Determining an Efficient Frontier in a Stochastic Moments Setting

Christian Johannes Zimmer*\\nBeat Matthias Niederhauser**

Resumo
Analisamos o problema de otimização de carteiras no caso que o investidor tem incerteza sobre a distribuição dos retornos dos ativos. Formulamos o problema de uma forma geral que usa o espaço produto entre as distribuições dos retornos e as distribuições de seus parâmetros. Mostramos como algumas das abordagens existentes se encaixam na formulação geral e propomos usar funções de penalização para resolver o problema. Concentramos-nos em duas propostas que favorecem uma característica com importância prática: a estabilidade das composições sob variações nos dados de entrada. Tendo em vista os altos custos de transação e a falta de liquidez em alguns mercados, esta característica é de grande importância na realidade brasileira. Finalmente, mostramos num exemplo real do mercado brasileiro que as propostas realmente aumentam a estabilidade o que indica que sejam alternativas interessantes cujo comportamento detalhado no mercado deve ser analisado em pesquisas futuras.

Abstract
We analyze the problem of portfolio optimization under uncertainty in the assets’ return distributions. After characterizing the problem using a general formulation involving the product space of the return distribution with the parameter distribution, we propose the use of penalty functions to solve the resulting program. The connection to some important existing approaches is shown, and we then focus on two specific proposals with an important practical feature: the stability of the resulting portfolio composition under changing input variables. With high transaction costs and missing liquidity in some Brazilian markets, this stability feature is of great practical relevance. Finally, we show with an example from the Brazilian market that the penalty function approach does indeed increase stability, and seems to be a promising alternative whose long-range performance should be analyzed.

Key words: portfolio optimization; model risk; penalty function; stable portfolios; resampling.

JEL codes: G11; G15.

1. Introduction

Assuming that economic agents prefer more to less, it is reasonable to postulate that they like to maximize the expected return while controlling the risk of the opportunities encountered. The traditional framework of Markowitz Mean-Variance trade-off (Markowitz, 1952) assumes that the agent knows the future distributions

*Banco Itaú/AMC/DAC/SPQ e USP/FEA. E-mail: christian.zimmer@itaú.com.br
**USP/IME. Supported by FAPESP through Grant Nr. 00/05134-5, present address: Banco Itaú/AMC/DAC/SPQ. E-mail: beat.niederhauser@itaú.com.br
of the assets and that these are normally distributed. With that knowledge, he determines the optimal portfolio composition with varying degrees of risk aversion. Applying this approach in practice leads to results that leave many asset managers with a certain discomfort: the optimizer is merciless and allocates the resources in extreme ways, in what has been called “error optimization” (see Michaud (1989)).

The most famous problem one encounters is the so-called substitution problem, where two assets with the same risk have slightly different expected returns. The optimizer in this case gives all weight to the higher expected return leading to extreme changes in the portfolio composition if expectations change slightly.

Several alternatives were proposed to avoid unrealistic portfolio allocations and we will discuss some of them briefly in Section 2. The tenor of the proposals on how to create realistic portfolio compositions is to take into account that the values used for risk and return are just estimates and should not be seen as fixed — they are stochastic. In fact, the whole future distribution might be seen as random, which means that we have to work with a class of possible future distributions, sometimes called “scenarios”. In this paper we formalize the idea of interpreting the expected return and the expected risk as random variables from a probabilistic point of view. However, we will restrict ourselves to distributions whose moments are specified by a parameter \( \theta \) coming from a set \( \Omega_\theta \).

Most part of the literature (see Section 2) is concerned with estimation of the risk and return variables. The reason for minimizing estimation risk is the fact that estimated returns are biased upwards whereas estimated risk is biased downwards, Jobson and Korkie (1980). The real-life behavior of asset managers seems to conform to with that point of view. The general impact of uncertainty in the estimations were reported by several authors: Dickenson (1979), Frankfurter et al. (1971), Jobson and Korkie (1980), Barry (1974), Bawa et al. (1979), Chopra and Ziemba (1993), Brown (1979) and Kalymon (1971).

It should be noted that the misestimation of returns (see Best and Grauer (1991), for a detailed analysis) is much more severe than that of the variances (when used as the risk measure), while the problem of a wrong correlation can be seen as intermediate, see Merton (1980). The problem of misspecification of the asset manager’s utility function is not that important either, as long as the level of the Arrow-Pratt absolute risk aversion is not misspecified (see Kallberg and Ziemba (1984)).

The procedure in most of the approaches is to estimate an alternative parameter \( \hat{\theta} \in \Omega_\theta \) by using the expected value of a loss function \( L \):

\[
\hat{\theta}^* = \arg\min_{\hat{\theta}} E_Q \left[ L \left( \hat{\theta}, \theta \right) \right],
\]

where \( Q \) is a probability measure on the measurable space \( (\Omega_\theta, \mathcal{F}_\theta) \) interpreted as the uncertainty about the parameter \( \theta \). One then applies the optimization to the resulting distribution \( P|\theta^* \).
\[
\max_{\pi} E_{\mathcal{F} | \theta} \left[ U \left( \langle X, \pi \rangle \right) \right],
\]

where \( \pi = (\pi_1, ..., \pi_n)' \) is a trading strategy and \( S = (S_1, ..., S_n)' \) the asset returns.\(^1\) Consider now \( \delta \in \mathbb{R} \) a constant representing a (correctly normalized) factor of aversion against uncertainty, and define a function

\[
B_\delta : \mathbb{R}^n \times \Omega_\theta \rightarrow \mathbb{R},
\]

\[
(\pi, \theta) \mapsto B_\delta (\pi, \theta).
\]

Then, instead of following the (not necessarily optimal) two-stage procedure of Equation (1), we propose to solve directly:

\[
\max_{\pi \in \mathcal{A}(X)} \int_{\Omega_\theta} E_{\mathcal{F} | \theta} \left[ U \left( \langle S, \pi \rangle \right) \right] - B_\delta (\pi, \theta) \, dQ (\theta),
\]

where \( \mathcal{A}(X) \) denotes the set of admissible strategies with respect to \( X \), the portfolio wealth, see (Karatzas and Shreve (1991), Section 5.8).

The advantage is that a discussion on which criterion to choose can be reduced to either a discussion on which utility function for uncertainty aversion to use or, alternatively, which penalty function to use. A concretization of the general problem is given by using penalty approaches. Via convex analysis results the penalty form of the objective function may sometimes be written in terms of utility functions, too. This means our proposals can be seen as approaches that use a utility function adapted for the uncertainties in the distribution parameters.

After a discussion of some of the main ideas proposed in the literature in the next section we move over to the presentation of our own proposals in Section 3. Section 4 is dedicated to a comparison of some of the key approaches in a specific market situation. We close the paper in Section 5 with a discussion of the results.

2. Existing Approaches to Incorporating Risk and Return Uncertainty in Portfolio Choice

In this section we will review the existing approaches in order to have a starting point for our analysis. The main focus of the existing approaches is the distortion of the initial distributions of the assets of the investment universe and the choice of a convenient one. The traditional optimizer is applied to the chosen alternative distribution.

One of the first propositions to deal with the unrealistic results caused by a direct application of the mean-variance optimizer was to impose limits on the allocation of some assets in the portfolio (e.g. Frost and Savarino (1988), Eichhorn

\(^1\)When we want to emphasize the dependence on the parameter, we also write \( S (\theta) = (S_1 (\theta), ..., S_n (\theta))' \).
et al. (1998), Grauer and Shen (2000)). The main effect is that the limits become binding which means, in the extreme case, that the asset manager imposes the portfolio composition with his limits. He then should ask himself why he wanted to use a portfolio optimization if he already knew what composition to use. The severe problem is that in a deterministic setting, portfolio performance is in general reduced by imposing limiting weights. In a stochastic setting for the return moments, they point out that the advantage of using constraints is that they may help to limit estimation errors. The principal idea is that the limits force the investor to diversify more and thus avoid large parts of the investment made in assets with misestimated distributions. The reason behind this is that returns with high estimation values are often overestimated whereas those with low values are underestimated, i.e., the estimations are biased. But, following this argument, as there is no obvious condition on what should be the limits for each asset, one might argue that the limit should be as low as possible, which would lead the asset manager to hold an equally weighted portfolio. This is only optimal when all assets have the same expected risk and return or when estimation risk is so high that the expected values seem to have this structure. Thus, for this approach to work correctly the level of constraint should depend on the asset manager’s aversion to uncertainty in the risk and return variables. Unfortunately, there exists no valuable method to determine the corresponding level yet. Neither does this approach allow one to decide which alternative values for the risk and return variables to choose. In the light of our general analysis to be developed, this means that it is neither clear which alternative measure for the efficient frontier to choose nor which utility function of aversion against uncertainty the asset manager assumes.

Giving the same limits to each asset, this approach is close to the equal weight rule analyzed, for example, by Brown (1979). Here one tries to

\[
\max_{\theta \in \Omega_\theta} \left( \max_{\pi_{\theta} \in A(\lambda_{\theta})} \left( E_{P_{\theta}} \left[ U \left( \langle S, \pi_{\theta} \rangle \right) \right] - \| \pi_{\theta} \|_2^2 \right) \right).
\]

This approach allows one to directly determine the optimal portfolio when the parameter \( \theta \) is specified in the traditional Markowitz mean-variance framework

\[
\pi_{\theta}^* = \arg \max_{\pi_{\theta} \in A(\lambda_{\theta})} \left( E_{P_{\theta}} \left[ U \left( \langle S, \pi_{\theta} \rangle \right) \right] - \lambda \| \pi_{\theta} \|_2^2 \right)
= - \arg \min_{\pi_{\theta}} \langle \pi_{\theta}, \Sigma_{\theta} \pi_{\theta} \rangle + \lambda \langle \pi_{\theta}, \pi_{\theta} \rangle - \lambda_1 \left( \langle \pi_{\theta}, \mu_{\theta} \rangle - \bar{\mu} \right)
- \lambda_2 \left( \langle \pi_{\theta}, 1 \rangle - 1 \right)
\]

where \( \Sigma_{\theta} \) and \( \mu_{\theta} \) are the covariance matrix and the mean vector, respectively, under the realization \( \theta \in \Omega_\theta \) and no further restrictions than the cone restriction

\[
\text{with } \epsilon = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}
\]

is assumed. The solution in this case is given by
\[ \pi_\theta^* = \frac{\hat{\Sigma}_\theta^{-1} \mu_\theta}{\langle e, \hat{\Sigma}_\theta^{-1} \mu_\theta \rangle} \]

with \( \hat{\Sigma}_\theta = (\Sigma_\theta + \lambda \cdot I) \).\(^2\) For some further discussion on this idea, see also the equal-weight penalty approach presented in Subsection 3.3.

Having accepted that the major problem is the existence of uncertainty within the distribution parameters (or even in the distribution as it is) one may be willing to take a look at a well-known statistical tool: the \textit{Loss-function}.\(^3\) The typical approach is to incorporate the impact of the existence of uncertainty about asset distributions by using a function from which optimal alternative asset distributions can be derived, and the uncertainty-adjusted optimal portfolio is exactly that one, which maximizes expected utility with respect to this new distribution.

The loss function measures the impact of misspecification by calculating how much utility the asset manager loses if he uses the misspecified values. The loss can be expressed in utility or in cash equivalent units. Chopra and Ziemba (1993) analyze the impact by calculating the cash equivalent loss assuming an exponential utility function for wealth \( W \) and a multidimensional normal distribution. The cash (or certainty) equivalent is the deterministic value (in benchmark) that leaves the asset manager equivalent of choosing the deterministic value or the stochastic payoff. It is given by the inverse of the utility function of the expected utility value:

\[ CE_\theta = U^{-1} \left( E_{\mathbb{P}|\theta} \left[ U \left( \langle S, \pi_\theta^* \rangle \right) \right] \right). \]

The loss of cash equivalent when using an alternative (parameterized by \( \hat{\theta} \)) instead of the correct distribution \( \theta \) can be calculated on a percentage basis:

\[ L \left( \hat{\theta}, \theta \right) = \frac{CE_\theta - CE_{\hat{\theta}}}{CE_\theta}. \]

Jorion (1986) focuses on the same type of loss function and analyzes it also in expected utility units:

\[ L \left( \hat{\theta}, \theta \right) = \frac{E_{\mathbb{P}|\theta} \left[ U \left( \langle S, \pi_\theta^* \rangle \right) \right] - E_{\mathbb{P}|\theta} \left[ U \left( \langle S, \pi_{\hat{\theta}}^* \rangle \right) \right]}{E_{\mathbb{P}|\theta} \left[ U \left( \langle S, \pi_\theta^* \rangle \right) \right]} \]

(3)

An alternative loss function was used by Horst et al. (2001) who prefer a mean-variance calculation:

\[ L \left( \hat{\theta}, \theta \right) = \langle \mu_\theta, \pi_\theta^* - \pi_{\hat{\theta}}^* \rangle - \frac{1}{2} \gamma \cdot \left( \pi_{\hat{\theta}}^* - \pi_\theta^* \right)' \left( \pi_{\hat{\theta}}^* - \pi_\theta^* \right). \]

\(^2\)Here \( I \) denotes the \((n \times n)\) identity matrix.

\(^3\)See for example Zellner (1987).
With the loss function specified, the agent tries to choose the alternative distribution \( P_{X|\hat{\theta}} \) such that the expected loss is minimized (see also Zellner (1987))

\[
\min_{\hat{\theta}} \int L(\hat{\theta}, \theta) \, dQ(\theta) = E_{Q} \left[ L(\hat{\theta}, \theta) \right] =: E_{\theta} \left[ L(\hat{\theta}, \theta) \right],
\]

where \( Q \) is the probability measure defined on the measurable space \((\Omega_{\theta}, \mathcal{F}_{\theta})\) for the parameter \( \theta \in \Omega_{\theta} \).

Although this calculates the impacts of misspecification of return and risk (here variance and covariance) in units of cash equivalent, no procedure is given on how to determine the optimal choice, i.e., which the optimal risk-return specification to use. Obviously, to do this, one must specify an optimization criterion for the choice of uncertainty implied by the estimated values.

When the portfolio strategy is parameterized by the risk aversion one might deduce the optimal risk aversion parameter from the optimal choice. In the framework of Horst et al. (2001), using to denote the level of risk aversion chosen, one starts with

\[
\pi^{\text{Alt}}(\gamma^{\text{Alt}}) = \langle \gamma^{\text{Alt}}, \rho \rangle^{-1} \cdot (\pi^{\text{Alt}} - \langle \rho, 1 \rangle).
\]

The authors then find that with a decreasing number of (historical) observations of return realizations \( T \) and an increasing number of assets \( K \) the optimal alternative risk aversion \( \gamma^{\text{Alt}} \) increases, being always higher than the original risk aversion:

\[
\gamma^{\text{Alt}^{*}} = \gamma \cdot \left( 1 + c \cdot \frac{K - 1}{T} \right).
\]

This result makes intuitively sense as the asset manager is assumed to be averse to uncertainty. Thus he should never end up with a lower risk aversion and the increase in risk aversion should be influenced just as described. It should also be intuitive from the point of view of portfolio risk: estimation risk increases portfolio risk (see for example Frost and Savarino (1986), but with decreasing impact as the time series used for estimation becomes longer, see Kempf et al. (2001)).

The same result was discovered independently by Lauprete (2001), but without using on the optimal value for the risk-aversion parameter. The optimal (asymptotic) loss is given by

\[
L(\pi^{\text{Alt}^{*}}) = \langle \pi^{\text{Orig}}, \rho \rangle \cdot \frac{K - 1}{T}.
\]

Further results have been obtained for the case of a mean absolute deviation as a risk measure and also for a multivariate Student-t distribution for the asset returns by Lauprete (2001) and can also be found in Lauprete et al. (2003). This alternative parameter of risk aversion can be interpreted as what should have been the investor's original risk aversion if he had already considered the uncertainty of the return distributions.
The principal idea in the Bayesian type approach is that all the available information should be incorporated into the estimation of expected return. When having multiple assets, the values of all assets may be used to estimate the expected return of every single asset. This leads to an estimated value different from that when only the values of the asset under analysis are used. The most popular estimator is the (James-) Stein-estimator which is also called a shrinkage estimator, as the pure expected return $\hat{\mu}$ of the asset is reduced by the factor $\alpha$:

$$\hat{\mu}_{\text{shrinkage}} = \alpha \cdot \hat{\mu} + (1 - \alpha) \cdot \mu_0,$$

where $\mu_0$ is an alternative estimation. Stein (1962) proposed an estimator that increases the performance of the estimation with respect to a quadratic loss function whereas, for example, Berger (1978) derives an optimal estimator for a polynomial loss function. One of the first applications of this method to portfolio optimization can be found in Jorion (1986). Observe that the use of the loss function can be seen as the core of the solution approach: We have to measure somehow whether (and maybe also how much) one distribution is better than an alternative one.

Another form of the Bayesian approach is to include additional information into the system, especially knowledge of specialists. Updating a prior (e. g. market) portfolio with the agents' expectations leads to a posterior vector of expectations, taking into account the uncertainty of the agent about his estimates, see Black and Litterman (1990) and Black and Litterman (1992). However, this approach is not free of criticism: Why is a possible uncertainty of the prior not taken into account? When taking the market as a prior, what is the market portfolio? Nevertheless, besides these specific points on the proposed approach, it should be remarked that the idea of mixing different possible distributions can be seen as a generalized method of using shrinkage estimators.

**Remark 1.** For the James-Stein framework, to our knowledge, no equivalent alternative parameter of risk aversion has been derived. But, alternatively, the optimal choice of the shrinkage factor was determined. Both approaches for an equivalent expression of the chosen alternative distribution assume a certain structure: a specific utility function in the first case and a prior as a mean estimation in the second.

Another criterion for measuring the deviation from the neutral return functional is proposed by Anderson et al. (2000) and further investigated by Pascal (2001), both in a dynamic setting. As there is no problem to cast the approach to the less general static setting, we will directly reformulate it here. The author proposes to measure the distance between the original, assumed, mean return and an alternative mean of the return by using relative entropy. Therefore he interprets the expectation operator as a functional that might be adjusted for the desire of robustness, i. e. the aversion to uncertainty inherent to the returns estimation. Writing the expected value of the final portfolio value $X$ as $T(X) = E[X]$ he considers alternative operators given by
\[ T^f(X) = \frac{E[f \cdot X]}{E[f]}, \]

where \( f \) is a strictly positive function. In the case of a constant function, the scaling by \( E[f] \) gives the original operator. The relative entropy for an alternative estimation can be written in terms of the original expectation operator as

\[ I(f) = \frac{T(f \cdot \log(f))}{T(f)} - \log(T(f)) \]
\[ = \frac{E[f \cdot \log(f)]}{E[f]} - \log(E[f]). \]

The optimal alternative expectation operator is then chosen by

\[
\min_f \frac{1}{\theta} I(f) + T^f(X) = \min_f \frac{1}{\theta} \left( \frac{E[f \cdot \log(f)]}{E[f]} - \log(E[f]) \right) - \frac{E[f \cdot X]}{E[f]}
\]

For calculation of the impact of estimation errors on optimal dynamic asset allocation see also Xia (2000). Balduzzi and Liu (2001) analyze the impact of misspecification given by non-cross inference estimation (see above, estimation methods) in a dynamic setting by evaluating the utility loss in terms of certainty equivalent losses.

Another common criterion of choice is to protect oneself against the worst outcome. In (dynamic) hedging of contingent claims, this is called the super-replication approach. In many, mostly engineering, applications the term robust optimization is used. For an application of this method to the financial context, see for example Goldfarb and Iyengar (2002), Iyengar (2002), and Costa and Paiva (2002). The principal idea of this approach is to have a portfolio that is optimal in the worst-case scenario, what is a max-min choice.

**Remark 2.** The word "robust" is also used in the environment of estimation, where a loss function is applied on the deviation of the alternative estimations from the initial estimation (for example population mean).

It is important to note that all of the present approaches only focus on the estimates of the risk and return variables of each asset return and do not analyze the whole efficient frontier as a convex combination of some or all asset returns. In practice this probably will not make much difference as the number of assets is finite, which allows us to construct an isomorphism between the probability spaces of individual asset returns and the equivalent efficient frontiers. This means that
after adjusting the risk and return variables for uncertainty, one could construct the efficient frontier.

Another typical characteristic of the approaches is their two-step procedure to find the optimal portfolio:

1. Adjust the original distribution $P_\theta$ to a new one $P_{\theta^*}$.

2. Apply the traditional optimization to the new distribution.

Contrary to this two-step methodology, Michaud (1998) proposed a way to derive an efficient frontier, which he calls the “resampled” frontier. It is built from the “resampled” efficient portfolios by averaging: For every sample $\theta \in \Omega_\theta$ the optimal portfolio $\pi^*_\theta$ with the same return rank as the populational (original) portfolio $\pi^*_{\theta^*}$ is chosen and its weight kept. Then, over all samples, these weights are averaged, leading to an “average” portfolio $\tilde{\pi}^*$. The average portfolio weights are used together with the original mean and variance values to construct the efficient frontier as the averaging is done stepwise for each fixed level of risk. Using our above introduced notation, we can write this approach as

$$\tilde{\pi}^* := \int_{\theta \in \Omega_\theta} \arg\max_{\pi_\theta \in A(X_\theta)} E_{P_{\theta}} \left[ U \left( \langle X, \pi_\theta \rangle \right) \right] dQ(\theta).$$

This approach is different from the one mentioned above, as the optimal portfolio is chosen first and then an averaging among the alternative distributions is carried out. An important fact to note is that during averaging the author does not explicitly take into consideration the agent’s aversion to uncertainty, i.e., he does not apply a utility function to the $\theta$-optimal portfolios. Bearing this in mind, we have the

**Proposition 2.1 (Rank equivalent optimal portfolio)** For a given parameter of risk the optimal stochastically equivalent portfolio is given by

$$\tilde{\pi}^* = \int V \left( \arg\max_{\pi_\theta \in A(X_\theta)} E_{P_{\theta}} \left[ U \left( \langle X, \pi_\theta \rangle \right) \right] \right) dQ(\theta),$$

where the agent’s aversion to uncertainty is expressed via the function

$$V : A(X_\theta) \to A(X_\theta),$$

$$V : (\pi^*_\theta(1), \ldots, \pi^*_\theta(n)) \mapsto (\tilde{\pi}^*_\theta(1), \ldots, \tilde{\pi}^*_\theta(n)),$$

such that for all $i = 1, \ldots, n$:

---

4Here, we define the set of admissible strategies $A(X_\theta)$ with respect to the portfolio wealth $X = X(\theta)$ in analogy to the standard set $A(X)$ without conditioning on the parameter $\theta$. 

99
An important point to consider in the way of constructing the "mean" portfolio is that the distribution of the weights is in many cases highly skewed, see Figure 1. The mean is thus not a statistic that – itself – describes well the distribution.

Figure 1
Frequency plot of the weights of the stock index (IBOVESPA) in the optimal portfolio

In order to account better for the asymmetries observable in the distribution of the portfolio weights, one might think of doing the averaging using the maximum likelihood:

$$\tilde{\pi}_{\theta}^{(i)} \left\{ \begin{array}{ll} \geq \pi_{\theta}^{*,(i)}, & \text{if } \pi_{\theta}^{*,(i)} \leq E_{\theta} \left[ \pi_{\theta}^{*,(i)} \right] \\ \leq \pi_{\theta}^{*,(i)}, & \text{if } \pi_{\theta}^{*,(i)} \geq E_{\theta} \left[ \pi_{\theta}^{*,(i)} \right] \end{array} \right. ,$$

$$i = 1, \ldots, n. \text{ The problem with this proposal is that the weights might not sum to one, i.e. } \tilde{\pi}_{\theta} = \sum_{i=1}^{n} \tilde{\pi}_{\theta}^{(i)} \neq 1, \text{ thus } \tilde{\pi}_{\theta} \notin A(X_{\theta}).$$

3. Incorporating Distributional Uncertainty with Penalty Approaches

In this section we discuss several proposals on how to account for uncertainty in the distribution parameters which use the idea of penalizing the expected utility.
value by a term that – in a flexible way – may represent the users’ aversion to uncertainty in the return distributions.

Our first approach works in a traditional way: the deviation of the parameter from a fixed (but in our case not previously determined) parameter is penalized with a quadratic function. The reason for such a choice of a penalty function can be seen in the use of variance (in the expected utility) as a measure for aversion to uncertainty in the distribution parameters. Having used the Markowitz framework for the optimization of the portfolio composition it might be understandable to assume the same kind of aversion function for the parameters as well.

Our second approach is divided into two subparts where we mainly just change the units of measurement. The main idea of our proposals is to find a composition that can be seen as stable. We will show how this characteristic is more like a by-product as the main focus is to minimize the distance to the result in the case of choosing the best portfolio. But, before entering our specific approaches we take a look at the problem in a general setting.

3.1. The general formulation of the penalty approaches

The abstract formulation for incorporating the uncertainty about the distributions into our construction of the efficient frontier is to be on a semi-product space. Let therefore \((\Omega, F, \mathbb{P})\) be the probability space for the asset return vector \(S = (S_1, ..., S_n)\). We assume, for simplicity, that the (joint) distribution \(P := \mathbb{P}^{-1}(S)\) can be parametrized by \(\theta = (1, ..., n)\). Let then \((\Omega_\theta, \mathcal{F}_\theta, \mathbb{P}_\theta)\) be the probability space for the parameter \(\theta \in \Omega_\theta\). As the two distributions are not independent, but \(P\) depends on \(Q\), we have a mixed space, writing it at \(Q \propto P\) indicating that \(Q\) influences \(P\). The investor's decision problem is then

\[
\max_{\pi \in A(X)} E_{Q \propto P} [V ((S, \pi))] ,
\]

where \(X\) is a generalized utility function that also accounts for the uncertainty in the asset parameter. This formulation is quite general but, unfortunately, also very far from practical. However, we can use it to reformulate the existing approaches from this point of view.

We may already point out that we apply the optimization (for \(\pi\)) after averaging over \(\theta\) whereas the typical approaches via loss functions first optimize and then average (see Equations (3), (4), and (5)). As mentioned before, the procedure in this case is to estimate an alternative parameter \(\tilde{\theta}\) by using the expected value of a loss function \(L\) :

\[
\theta^* = \arg\min_{\tilde{\theta}} E_Q \left[ L \left( \tilde{\theta}, \theta \right) \right] ,
\]

and then to apply the optimization to the chosen distribution

\[
\max_{\pi \in A(X_{\theta^*})} E_{P|\theta^*} [U ((S, \pi))] .
\]
As the loss function $L$ represents the agent’s aversion to uncertainty in the distribution parameters and $U$ the risk aversion, together they do a similar job as $V$ does for the combined optimization in Equation (6).

As the two-step procedure of first choosing an alternative distribution and then optimizing the portfolio does not necessarily lead to an optimal solution for (6) we want to go a slightly different way. Applying results from convex duality, the original problem (6) might in some cases of the generalized utility function $V$ be rewritten in a form involving a penalty function $B_\delta$ defined as in (2):

$$\max_{\pi \in \mathcal{A}(X)} \int_{\mathcal{Q}} E_{\mathcal{P}|\theta} \left[ U \left( \langle S, \pi\theta \rangle \right) \right] - B_\delta \left( \pi, \theta \right) \, d\mathcal{Q} \left( \theta \right). \quad (6)$$

In what follows, we give two types of specifications and discuss their consistency with the intuitively expected results in some special cases. We will consider a simple example called the substitution problem. It is defined as follows. Let $X_1$ and $X_2$ be two assets with same return variances $\sigma_{11} = \sigma_{22}$ and complete correlation. Their expected returns $\mu_1, \mu_2$ are independent normal random variables with means $m_1, m_2$, and variances $s_1^2, s_2^2$. In the traditional problem, where $\mu_1$ and $\mu_2$ are deterministic, the portfolio decision would be $\pi_1 = 1_{\{\mu_1 \geq \mu_2\}} = 1 - \pi_2$, leading to possible extreme shifts in portfolio weights due to the discontinuous nature of the indicator function. We shall see that our proposals avoid this problem.

3.2. Quadratic uncertainty aversion

One of the most typical ways of calculating the deviation from the mean value is to calculate the variance of a random variable, thus penalizing any deviation with a quadratic function. The variance dislikes in a symmetric way the deviation from the mean:

$$B_\delta \left( \pi, \theta \right) := \delta \cdot \left\{ E_{\mathcal{P}|\theta} \left[ U \left( \langle S, \pi \rangle \right) \right]^2 - \left( \int_{\mathcal{E}} E_{\mathcal{P}|\theta} \left[ U \left( \langle S, \pi \rangle \right) \right] \, d\mathcal{Q} \left( \theta \right) \right)^2 \right\}. \quad (7)$$

Example (Substitution Problem). In the simple two-asset case with $\pi = (\pi_1, \pi_2) = (\pi_1, 1 - \pi_1)$, and $\mu_i \sim \mathcal{N} \left( m_i, s_i^2 \right)$, $i = 1, 2$ independent, with $U = Id$ for the optimal portfolio, we get

$$\pi_\delta^* = \arg \max_{\pi} \left( \pi_1 m_1 + \pi_2 m_2 - \delta \left( \pi_1^2 s_1^2 + \pi_2^2 s_2^2 \right) \right)$$

and

$$\pi_1^* = \left( s_1^2 + s_2^2 \right)^{-1} \cdot \left( \frac{m_1 - m_2}{2\delta} + s_2^2 \right).$$

The expected return for a given $\theta$ is the Gaussian random variable

102
\[
E_{\mathcal{P}^{\theta}} [U ((S, \pi))] = \pi_1^\ast \cdot \mu_1 + \pi_2^\ast \cdot \mu_2
\sim \mathcal{N} \left( \pi_1^\ast m_1 + \pi_2^\ast m_2, (\pi_1^\ast)^2 \cdot s_1^2 + (\pi_2^\ast)^2 \cdot s_2^2 \right).
\]

In the special case of \(m_1 = m_2 = m\),

\[
\pi_1^\ast = \frac{s_2^2}{s_1^2 + s_2^2},
\]

hence

\[
E_{\mathcal{P}^{\theta}} [U ((S, \pi))] \sim \mathcal{N} \left( \frac{s_1^2}{s_1^2 + s_2^2} \cdot \frac{s_2^2}{s_1^2 + s_2^2} \cdot m \right).
\]

This means that diversification decreases the risk measured by variance.

### 3.3. Determining an optimal portfolio with a stability feature

Recall that Michaud (1998) calculated the mean portfolio composition among all the resampled distributions. Although, as mentioned before, the expected value may often not be a very informative statistic, it is important to note the intuition behind this approach: The mean portfolio is seen as the portfolio that has somehow minimal distance from all possible portfolios. This means that on average no other portfolio will be that close to all possible realizations than the chosen one. Our proposals focus on this idea of being close to all possible outcomes. This allows the investor to be comfortable with any future realization.

**Minimizing the Deviation from the Best Choice in Expected Utility Units**

In order to analyze the suboptimal decision we take into account that the often determined strategy can not be optimal for all scenarios. We define

\[
B_\delta (\pi, \theta) := \delta \cdot \left\| E_{\mathcal{P}^{\theta}} [U ((S, \pi))] - \max_{\pi_\theta \in A(X_\theta)} E_{\mathcal{P}^{\theta}} [U ((S, \pi))] \right\|_2^2 \quad (7)
\]

This approach also follows somehow the spirit of Jorion (1986) by measuring the deviation from the best result. We are forced to be closer to the portfolios that are optimal in the scenarios with high expected utility. There exist now two possibilities: We can insert this penalty function into the optimization problem (6). Alternatively, we calculate the optimal portfolio composition directly from (7).

\[
\pi^\ast = \arg \min_{\pi} \int_{\theta \in \Omega_\theta} \left\{ E_{\mathcal{P}^{\theta}} [U ((S, \pi))] - \max_{\pi_\theta \in A(X_\theta)} E_{\mathcal{P}^{\theta}} [U ((S, \pi))] \right\}^2 dQ (\theta).
\quad (8)
\]
Example (Substitution Problem, cont’d). In our simple example, the optimal portfolio is given by

$$
\pi^* = \arg \min_\pi \int dQ(\theta) \left( E[U_\gamma(\langle \pi_\theta, S \rangle)] - EU_\gamma(\langle \pi_\theta^*, S \rangle) \right)^2,
$$

(9)

where $\pi_\theta^*$ is the optimal portfolio for a given $\theta$. In our example, we have $\pi_\theta^* = (1, 0)$, if $\mu_1 > \mu_2$, respectively $\pi_\theta^* = (0, 1)$, if $\mu_2 > \mu_1$. The corresponding expected utility is given by

$$
E[U_\gamma(\langle \pi_\theta, S \rangle)] = \max(\mu_1, \mu_2) - \sigma_{11}.
$$

(10)

Denote the right-hand side of (9) by $I(\pi)$. A straightforward calculation then yields

$$
\frac{dI(\pi)}{d\pi_1} = 2 \int dQ(\mu_1, \mu_2) \left\{ \pi' \cdot \mu - \max(\mu_1, \mu_2) \right\} (\mu_1 - \mu_2).
$$

(11)

Setting the above equal to zero as the first order condition for the optimum leads to the solution

$$
\pi_1^* = \frac{E[(\mu_1 - \mu_2) \langle \mu_1 - \mu_2 \rangle]}{s_1^2 + s_2^2} = \frac{E[(\mu_1 - \mu_2)^2 1_{\{\mu_1 > \mu_2\}}]}{s_1^2 + s_2^2}.
$$

(12)

Let us observe that in the case $\pi_1 = \pi_2$, the optimal portfolio allocates exactly half of the resources to each of the assets, independently of the respective variances $s_1^2$ and $s_2^2$ of the expected returns. The reason is that only the mass above/below the mean expected return is considered and not its dispersion.

Remark that in the case of a quadratic utility function with $\gamma \in \mathbb{R}^+$ the coefficient of risk aversion, we have

$$
E_{\theta|\eta}[U(\langle S, \pi \rangle)] = \langle \mu_\theta, \pi \rangle - \frac{\gamma}{2} (\pi' \Sigma_\theta, \pi),
$$

the formula looks similar to that of Horst et al. (2001):\(^5\)

\[
\pi^* = \arg \min_\pi \int_{\theta \in \Omega_\eta} \left\{ \langle \mu_\theta, \pi \rangle - \frac{\gamma}{2} (\pi' \Sigma_\theta, \pi) \\
- \max_{\pi_\theta} \left( \langle \mu_\theta, \pi_\theta \rangle - \frac{\gamma}{2} (\pi_\theta' \Sigma_\theta, \pi_\theta) \right) \right\}^2 dQ(\theta)
\]

\[
= \arg \min_\pi \int_{\theta \in \Omega_\eta} \left\{ \langle \mu_\theta, \pi - \pi_0^* \rangle - \frac{\gamma}{2} ((\pi' - \pi_0^*)' \Sigma_\theta, (\pi' - \pi_0^*)) \right\}^2 dQ(\theta).
\]

\(^5\)We denote again with $\pi_0^*$ the portfolio that is optimal for the expected utility maximization under $P_{X|\theta}$. 
though they want to

$$
\arg \min_{\theta} \int_{\theta \in \Omega_{\theta}} L \left( \hat{\theta}, \theta \right) dQ(\theta)
= \arg \min_{\theta} \int_{\theta \in \Omega_{\theta}} \left\{ \left( \mu_{\theta} - \pi_{\theta}^{*} - \pi_{\theta}^{*} \right) - \frac{\gamma}{2} \cdot \left( \pi_{\theta}^{*} - \pi_{\theta}^{*} \right) \cdot \left( \pi_{\theta}^{*} - \pi_{\theta}^{*} \right) \right\} dQ(\theta).
$$

Minimizing the Deviation from the Best Choice in Portfolio Weights

Another way of analyzing the deviation of a given portfolio from the optimal portfolio for some $\theta$ is to measure the distance directly in the portfolio weights instead of in expected utility units. In this case, our penalty function would be

$$
B_\delta(\pi, \theta) := \delta \cdot ||\pi_{\theta}^{*} - \pi||^2_2 .
$$

Again, we calculate the optimal portfolio directly from optimizing this function averaged over the parameters $\theta$ instead of involving the whole optimization (6):

$$
\pi^{*} = \arg \min_{\pi} \int_{\theta \in \Omega_{\theta}} ||\pi_{\theta}^{*} - \pi||^2_2 dQ(\theta) .
$$

Example (Substitution problem, cont’d). Let us determine the optimal portfolio in the case of the substitution problem. Assume $\theta = (\mu_{1}, \mu_{2})$ with any distribution $Q$, then the problem of finding a minimal sensitive portfolio can be split into two cases:

$$
\pi^{*} = \arg \min_{\pi} \int_{\theta \in \Omega_{\theta}} ||\pi_{\theta}^{*} - \pi||^2_2 dQ(\theta)
= \arg \min_{\pi} \left\{ \int_{\theta \in \Omega_{\theta}} ||\pi_{\theta}^{*} - \pi||^2_2 \cdot \mathbb{1}_{\{\mu_{1} \geq \mu_{2}\}} dQ(\theta)ight\}
+ \int_{\theta \in \Omega_{\theta}} ||\pi_{\theta}^{*} - \pi||^2_2 \cdot \mathbb{1}_{\{\mu_{1} < \mu_{2}\}} dQ(\theta)
\right\}
= \arg \min_{\pi} \left\{ ||\pi_{\theta}^{*} - \pi||^2_2 \cdot \int_{\theta \in \Omega_{\theta}} \mathbb{1}_{\{\mu_{1} \geq \mu_{2}\}} dQ(\theta)\right\}
$$
\[ + \|\pi_0^\ast - \pi\|^2_2 \int_{\theta \in \Omega_0} 1_{\{\mu_1 < \mu_2\}} d\mathbb{Q}(\theta) \]

\[ = \arg\min_{\pi} \left\{ \|\pi_0^\ast - \pi\|^2_2 \mathbb{Q}\{\{\mu_1 \geq \mu_2\} \right\} \\
+ \|\pi_0^\ast - \pi\|^2_2 \mathbb{Q}\{\{\mu_1 < \mu_2\} \right\} \right\} \\
= \arg\min_{\pi} \left\{ \left\| \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} \right\|^2_2 \mathbb{Q}\{\{\mu_1 \geq \mu_2\} \right\} \\
+ \left\| \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} \right\|^2_2 \mathbb{Q}\{\{\mu_1 < \mu_2\} \right\} \right\} \right\} \\
= \arg\min_{\pi} \left\{ \left(1 - \pi_1\right)^2 + \pi_2^2 \mathbb{Q}\{\{\mu_1 \geq \mu_2\} \right\} \\
+ \left(1 - \pi_2\right)^2 + \pi_1^2 \mathbb{Q}\{\{\mu_1 < \mu_2\} \right\} \right\} \right\} .

Using the restriction that \(\pi_2 = 1 - \pi_1\), and letting \(\alpha := \mathbb{Q}\{\{\mu_1 \geq \mu_2\}\}\) we get

\[ \pi^\ast = \arg\min_{\pi} \left\{ 2 \cdot \left(\pi_1^2 - 2 \cdot \alpha \pi_1 \right) + 2 \cdot \alpha \right\} . \]

Calculating the gradient, we find that

\[ \pi^\ast = \begin{pmatrix} \alpha \\ 1 - \alpha \end{pmatrix} . \]

In the deterministic case with \(\mu_1 > \mu_2\) we thus recover the typical result of \(\pi^\ast = \begin{pmatrix} 1 \\ 0 \end{pmatrix} . \)

The Feature of Minimal Sensitivity

Minimal sensitivity accounts for the problem of changing strategies and we try to find a more balanced portfolio. In the approach of equally weighted assets large weights in a few assets are penalized as the \(L_2\)-norm increases in these cases. The more equally weighted, the smaller the penalty. This leads to a portfolio that somehow lies in the middle of all possible portfolios. Thus, we may also see the resulting portfolio as one that is insensitive to changes in the return distributions. We may then formulate the objective function with penalty as

\[ \max_{\pi \in \mathcal{A}} \int_{\theta \in \Omega_0} E_{P|\theta} [U(S, \pi)] - \|\pi\|_2^2 d\mathbb{Q}(\theta) . \] (15)

As mentioned within the general formulation of our penalty approach, the optimization is applied after averaging which makes this formulation different from the problem typically analyzed.
Furthermore, we note that the use of the norm of the portfolio weights as the penalty is an elegant way of incorporating the aversion to high weights in the portfolio, although it is not the typical way of measuring sensitivity.

Let us briefly comment on why our proposal incorporates the feature of high stability. A natural first approach to penalizing high variation in the expected utility due to slight changes in the parameters might be based on analyzing the gradient, choosing for example the following function

$$B_\delta (\pi, \theta) := \delta \cdot \left\| \nabla_{\theta} E_{\mathbb{P}|\theta} [U (\langle S, \pi_{\theta} \rangle)] \right\|_2^2.$$ 

In the special case of normally distributed means, $\theta \equiv \mu, X \sim \mathcal{N} (\mu, \Sigma)$, we can write this as

$$B_\delta (\pi, \theta) = 2\delta (\Sigma^{-1})^2 \cdot \mathbb{E}_\theta \left[ E_{\mathbb{P}|\theta} [U (\langle S, \pi_{\theta} \rangle)]^2 \right],$$

leading to an objective function similar to that in the variance penalty

$$f = \max_{\pi \in \mathcal{A}(X)} \int_{\Omega_\theta} E_{\mathbb{P}|\theta} [U (\langle S, \pi_{\theta} \rangle)] d\mathbb{Q} (\theta)$$

$$-2\delta (\Sigma^{-1})^2 \int_{\Omega_\theta} E_{\mathbb{P}|\theta} [U (\langle S, \pi_{\theta} \rangle)]^2 d\mathbb{Q} (\theta).$$

When penalizing for high variation in the portfolio composition, we would have

$$B_\delta (\pi, \theta) := \delta \cdot \left\| \nabla_{\theta} \pi_{\theta} \right\|_2^2.$$

The problem with measuring the sensitivity in this traditional form is that we are confronted with compositions that change abruptly at the points where one parameter starts to dominate the other, as the maximum is not a continuous function. Let us visualize this in the substitution problem again: With $\sigma_1 = \sigma_2$ the optimal portfolio is given by

$$\pi^* = \arg \min_{\pi} \int_{\Omega_\theta} \left\| \nabla_{\theta} \pi_{\theta} \right\|_2^2 d\mathbb{Q} (\theta).$$

Since for fixed $\mu = (\mu_1, \mu_2)$ we have $\pi_1 = 1_{\{\mu_1 \geq \mu_2\}} = 1 - \pi_2$, the gradient satisfies

$$\nabla_{\theta} \pi_{\theta} = 0, \mathbb{Q}-a.e.$$ 

which means that the gradient is, even in this simple example, not a very good measure of sensitivity and should be substituted by a measure of distance (after normalizing) $\| \nabla_{\theta} \pi_{\theta} \| \sim \| \pi^*_\theta - \pi \|$, where $\pi^*_\theta$ is a fixed portfolio, in our case the optimal one for a given parameter $\theta$. Observe that we did explicitly not specify the
norm, as the choice of it may be left to the investor. For example we can allow the supremum-norm. This means that the use of

$$\pi^* = \arg \min_\pi \int_{\Omega^\theta} \|\pi^*_\theta - \pi\|_\infty d\mathbb{Q}(\theta)$$

leads to a portfolio $\pi^*$ that has minimal maximum distance between any two portfolio weights and hence leads to equivalent results as the approach of an equally weighted composition (15). The $L_2$-norm was applied in our second proposal in the previous section. When the sensitivity is not analyzed in terms of portfolio composition but in terms of expected utility, our proposal minimizes the deviation from the best choice in expected utility units.

These observations about the form of how to describe the sensitivity of the portfolio composition classify our proposals as approaches that have a certain stability feature as they minimize the sensitivity.

3.4. Advantages and disadvantages of the approaches

The most important advantage over the conventional approaches with loss functions is that we do not have to specify a prior to measure the impact of the uncertainty in the distribution. This is an important point for the determination of an optimal portfolio as it is often hard for an asset manager to set a prior as well as his individual distribution too.

Secondly, the penalty approach, as seen with the proposals for incorporating the sensitivity of the composition, allows for the specification of a wide range of objectives of the asset manager in a simple form. The problem will nevertheless be on how to specify the factor of aversion to uncertainty about the return distributions. Approaches where the expected utility decreases fast with little changes in the aversion factor may be dangerous to use when specification errors may arise.

Finally, from an operational point of view, our penalty approaches allow us to avoid the typical two-step procedure of first finding an optimal parameter $\theta^*$ and then determining the optimal portfolio $\pi^*_{\theta^*}$ with respect to this parameter.

A problem may arise in practice when one is obliged to evaluate the double integral with Monte-Carlo methods, since calculations may become time-consuming.

4. Comparison of the Approaches in a Real Market Setting

We show the effects of some of the approaches in an example in the Brazilian financial market. The assets with their populational data are given as in Table 1 where variance is used as the measure of risk; normal marginal distributions are assumed.\footnote{Data as of 05.11.2002. Mean and variance are estimated traditionally from historical monthly data. SELIC=Selic interest rate; RF curto=Pre-fixed interest rate, 3 months; RF longo=Pre-fixed interest rate, 6 months; Equity=IBOVESPA stock index; T. Camb.=dollar-denominated interest rate title. 6}
Table 1
Assets included in the portfolio optimization

<table>
<thead>
<tr>
<th>Asset</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELIC</td>
<td>17.6%</td>
<td>0.3%</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF corto</td>
<td>19.4%</td>
<td>1.2%</td>
<td>0.15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF longo</td>
<td>19.1%</td>
<td>2.9%</td>
<td>0.17</td>
<td>0.96</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equity</td>
<td>24.2%</td>
<td>35.3%</td>
<td>-0.03</td>
<td>0.60</td>
<td>0.67</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T. Conv.</td>
<td>41.3%</td>
<td>16.0%</td>
<td>-0.15</td>
<td>-0.49</td>
<td>-0.60</td>
<td>-0.68</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGP-M</td>
<td>23.5%</td>
<td>3.9%</td>
<td>0.02</td>
<td>-0.08</td>
<td>-0.12</td>
<td>-0.07</td>
<td>0.37</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>LFT</td>
<td>16.5%</td>
<td>0.6%</td>
<td>0.29</td>
<td>0.31</td>
<td>0.43</td>
<td>0.21</td>
<td>-0.23</td>
<td>-0.13</td>
<td>1</td>
</tr>
</tbody>
</table>

When we drop the assumption of fixed, deterministic means, and assume a normal distribution for them with the populational means as expected values and the returns covariance matrix as the covariance matrix of the means, we work under the assumptions of Horst et al. (2001). The set of resampled frontiers for this setting, as proposed by Michaud (1998), is shown in Figure 2.

![Figure 2](image-url)

Set of resampled frontiers

For our purpose of simulation we follow mainly the typical approach starting with a vector of means and covariances that are assumed to be multinomially distributed see Jobson and Korkie (1980). In addition to this setting we also sample the correlation structure which lets the covariance matrix follow a Wishart distribution, defined as the sum of independent products of multivariate normal random vectors see Johnsons and Wichern (1992).

months; IGP-M=inflation-linked interest title, 12 months; LFT=Letra Financeira Tesouro Nacional, 12 months.
4.1. Pure Markowitz

When we directly apply the mean-variance optimization to the original data set we get the results in Table 2.

<table>
<thead>
<tr>
<th>Asset</th>
<th>Original Data</th>
<th>Alternative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Weights</td>
</tr>
<tr>
<td>SELIC</td>
<td>17.6%</td>
<td>40%</td>
</tr>
<tr>
<td>RF curto</td>
<td>19.4%</td>
<td>182%</td>
</tr>
<tr>
<td>RF longo</td>
<td>19.1%</td>
<td>39%</td>
</tr>
<tr>
<td>Equity</td>
<td>24.2%</td>
<td>1%</td>
</tr>
<tr>
<td>T. Camb.</td>
<td>41.3%</td>
<td>3%</td>
</tr>
<tr>
<td>IGP-M</td>
<td>23.5%</td>
<td>3%</td>
</tr>
<tr>
<td>LFT</td>
<td>16.5%</td>
<td>71%</td>
</tr>
</tbody>
</table>

The problem of this result can be seen when taking a look at the result after a slight change in the expected mean for the SELIC and for the LFT. Their returns are so close that none of them could arguably exclude a misspecification of 0.6%. Unfortunately, the portfolio composition in this case suffers extreme changes, which hardly can be realized by bigger funds in the Brazilian market. We use this example as our “benchmark” for our other approaches.

4.2. Limiting the portfolio weights

When limiting the portfolio weights by prohibiting leverage and short-selling, we reduce the extreme changes in the optimal weights.

<table>
<thead>
<tr>
<th>Asset</th>
<th>Original Data</th>
<th>Alternative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Weights</td>
</tr>
<tr>
<td>SELIC</td>
<td>17.6%</td>
<td>33%</td>
</tr>
<tr>
<td>RF curto</td>
<td>19.4%</td>
<td>57%</td>
</tr>
<tr>
<td>RF longo</td>
<td>19.1%</td>
<td>0%</td>
</tr>
<tr>
<td>Equity</td>
<td>24.2%</td>
<td>1%</td>
</tr>
<tr>
<td>T. Camb.</td>
<td>41.3%</td>
<td>5%</td>
</tr>
<tr>
<td>IGP-M</td>
<td>23.5%</td>
<td>4%</td>
</tr>
<tr>
<td>LFT</td>
<td>16.5%</td>
<td>0%</td>
</tr>
</tbody>
</table>

4.3. Bayesian distortion of the distributions’ parameters: the Black-Litterman proposal

When we apply the Bayesian idea of updating our prior distribution, parameterized by the default setting, we need to specify an alternative distribution. As we assume the setting that the expectations resulted from the views of the firm's specialists and not from estimations out of historical data, we can not apply the James-Stein estimator. Furthermore, no explicit prior distribution is given, therefore we use the original data as prior and interpret the previously used alternative
distribution as our updating distribution. This leads to the distribution used as alternative data in the following setting:

Table 4
Portfolio weights according to Black-Litterman

<table>
<thead>
<tr>
<th>Asset</th>
<th>Original Data</th>
<th>Alternative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Weights</td>
</tr>
<tr>
<td>SELIC</td>
<td>17.6%</td>
<td>40%</td>
</tr>
<tr>
<td>RF curto</td>
<td>19.4%</td>
<td>182%</td>
</tr>
<tr>
<td>RF longo</td>
<td>19.1%</td>
<td>-59%</td>
</tr>
<tr>
<td>Equity</td>
<td>24.2%</td>
<td>1%</td>
</tr>
<tr>
<td>T. Camb.</td>
<td>41.3%</td>
<td>3%</td>
</tr>
<tr>
<td>IGP-M</td>
<td>23.5%</td>
<td>3%</td>
</tr>
<tr>
<td>LFT</td>
<td>16.5%</td>
<td>-71%</td>
</tr>
</tbody>
</table>

Observe that a change in the means of SELIC and LFT caused a change in all means of the mixed distribution. We see that there were fewer changes in the weights than in the first example. In the calculation of the update we set the original variances as the variances of the views of the alternative distribution assuming independence between the views. The views represent the assets, i.e., no combination is used.

4.4. Resampling the frontier

For an analysis of Michaud’s method of the resampled frontier, we now take a look at the portfolios with several risk aversion factors. Table 5 gives the compositions along the frontier using the original data, whereas Table 6 shows the results for the alternative data.

Table 5
Resampled efficient frontier with original data

<table>
<thead>
<tr>
<th>Asset</th>
<th>50,000</th>
<th>14,881.8</th>
<th>4,429.3</th>
<th>1,312.3</th>
<th>392.4</th>
<th>172.5</th>
<th>63.7</th>
<th>19</th>
<th>5.6</th>
<th>1.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELIC</td>
<td>0.3%</td>
<td>0.20%</td>
<td>0.83%</td>
<td>0.74%</td>
<td>0.30%</td>
<td>0.25%</td>
<td>0.09%</td>
<td>0.09%</td>
<td>0.09%</td>
<td>0.09%</td>
</tr>
<tr>
<td>RF curto</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>RF longo</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Equity</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>T. Camb.</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>IGP-M</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>LFT</td>
<td>13%</td>
<td>11%</td>
<td>6%</td>
<td>8%</td>
<td>6%</td>
<td>8%</td>
<td>6%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Leg-Norn</td>
<td>0.687</td>
<td>0.684</td>
<td>0.767</td>
<td>0.601</td>
<td>0.444</td>
<td>0.593</td>
<td>0.234</td>
<td>0.541</td>
<td>0.735</td>
<td>0.884</td>
</tr>
</tbody>
</table>

7Remark that this procedure is not how one would apply the Black-Litterman approach in praxis, but is only an example to illustrate the result of mixing distributions.

8The numbers in the first line are the risk aversions along the portfolio frontier.
Table 6
Resampled efficient frontier with alternative data

<table>
<thead>
<tr>
<th>Asset</th>
<th>30.000</th>
<th>14.881.8</th>
<th>4.429.3</th>
<th>1.318.3</th>
<th>392.4</th>
<th>172.5</th>
<th>63.7</th>
<th>19</th>
<th>5.6</th>
<th>1.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELIC</td>
<td>83%</td>
<td>81%</td>
<td>75%</td>
<td>61%</td>
<td>11%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>RF-cuan</td>
<td>4%</td>
<td>5%</td>
<td>9%</td>
<td>25%</td>
<td>70%</td>
<td>77%</td>
<td>72%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>RF-longo</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>21%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Equity</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>8%</td>
<td>22%</td>
<td>16%</td>
<td>0%</td>
</tr>
<tr>
<td>T. Canh</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>6%</td>
<td>11%</td>
<td>29%</td>
<td>71%</td>
<td>84%</td>
<td>0%</td>
</tr>
<tr>
<td>IGP-M</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>LFT</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
<td>10%</td>
<td>8%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

$L_2$-Norm 0.686 0.650 0.619 0.642 0.510 0.600 0.234 0.551 0.723 0.719

We see that the portfolios at the extremes of the frontier are relatively stable. However, in the middle we observe heavy reallocation activities with a maximum change of 19% in SELIC. The last line of the table shows the $L_2$-norm of the composition and their sum along the frontiers are respectively 6.039 and 5.841, an indicator for the concentration of the portfolio. The problem of the large changes in the middle part is due to the relatively high concentration in the assets SELIC and pré-curto RF. Note that we used the restriction of no-leverage.

Although not stated explicitly in the previous approaches, one of the advantages of the resampling is the relatively higher stability of the composition as we somehow chose a "mean portfolio".

4.5. Penalty approaches

We calculated the optimum portfolio for the penalty functions (8) and (14) without taking into account the expected utility as we did in Subsection 3.3. The following tables show the resulting compositions and their $L_2$-norms for both the initial and the alternative data. Tables 7 and 8 give the efficient frontier using expected utility, whereas Tables 9 and 10 show the results for the case with portfolio composition penalty.

Table 7
Efficient frontier using (8) with original data

<table>
<thead>
<tr>
<th>Asset</th>
<th>30.000</th>
<th>14.881.8</th>
<th>4.429.3</th>
<th>1.318.3</th>
<th>392.4</th>
<th>172.5</th>
<th>63.7</th>
<th>19</th>
<th>5.6</th>
<th>1.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELIC</td>
<td>82%</td>
<td>86%</td>
<td>74%</td>
<td>58%</td>
<td>23%</td>
<td>19%</td>
<td>15%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>RF-cuan</td>
<td>11%</td>
<td>6%</td>
<td>0%</td>
<td>7%</td>
<td>25%</td>
<td>19%</td>
<td>15%</td>
<td>6%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>RF-longo</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Equity</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>T. Canh</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>IGP-M</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>LFT</td>
<td>6%</td>
<td>3%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

$L_2$-Norm 0.688 0.742 0.610 0.444 0.248 0.167 0.136 0.283 0.640 0.978

Table 8
Efficient frontier using (8) with alternative data

<table>
<thead>
<tr>
<th>Asset</th>
<th>30.000</th>
<th>14.881.8</th>
<th>4.429.3</th>
<th>1.318.3</th>
<th>392.4</th>
<th>172.5</th>
<th>63.7</th>
<th>19</th>
<th>5.6</th>
<th>1.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELIC</td>
<td>89%</td>
<td>82%</td>
<td>75%</td>
<td>53%</td>
<td>22%</td>
<td>15%</td>
<td>15%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>RF-cuan</td>
<td>5%</td>
<td>7%</td>
<td>6%</td>
<td>7%</td>
<td>25%</td>
<td>19%</td>
<td>15%</td>
<td>7%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>RF-longo</td>
<td>8%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Equity</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>T. Canh</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>IGP-M</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>LFT</td>
<td>6%</td>
<td>12%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

$L_2$-Norm 0.706 0.687 0.613 0.424 0.247 0.167 0.136 0.287 0.643 0.978

112
Comparing these results with those obtained from resampling, we observe that our two approaches lead to more stable portfolios. This can be seen not only in the $L_2$-norm of the differences between the portfolios, but also in the maximum change in the portfolio weights. Observe that the sum of the $L_2$-norm of the differences of the portfolio changes, i.e., how much the weights of the optimal portfolios change, in the case of the resampled frontier the weights are nearly three times larger than in the case of expected utility penalty.

Interestingly, the fact that we did not consider the expected utility in the optimization procedure does not lead to a significant loss in the expected return. This, of course, is due to the fact that we compare, for every realization, the candidate portfolio to the best one.

5. Conclusion and final remarks

The aim of this paper was to point out how to formulate the portfolio optimization in a general way, taking into account uncertainty about the underlying distribution parameters.

We proposed to use penalty function-like approaches to solve the general problem, focusing on two special decision criteria. In an applied example, numerical results were obtained, giving the hindsight that his approach might be quite promising. The comparison of our proposals with some traditional approaches in a real-market setting showed that they indeed result in more stable portfolios without a significant reduction in expected utility.

As was already pointed out by Scherer (2002) and Markowitz and Uskun (2003), approaches taking into consideration distribution uncertainty show better real-world performance, it might be fruitful to run a full-fledged empirical test for the Brazilian market in some future research.
References


