin{aligned}\hat{\rho}_{ij} &= \bar{y}'(aI + (1-a)P_i)'\Sigma^{-1}(aI + (1-a)P_j)\bar{y}\\ &= \bar{y}'(a^2\Sigma^{-1} + a(1-a)\Sigma^{-1}(P_i + P_j) + (1-a)^2 P_i'\Sigma^{-1}P_j)\bar{y}\\ &= a^2\bar{\rho} + a(1-a)\bar{\rho}_{ii} + a(1-a)\bar{\rho}_{jj} + (1-a)^2\bar{\rho}_{ij},\end{aligned} \tag{A43}$$

where $\bar{\rho} = \bar{y}'\Sigma^{-1}\bar{y}$. The last equality in (A43) uses the definition of $\bar{\rho}_{ij}$ and the projection matrix property $\Sigma^{-1}P = P'\Sigma^{-1}P$. Inequality (A42) is thus equivalent to

$$\begin{gathered}(1-a)^2\bar{\rho}_{12} + a(1-a)\bar{\rho}_{11} + a(1-a)\bar{\rho}_{22} < \min((1-a^2)\bar{\rho}_{11}, (1-a^2)\bar{\rho}_{22})\\ \Updownarrow\\ (1-a)\bar{\rho}_{12} + a\bar{\rho}_{11} + a\bar{\rho}_{22} < \min((1+a)\bar{\rho}_{11}, (1+a)\bar{\rho}_{22})\\ \Updownarrow\\ (1-a)\bar{\rho}_{12} < \min(\bar{\rho}_{11} - a\bar{\rho}_{22}, \bar{\rho}_{22} - a\bar{\rho}_{11})\end{gathered} \tag{A44}$$

Analogously to the normalisation that transforms (12) to (13) in proposition 2 we can divide both sides of (A44) by $\sqrt{\bar{\rho}_{11}\bar{\rho}_{22}}$ and use the definitions of $\bar{r}_{12}$ and $\bar{\kappa}$ in the proposition to obtain

$$(1-a)\bar{r}_{12} < \min\left(\bar{\kappa} - a\frac{1}{\kappa}, \frac{1}{\kappa} - a\bar{\kappa}\right), \tag{A45}$$

which directly leads to the inequality in the proposition.

Using proposition 2 once more gives the expression for λ. In particular,

$$\begin{aligned}\hat{\rho}_{jj} - \hat{\rho}_{ij} &= a(1-a)\bar{\rho}_{jj} + (1-a)^2\bar{\rho}_{jj} - a(1-a)\bar{\rho}_{ii} - (1-a)^2\rho_{ij} \\ &= (1-a)(\bar{\rho}_{jj} - a\bar{\rho}_{ii} - (1-a)\bar{\rho}_{ij}) \end{aligned} \tag{A46}$$

and thus

$$\begin{aligned}\hat{\rho}_{11} + \hat{\rho}_{22} - 2\hat{\rho}_{12} &= (\hat{\rho}_{22} - \hat{\rho}_{12}) + (\hat{\rho}_{11} - \hat{\rho}_{12}) \\ &= (1-a)^2 (\bar{\rho}_{11} + \bar{\rho}_{22} - 2\bar{\rho}_{12}) \end{aligned} \tag{A47}$$

Substituting (A46) and (A47) in (12) gives

$$\lambda = \frac{\bar{\rho}_{22} - a\bar{\rho}_{11} - (1-a)\bar{\rho}_{12}}{(1-a)(\bar{\rho}_{11} + \bar{\rho}_{22} - 2\bar{\rho}_{12})} \tag{A48}$$

Again this can be rewritten in terms of $\bar{r}_{12}$ and $\bar{\kappa}$ by dividing both numerator and denominator by $\sqrt{\bar{\rho}_{11}\bar{\rho}_{22}}$,

$$\begin{aligned}\lambda &= \frac{1/\bar{\kappa} - a\bar{\kappa} - (1-a)\bar{r}_{12}}{(1-a)(\bar{\kappa} + 1/\bar{\kappa} - 2\bar{r}_{12}} \\ &= \tfrac{1}{2} + \frac{1/\bar{\kappa} - a\bar{\kappa} - (1-a)\bar{r}_{12}}{(1-a)(\bar{\kappa} + 1/\bar{\kappa} - 2\bar{r}_{12}} - \tfrac{1}{2} \\ &= \tfrac{1}{2} + \tfrac{1}{2}\frac{1+a}{1-a}\ \frac{1/\bar{\kappa} - \kappa}{1/\bar{\kappa} + \bar{\kappa} - 2\bar{r}_{12}}, \end{aligned} \tag{A49}$$

which proves the last part of the proposition.

Appendix B: Bootstrap statistics

For each of $\ell = 1, \ldots, L$ bootstrap samples the experts compute their posterior means $\hat{\mu}_j^\ell$ and $\hat{\Sigma}_j^\ell$. The investor receives these estimates and constructs the optimal portfolios $\hat{w}_j^\ell$ conditional on each expert's opinion, and also the optimal robust portfolio $\hat{w}_R^\ell$. For the robust portfolio, $j = R$, the expected return and variance are defined through the optimality conditions in theorem 1,

$$\begin{aligned}\hat{\mu}_R^\ell &= \sum_j \hat{\lambda}_j^\ell \hat{\mu}_j^\ell \\ \hat{\Sigma}_R^\ell &= \sum_j \hat{\lambda}_j^\ell \hat{\Sigma}_j^\ell \end{aligned} \tag{B1}$$

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

$$E_j = \frac{1}{L}\sum_{\ell}\left(\hat{w}_j^{\ell}\right)'\hat{\mu}_j^{\ell},$$
$$V_j^2 = \frac{1}{L}\sum_{\ell}\left(\hat{w}_j^{\ell}\right)'\hat{\Sigma}_j^{\ell}\hat{w}_j^{\ell}, \tag{B2}$$

and the ex-ante expected utility as

$$\begin{aligned} Q_j &= E_j - \frac{\gamma}{2}V_j^2 \\ &= \frac{1}{2\gamma}\,\frac{1}{L}\sum_{\ell}\left(\hat{\mu}_j^{\ell}\right)'\left(\hat{\Sigma}_j^{\ell}\right)^{-1}\hat{\mu}_j^{\ell} \end{aligned} \tag{B3}$$

The second line in (B3) follows from substituting the optimal portfolios $\hat{w}_j^{\ell} = (\hat{\Sigma}_j^{\ell})^{-1}\hat{\mu}_j^{\ell}/\gamma$ in (B2).

For the ex-post performance we evaluate the mean and variance under the true parameters as

$$E_{0j} = \frac{1}{L}\sum_{\ell}\left(\hat{w}_j^{\ell}\right)'\mu_0 = \bar{w}_j'\mu_0,$$
$$V_{0j}^2 = \frac{1}{L}\sum_{\ell}\left(\hat{w}_j^{\ell}\right)'\Sigma_0\hat{w}_j^{\ell} = \bar{w}_j'\Sigma_0\bar{w}_j + \operatorname{tr}\left(S_w\Sigma_0\right), \tag{B4}$$

where

$$\bar{w}_j = \frac{1}{L}\sum_{\ell}\left(\hat{w}_j^{\ell}\right)$$
$$S_w = \frac{1}{L}\sum_{\ell}\left(\hat{w}_j^{\ell} - \bar{w}_j\right)\left(\hat{w}_j^{\ell} - \bar{w}_j\right)'$$

are the average and covariance matrix of the portfolio weights over the simulations. The subscript 0 in (B4) indicates evaluation under the true values. The quantities E_{0j} and V_{0j} define the ex-post utilities

$$Q_{0j} = E_{0j} - \frac{\gamma}{2}V_{0j}^2 \tag{B5}$$

Expected loss is calculated as

$$\begin{aligned} L_{0j} &= Q_0 - Q_{0j} \\ &= \frac{\gamma}{2}\,\frac{1}{L}\sum_{\ell}\left(\left(\hat{w}_j^{\ell}\right)' - w_0\right)\Sigma_0\left(\left(\hat{w}_j^{\ell}\right)' - w_0\right) \\ &= \frac{\gamma}{2}((\bar{w}_j - w_0)'\Sigma_0(\bar{w}_j - w_0) + \operatorname{tr}(\Sigma_0 S_w)), \end{aligned} \tag{B6}$$

where Q_0 is the utility for the optimum portfolio under the true values. The second and third line in (B6) define expected loss analogously to (7) and the decomposition of proposition 5. As in (27) the first term represents the bias while the second term is the variance component of the expected loss.

We estimate the expected disappointment as the difference between what an expert promises and what could have been achieved:

$$D_{0j} = Q_j - Q_{j0} = L_{0j} + (Q_j - Q_0) \tag{B7}$$

As characteristics of the portfolios we calculate the fraction of wealth invested in risky assets and the cross-sectional standard deviation of portfolio weights:

$$\begin{aligned} \mathsf{E}w_j &= \frac{1}{L}\sum_{\ell} \iota' \hat{w}_j^{\ell} \\ s^2(w_j) &= \frac{1}{L}\sum_{\ell} \left(\hat{w}_j^{\ell}\right)' \hat{w}_j^{\ell} - \mathsf{E}w_j^2 \end{aligned} \tag{B8}$$

References

Avramov, D. (2002) Stock Return Predictability and Model Uncertainty, *Journal of Financial Economics* **64**, 423–458.

Black, F. and Litterman, R. (1992) Global Portfolio Optimization, *Financial Analysts Journal* **48**, 28–43.

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

Cochrane, J. (2005) *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. (2007) Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?. *Review of Financial Studies* forthcoming.

Fama, E. and French, K. (1996) Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* **51**, 55–84.

Garlappi, L., Uppal, R. and Wang, T. (2007) Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach, *Review of Financial Studies* **20**, 41–81.

Gilboa, I. and Schmeidler, D. (1989) Maxmin Expected Utility Theory with Non-unique Prior, *Journal of Mathematical Economics* **18**, 141–153.

Goldfarb, D. and Iyengar, G. (2003) Robust Portfolio Selection Problems, *Mathematics of Operations Research* **28**, 1–38.

Jagannathan, R. and Ma, T. (2003) Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps, *Journal of Finance* **58**, 1651–1683.

Jobson, J. and Korkie, B. (1981) Putting Markowitz Theory to Work, *Journal of Portfolio Management* **7**, 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**, 279–292.

Kan, R. and Zhou, G. (2006) Optimal Portfolio Choice with Parameter Uncertainty, *Journal of Financial and Quantitative Analysis* **42**, 621–656.

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

Ledoit, O. and Wolf, M. (2003) Improved Estimation of the Covariane Matrix of Stock Returns with an Application to Portfolio Selection, *Journal of Empirical Finance* **10**, 603–621.

Michaud, R. (1989) The Markowitz optimization enigma: Is optimized optimal?, *Financial Analysts Journal* **45**, 31–42.

Pastor, L. and Stambaugh, R. (2000) Comparing Asset Pricing Models: An Investment Perspective *Journal of Financial Economics* **56**, 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**, 1591–1621.

Wang, Z. (2005) A Shrinkage Approach to Model Uncertainty and Asset Allocation, *Review of Financial Studies* **18**, 673–705.