Subjective views and stress-tests for optimization of credit risk of loan portfolio

Master Thesis - Quantitative Finance
Erasmus University Rotterdam

Liudvika Liubeckiene
325804
liudvika.liubeckiene@gmail.com

Arco van Oord vanoord@ese.eur.nl
Dr. D.J.C. van Dijk djvandijk@ese.eur.nl

October 2, 2009

Abstract

I apply the methods to express subjective views in the spirit of Black-Litterman in the optimization of CVaR of a loan portfolio in order to explore the effects of the stress-tests of parameters defining the risk and return of a loan portfolio. The loan portfolio returns distribution is characterized by the joint default distribution, where each individual loan has a high probability of small profit (no default) and a low probability of big losses (default). Normality assumption does not hold thus the standard mean-variance optimization and Black-Litterman views cannot be applied. I simulate the joint distribution of the times to default and express views on different properties of the joint distribution of discrete default events: means, correlation, tail dependence. I optimize the CVaR of the original and stressed portfolios and compare the weights.

Contents

1 Introduction

The key idea of this work is to implement a subjective view expression on the joint default distribution with the purpose of optimizing the credit risk of a loan portfolio. Subjective views allow stress-testing: investigating how does the optimal portfolio change with respect to certain changes in the default distribution.

A loan is a financial instrument with the profit or loss it generates and the risk it carries. The risk is that the borrower defaults and is not able to repay all or a part of its debt. The risk of the portfolio of loans depends not only on the quality of individual borrower but also on the dependence between their probabilities to default. If we choose a risk measure and are able to determine each loan's contribution to overall portfolio risk, we evaluate how good certain loans are to our risk profile. There is nothing to prevent us to look at loan portfolio as something to be optimized with the help of the methods used in optimization of stocks and other instruments.

Optimizing the loan portfolio for the bank can result, most directly, in selling the least optimal (in chosen risk-return measure) loans. Results may facilitate rethinking of the lending strategy, by focusing to the sectors where loans have a better risk/return profile or by modifying the margin structure. ”Good” loans may also be bought.

”Return” of the loan is a variable which depends on whether the obligor defaults or not within a given time frame, usually a year. If not, we get the margin. The margin of the loan is the difference between the interest rate that bank borrows and the interest rate that bank lends. In the framework of risk-return optimization the margin is the equivalent of stock excess returns. If the obligor defaults, we lose the money, recovering only the amount specified by the recovery rate. With this definition of returns, the returns are already risk adjusted. The distribution of returns is therefore such that with high probability we gain small amounts and with low probability we suffer large losses. This distribution is highly skewed to the left - is far from normal and just as far from being elliptical. This obstacle must be overcome before applying methods from stock portfolio optimization directly.

In particular, we replace variance as risk measure with a risk measure which respects the assymmetry, for instance, Conditional Value-At-Risk (CVaR). In this thesis I discuss the issues related to the optimization of CVaR as a risk measure: what is CVaR, how we can write down the optimization problem, how the optimization problem is linearized, what are the features of the mean-CVaR optimal portfolio, especially the diversification, how the mean-CVaR efficient frontier is different from the mean-variance efficient frontier and what is the sensitivity of the composition of the optimal portfolio to the changes in parameters.

The joint distribution of loan returns is characterized by the joint distribution of defaults. Following Li (2000) [4], instead of working with the joint distribution of discrete default events within a year, I analyze the joint distribution of the times to default, which is conveniently split into marginals and a copula. I assume exponential marginals with fixed hazard rate parameters and a t-copula (as opposed to normal) for the dependence structure. A simulation panel consists of a set of scenarios of the times to default of each obligor which maintain the dependence structure imposed by the copula. The hazard rate parameters of the individual times to deault distributions are estimated from the known probabilities of default. The parameters of the t-copula - the degrees of freedom and the correlation matrix - are estimated from the dataset of the historical equity returns of the obligors. Here I employ the assumption introduced and verified by Mashal, Naldi and Zeevi (2003) [6] that the dependence structure of the times to default follows the dependence structure between the historical equity returns. Therefore, given the easily accesible equity returns, we can parametrize the distribution of the times to default, which in turn defines the discrete default events distribution, and further work with the latter.

The straightforward optimization of a loan portfolio already benefits the banks, but, just like with stocks, further advances are possible. One such methodology is to allow analysts to supplement the direct optimization with their own subjective opinions. Not only the result can be more understandable and realistic, it often becomes more robust and less sensitive to changes in inputs. In the area of stock portfolio management the method to input subjective views is widely established. The technique by Black and Litterman (1992) [8] allows portfolio managers to compute a posterior market distribution that smoothly blends their subjective views on the market with a prior market distribution, under the assumption that all the distributions involved are normal. Given that the loan return distribution is not normal, the Black-Litterman method cannot be applied here.

Meucci (2008) [1] Entropy Pooling (EP) is a way for a risk manager to express subjective views on the joint portfolio default distribution. EP extends Meucci (2006) [2] Copula Opinion Pooling technique, which provides an intuitive blending of the prior distribution and the views of the manager without requiring the normality assumption. The intuition behind their approach is to use opinion pooling criteria to determine the marginal distribution of each view separately, whereas the joint co-dependence, i.e. the copula is directly inherited from the prior market structure. A suitable change of coordinates translates the joint distribution of the views into a joint posterior distribution for the market. The market is represented by a large number of Monte-Carlo generated scenarios of dependent risk factors. The simulation of the market is easy once we have separated the joint multivariate distribution of risk factors into marginal distributions plus a copula. Entropy Pooling extends the Copula Opinion Pooling by permitting the views to be expressed on any property of the joint default distribution. In this thesis I apply Entropy Pooling on the joint default distribution to express views about its certain properties: mean, correlation, tail dependence.

I use scenario simulations to estimate portfolio CVaR. Each simulation scenario represents one particular instance of joint times to default. After the simulation, we transform the simulation panel of the times to default to a panel of joint discrete default events within one year. Before stress-testing, each scenario has equal probability of occurring. Stress-tests twist the probabilities - increases the likelihood of certain scenarios and decreases the likelihood of others, depending on what feature of distribution is represented by a view (stress-test), while the simulation panel remains the same. Thus the calculation of CVaR is computationally just as effective as for the original market. The prior and stressed posterior default distributions are useful for numerous purposes in risk management and portfolio optimization: the pricing of CDO and other multiname credit risk derivatives and economic capital calculations, among others. I concentrate on the optimization of the loan portfolio before and after stress testing.

I analyze 3 cases of stress-testing of the joint default distribution. First, increase probabilities of default, corresponding to a view on a mean as a parameter of the marginal distribution. Next, I stress the correlation matrix of the joint distribution of discrete default events. Finally, I stress the tail dependence of the joint distribution of discrete default events, allowing the default events to be related closer while the rest of the distribution stays the same. I compare the minimum CVaR portfolio of the original unstressed case and all the stress tests. The results on an artificial loan portfolio of real obligors with real default parameters but generated loan margins show that Entropy Pooling works well and the changes in the optimal portfolio indeed reflect the stress tests in the parameters. When the dependence between tail event increases, the least connected obligors increase their share in the portfolio, despite possibly lower returns. Besides, Entropy Pooling is very effective computationally, making it cheap to run many different tests and explore the effects in detail.

Moreover, I check how sensitive the CVaR optimum is with respect to the input returns. It is well known that the stock portfolio mean-variance optimum is highly sensitive to assumptions on returns and results in extremely different portfolios given a slight change in returns. For this reason the Black-Litterman method starts with ”reverse optimization” to determine the market equilibrium returns: the returns derived from the assumption that the market is in the mean-variance equilibrium. Market equilibrium returns instead of the historical returns are considered the prior - this is the other contribution of the Black-Litterman model besides the capability to express the views. While I allow the expression of the views on the loan returns distribution using Entropy Pooling as the fitting alternative to the Black-Litterman model, I also put effort in determining mean-CVaR equilibrium returns by the means of linear programming inverse optimization techniques. The results of this part are incomplete and inconclusive, suggesting that further research is needed.

My main contribution to the existing theory is the idea of applying Entropy Pooling to express the views on the joint default distribution to allow the stress-testing of the loan portfolio. It is not easy for banks to obtain the correlation matrix of default events and have confidence in it. I show that we can start with the more robustly parametrized joint distribution of the times to default (where the dependence structure is inherited from the equity returns), smoothly translate it to the distribution between the discrete default events, and stress test the latter with intuitive changes in its properties. The work carries practical character and is readily applicable for bank practitioners wishing to strengthen the lending strategy against financial crises.

The remainder of the thesis is organized as follows. Section 2 describes the CVaR as the risk measure and discusses the methodology of CVaR optimization. In section 3 I show how to estimate the parameters of the t-copula. Section 4 deals with subjective views: Black-Litterman, Copula Opinion Pooling, Entropy Pooling. Section 5 gives a short summary of efforts to derive mean-CVaR equilibrium returns (reader may check appendix B for detailed analysis). Section 6 describes the data and section 7 presents numerical results. Finally, main conclusions and limitations are listed in section 8.

2 Risk of a loan portfolio

In this section we present a typical loan portfolio and how to measure and minimize its risks.

2.1 CVaR as a risk measure

Next we decide how to measure the risk of a loan portfolio. Current regulations for finance businesses (Basel II accord in particular) formulate some of the risk management requirements in terms of percentiles of loss distributions. An upper percentile of the loss distribution is called Value-at-Risk (VaR). The popularity of VaR is mostly related to a simple and easy to understand representation of high losses. VaR can be quite efficiently estimated and managed when underlying risk factors are normally (log-normally) distributed. However, for non-normal distributions, VaR may have undesirable properties (Artzner at al. (1997, 1999) [18]) such as lack of sub-additivity, i.e., VaR of a portfolio with two instruments may be greater than the sum of individual VaRs of these two instruments. Intuitively, we expect that the risk of the portfolio with two instruments decreases inversely proportional to the correlation between the instruments, and VaR does not necessarily reflect this insight. Also, VaR is difficult to control/optimize for discrete distributions, when it is calculated using scenarios. In this case, VaR is non-convex and non-smooth as a function of positions, and has multiple local extrema (Rockafeller and Uryasev (2000) [9]).

VaR is also criticized for not taking into account the magnitude of losses when the VaR is exceeded. VaR provides no insight into what would happen to a bank if a 1 in 1000 chance event occurs. Since Rockafellar and Uryasev (2000) [9] suggested to use, as a supplement (or alternative) to VaR, another percentile risk measure which is called Conditional Value-at-Risk, it has gained wide popularity and acceptance. The CVaR risk measure is closely related to VaR. For continuous distributions, CVaR is defined as the conditional expected loss under the condition that it exceeds VaR. For general distributions, including discrete distributions, CVaR is defined as the weighted average of VaR and losses strictly exceeding VaR, see Rockafellar and Uryasev (2000) [9].

CVaR has more attractive properties than VaR. CVaR is sub-additive and convex (Rockafellar and Uryasev, 2000) [9]. Moreover, CVaR is a coherent measure of risk in the sense of Artzner et al. (1997, 1999) [18]. We also hypothesize that CVaR is a more robust risk measure than VaR. The reason is that all values in the tail of the distribution of returns are considered when estimating CVaR, compared to just the number of values for the case of VaR. For example, if the tail consists of the returns -8%, -9.5% and -11%, all these values are taken into account when estimating CVaR. When VaR is estimated, the most important feature about the tail is that it consists of three different (in this example) values. It should be noted, that it is the tail of the distribution that is important for most risk measures, since the tail in some sense defines the risk. The variance of the estimate of a mean (e.g. CVaR) should, intuitively, be less than the variance of the estimation of a single point (e.g. VaR), and therefore the variance should be lower for CVaR compared to VaR. Hence CVaR ought to be a more robust measure of risk than VaR.

Another risk measure CVaR outcompetes is variance. Despite common usage of the variance as a risk measure for stock portfolio optimization, instead of optimizing according to the mean-variance model, a portfolio can be optimized in other frameworks. Krokhmal, Palmquist and Uryasev (2002) [11] illustrate the relation of the mean-CVaR approach to the standard Markowitz mean-variance (MV) framework. Rockafellar and Uryasev (2000) [9] show that for normally distributed loss functions these two methodologies are equivalent in the sense that they generate the same efficient frontier. However, in the case of nonnormal, and especially non-symmetric distributions, CVaR and MV portfolio optimization approaches may reveal significant differences. Indeed, the CVaR optimization technique concentrates on one tail of the loss distribution, which corresponds to high losses, and does not account for the opposite tail representing high profits. On the contrary, the Markowitz approach defines the risk as the variance of the loss distribution, and since the variance incorporates information from both tails, it is affected by high gains as well as by high losses. Due to the asymmetry of loan returns, the traditional mean-variance framework may produce misleading results.

Mathematically, we define VaR as

$$P[x \leq \mathrm{VaR}_\beta] = 1 - \beta, \tag{1}$$

or equivalently,

$$\int_{-\infty}^{\mathrm{VaR}_\beta} f_X(x)\mathrm{d}x = 1 - \beta, \tag{2}$$

where $f_X(x)$ is a probability distribution function of returns x and β is the desired VaR level (95%, 99%).

We define CVaR as

$$\mathrm{CVaR}_\beta(x) = E[x | x \leq \mathrm{VaR}_\beta], \tag{3}$$

or equivalently,

$$\mathrm{CVaR}_\beta(x) = \frac{1}{1-\beta}\int_{-\infty}^{\mathrm{VaR}_\beta} x f_X(x)\mathrm{d}x \tag{4}$$

2.2 CVaR optimization

We implement the scenario-based model proposed by Rockafellar and Uryasev (2000) [9] for portfolio CVaR minimization.

Let $f(\mathbf{x}, y)$ be the loss associated with the decision vector $\mathbf{x}$, to be chosen from a certain subset $\mathbf{X}$ of $\mathbf{R}^n$, and the random vector y in $\mathbf{R}^m$. The vector $\mathbf{x}$ can be interpreted as representing a portfolio, with $\mathbf{X}$ as the set of available portfolios (subject to various constraints), but other interpretations could be made as well. The vector y stands for the uncertainties, e.g. in market parameters, that can affect the loss. Of course the loss might be negative and thus, in effect, constitute a gain. For each $\mathbf{x}$, the loss $f(\mathbf{x}, y)$ is a random variable having a distribution in $\mathbf{R}$ induced by that of y. The underlying probability distribution of $y \in \mathbf{R}^m$ is assumed for convenience to have density, which we denote by $p(y)$. However, as it will be shown later, an analytical expression $p(y)$ for the implementation of the approach is not needed. It is enough to have an algorithm (code) which generates random samples from $p(y)$.

It is difficult to work with CVaR as per (4), because of the VaR function

$$\alpha_\beta(\mathbf{x}) = \min_{\alpha \in R}(\text{loss} \geq \beta) \tag{5}$$

involved in its definition. Rockafeller and Uryasev (2000) [9] propose much simpler function $F_\beta(\mathbf{x}, \alpha)$:

$$F_\beta(\mathbf{x}, \alpha) = \alpha + \frac{1}{1-\beta} \int_{y \in \mathbf{R}} [f(\mathbf{x}, y) - \alpha]^+ p(y) \mathrm{d}y \tag{6}$$

I.e., α is the β-VaR: the losses at percentile β of the loss function. Then β-CVaR of the loss associated with any $\mathbf{x} \in \mathbf{X}$ can be determined by the formula

$$\text{CVaR}_\beta(\mathbf{x}) = \min_{\alpha \in R} F_\beta(\mathbf{x}, \alpha) \tag{7}$$

In particular, it can be proven that one always has

$$\alpha_\beta(\mathbf{x}) = \min_{\alpha \in R} F_\beta(\mathbf{x}, \alpha), \quad \text{and} \quad \text{CVaR}_\beta(\mathbf{x}) = F_\beta(\mathbf{x}, \alpha_\beta(\mathbf{x})) \tag{8}$$

Though it may seem clear that $F_\beta(\mathbf{x}, \alpha)$ is equal to $\text{CVaR}_\beta(\mathbf{x})$, the demonstration of it is not trivial and can be found in Rockafellar and Uryasev (2000) [9].

Furthermore, the integral in the definition (6) of $F_\beta(\mathbf{x}, \alpha)$ can be approximated in various ways. For example, this can be done by sampling the probability distribution of y according to its density $p(y)$. If the sampling generates a collection of vectors $y_1, y_2, \ldots y_J$, then the corresponding approximation to $F_\beta(\mathbf{x}, \alpha)$ is

$$\tilde{F}_\beta(\mathbf{x}, \alpha) = \alpha \frac{1}{1-\beta} \sum_{k=1}^{J} p(y)[f(\mathbf{x}, y) - \alpha]^+ \tag{9}$$

Finally, minimizing CVaR of the loss associated with $\mathbf{x}$ over all $\mathbf{x} \in \mathbf{X}$ is equivalent to minimizing $F_\beta(\mathbf{x}; \alpha)$ over all $(\mathbf{x}, \alpha) \in (\mathbf{X} \times R)$, in the sense that

$$\min_{\mathbf{x} \in \mathbf{X}} \mathrm{CVaR}_\beta(\mathbf{x}) = \min_{(\mathbf{x},\alpha) \in (\mathbf{X}, R)} F_\beta(\mathbf{x}, \alpha) \tag{10}$$

When the analytical expression of density $p(y)$ is not known, we use historical observations of prices of instruments or we may use Monte Carlo simulations. In this case we use the approximation $\tilde{F}_\beta(\mathbf{x}, \alpha)$. If the function $f(\mathbf{x}, y)$ is convex with respect to $\mathbf{x}$, then $\tilde{F}_\beta(\mathbf{x}, \alpha)$ is a convex nonsmooth function with respect to the vector $(\mathbf{x}, \alpha)$. Therefore, if the feasible set $\mathbf{X}$ is convex, the optimization problem with the CVaR performance function can be solved using nonsmooth optimization techniques.

Moreover, if the function $f(\mathbf{x}, y)$ is linear with respect to $\mathbf{x}$, these problems can be solved using Linear Programming (LP) techniques (Rockafeller and Uryasev (2000) [9]). The possibility of such reduction to linear programming does not depend on a special distribution of y, such as a normal distribution; it works for nonnormal distributions just as well. However, linearization adds a variable for each scenario, thus for problems involving low number of instruments but many scenarios nonsmooth techniques compete with linear programming in terms of the computation time.

I minimize CVaR with respect to a constraint on returns. I build the efficient portfolio frontier in the mean-CVaR space by repeating the CVaR minimization procedure for different values of the expected return of the portfolio, r.

2.3 Linearization of CVaR optimization problem

In this section and onwards $f(\mathbf{x}, y)$ denotes the returns, not the loss as per Rockafeller and Uryasev (2000) [9] definition, because it is easier to relate to returns distribution. Consequently, if VaR is defined as a quantile of the returns distribution instead of the loss distribution, we must invert its sign as well. ζ is minus α of the previous section.

From the linearization of the CVaR function (9) we obtain the linear form of the CVaR minimization problem.

$$\begin{array}{rll} Minimize & \zeta + \frac{1}{1-\beta} \sum_{j=1}^{J} p_j z_j & \\ s.t. & z_j + \zeta + f(\mathbf{x}, y_j) \geq 0 & \forall j \in J \\ & f(\mathbf{x}, y) \geq r & \\ & z_j \geq 0 & \forall j \in J \end{array} \tag{11}$$

where r is the desired level or returns and z_j are artificial linearization variables:

$$z_j = [\zeta + f(\mathbf{x}, y_j)]^+ \tag{12}$$

which together with ζ become the decision variables.

Alternatively, we could fix the CVaR at a desired level ω and maximize the portfolio return $f(\mathbf{x}, y)$:

$$\begin{array}{rl} Maximize & f(\mathbf{x}, y) \\ s.t. & \zeta + \frac{1}{1-\beta} \sum_{j=1}^{J} p_j z_j \leq \omega \\ & z_j + \zeta + f(\mathbf{x}, y_j) \geq 0 \quad \forall j \in J \\ & z_j \geq 0 \quad \forall j \in J \end{array} \tag{13}$$

The real portfolio holdings are $\mathbf{x}$ - we must expand the $f(\mathbf{x}, y)$ in order to specify the $\mathbf{x}$. In the scenario based approach every scenario is a certain outcome of the market, dependent on the uncertainties y, which determine the multivariate returns. Let us also assume we simulate market outcome panel $\mathbf{M}$: $J \times K$ matrix, where J is the number of simulations and K is the number of obligors. The elements of panel $\mathbf{M}$ are $m_{ij},\ i = 1, \ldots, K,\ j = i = 1, \ldots, J$. One scenario can also be presented as vector $\mathbf{x} = (m_{1j}, \ldots, m_{Kj})$. Then

$$f(\mathbf{x}, y) = \sum_{j=1}^{J} p_j (\sum_{i=1}^{K} w_i m_{ij}) \tag{14}$$

In other words, the returns is the average over all scenarios of market outcomes per obligor multiplied by the weights of the obligors. $f(\mathbf{x}, y_j)$ is the j-th scenario's portfolio return:

$$f(\mathbf{x}, y_j) = \sum_{i=1}^{K} w_i m_{ij} \tag{15}$$

Usually, but not necessarily, $p_j = 1/J$. We will encounter twisted (stressed) p_j in the Entropy Pooling problem where we reuse the same simulated market outcome scenarios with new probabilities to express the stressed joint default distribution.

$f(\mathbf{x}, y)$ expanded into a function of the market outcome panel $\mathbf{M}$ I write (11) specifying all the $K + 1 + J$ variables involved: K portfolio holdings w_i, 1 ζ for VaR and J dummies z_j.

$$\begin{array}{rll} Minimize & \zeta + \frac{1}{1-\beta} \sum_{j=1}^{J} p_j z_j & \\ s.t. & z_j + \zeta + \sum_{i=1}^{K} w_i m_{ij} \geq 0 & \forall j \in J \\ & \sum_{j=1}^{J} p_j (\sum_{i=1}^{K} w_i m_{ij}) \geq r & \\ & \sum_{i=1}^{K} w_i = 1 & \\ & w_i \geq 0,\ \zeta \geq 0,\ z_j \geq 0 & \forall i \in K,\ \forall j \in J \end{array} \tag{16}$$

Two notable differences from (11) have appeared:

- Full investment constraint $\sum_{i=1}^{K} w_i = 1$.

- Requirement $w_i \geq 0$. In business terms it means that we deny negative loans, for instance, when bank borrows from another bank. Negative loan is similar to short selling, we would hope for lender to default.

The size of the decision variable now is $K+1+J+1+J$. K is the number of obligors, so first K of the final weight vector will be the true portfolio holdings. Next, $K+1$'th entry is VaR, referred to as ζ. Next J entries are the z_j, the dummies resulting from CVaR function linearization. The first K entries of the solution $\mathbf{x}^* = (x_1, \ldots, x_K)$ is the vector of weights that define mean-CVaR optimal portfolio.

2.4 CVaR optimization examples

Usually the CVaR optimization results in a balanced and sensible portfolio, given that the constituents of the portfolio present a balanced trade-off between risk and return. The higher the default probability, the higher the return (the margin bank sets lending to this obligor) and vice versa. The higher the desired level of returns, the higher the investment into high risk-high return obligor, and vice versa. However, there is some research suggesting that sometimes the portfolios optimized in CVaR sense can be very concentrated. For instance Romero-Meza and Laengle (2007) [17] find that half of the assets in their portfolio receive zero weights and hypothesize that this is due to instablility of foreign currencies against Chilean Peso.

Another reason could be a tendency of CVaR optimization to produce concentrated portfolios if risk-return balance is a little unfavourable for certain assets.

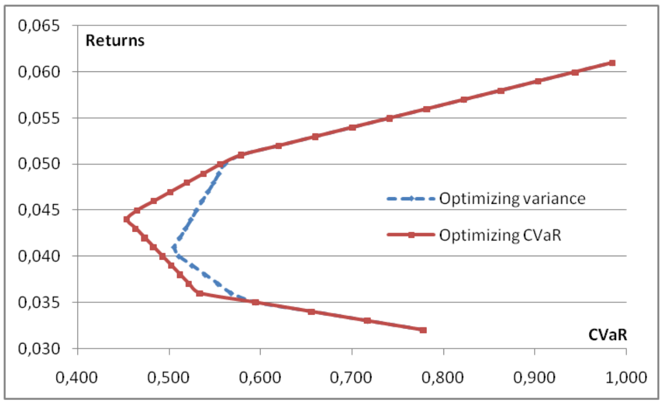
I start with a simple example with just 3 obligors in order to see what happens when the portfolio is simplified to the very minimum. This way I can disentangle the effects of all the parameters. Assume we have 3 obligors. Correlations equal (=0.3367 off-diagonal), recovery rates equal (=0%). Returns and PDs are all different and calibrated to represent realistic trade-off between risk and return. Table 1 gives PDs and returns of 3 obligors.

Obligor	Returns before risk adjustment	PD	Risk adjusted returns
1	9.00%	2.62%	6.14%
2	5.00%	0.90%	4.06%
3	4.00%	0.81%	3.15%

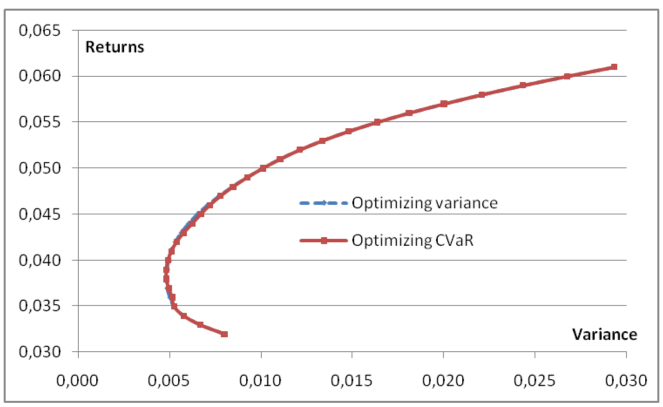
Table 1: 3 obligor problem: risk and return

The risk adjusted returns are derived from the panel of the joint default scenarios, where the dependence between defaults is modelled by a t-copula with 3 degrees of freedom. In every scenario each obligor either defaults or not. If not, we get the expected returns (9%, 5%, 4% respectively). If defaults, we lose the loan, recovering only the amount specified by the recovery rate, which in this case is 0%. The distribution of returns is therefore such that with high probability we gain small amounts and with low probability we suffer large losses.

I illustrate the efficient frontier of this portfolio in the mean-CVaR setup in figure 1. For comparison I also show the mean-variance frontier.



(a) Mean-CVaR efficient frontier



(b) Mean-variance efficient frontier

Figure 1: Efficient frontiers. Solid line in both pictures is obtained optimizing CVaR, and dotted line optimizing variance.

We see that when we optimize the CVaR, resulting portfolios' variance is practically indistinguishable from minimal. However if the variance is optimized, resulting CVaR can be much further away from optimal. At a nearly minimum CVaR point (returns: 0.044) CVaR of the minimum variance portfolio is higher than the CVaR of the CVaR-optimal portfolio by 15%.

The diversification of the 3 obligor portfolio is displayed in figure 2: it shows which parts of the mean-CVaR efficient frontier are constructed of which obligors. In the redundant bottom branch where returns are low but CVaR is high only obligors 2 and 3 (low returns, lower but not so low risk) are used. The upper part is contributed by obligors 2 and 3 (higher returns, higher risk - the typical trade-off). The middle portfolios use all 3 obligors. In this case the equal weights (=1/3) portfolio would be quite close to the tip point of the efficient frontier. It means that all the obligors present a healthy mix of risk and return. Thus we observe that in a healthy (and already quite optimal) case where different obligors do not stand out by their good or bad risk profile, but contribute to trade-off in a proportionate manner, the minimum CVaR portfolio is quite well diversified. Concentrated portfolios most likely arise from the fact that discarded assets have a (slightly) suboptimal risk-return profile.

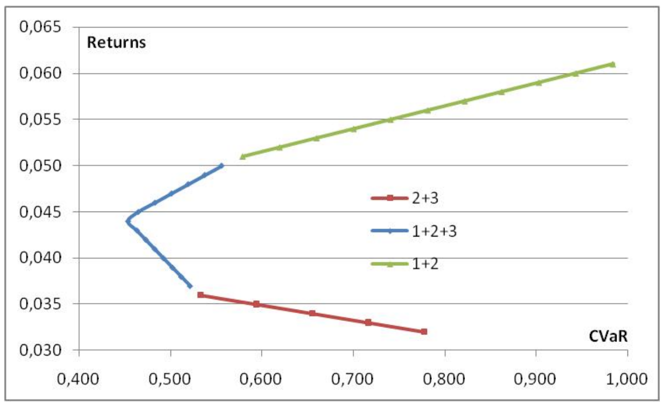


Figure 2: How many obligors contribute to portfolio at different return-CVaR levels

Also, note that efficient frontier is indeed convex and piecewise-linear, at least in 3 obligors case. It should enable us to find a point where CVaR is lowest overall - similarly to the lowest variance point in the mean-variance frontier.

Next I try to reproduce the research of DiClemente and Romano (2004) [7] who analyze loan portfolio optimization, show significant benefits of performing it and present a balanced portfolio as a result. The authors present all the necessary data components in the paper. Table 2 presents the findings in terms of the weights of the optimal portfolios with different return constraints. To be true to the original article, DiClemente and Romano (2004) [7] consider that the returns in this context are pure loan margins without default events, i.e. returns before risk adjustment. In this case 7.177% is the return of the equal weights portfolio. This paradigm is hard to justify in business sense, as it means that in the case of obligor default we still earn the margin, but it makes certain computational sense because this way we can separate risk from return. Still, the difference in return level specification does not impact the weights of the portfolio: if I fix the returns before risk adjustment at the level that corresponds the risk adjusted returns level of say 1%, I get the same weights as displayed in column 1 where the risk adjusted returns are constrained.

I would expect column 3 to match column 6 exactly, i.e. DiClemente and Romano (2004) [7] results to be reproducible, but this was not the case. I cannot find the diversification. The portfolio consists purely of the lowest risk/lowest return obligor

	returns					
					With 29.35% upper limit	DR
obligor	1%	2%	2.93%	3%	2.93%	7.177%
1	0.9089	0.5591	0.2326	0.2094	0.2053	0.2612
2	0	0	0	0	0.0075	0.0020
3	0	0	0	0	0.1594	0.1242
4	0	0	0	0	0.1441	0.1618
5	0	0	0	0	0	0
6	0	0	0	0	0.0064	0
7	0	0	0	0	0.1805	0.1573
8	0.0911	0.4408	0.7674	0.7906	0.2935	0.2935
9	0	0	0	0	0	0
10	0	0	0	0	0.0032	0

Table 2: DiClemente and Romano (2004) [7] portfolios. First 4 columns are plain CVaR minimum weights subject to the given return level. The level 2.93% corresponds to the risk adjusted return of the portfolio with equal weights, which is a good benchmark. Column 5 contains CVaR minimum weights with an upper limit on investment into a single asset, namely 29.35%. This number is the level of largest investment in column 6, which reprints results from the article of DiClemente and Romano (2004) [7] where 7.177% is the returns before risk adjustment.

and the highest risk/highest return obligor. The pattern remains over all set levels of returns, only the ratio changes in favor of the low risk obligor when we request a lower return overall. I cannot find the reason for the discrepancy.

I hypothesize that other obligors may have less favourable risk/return trade-off - then the concentration is quite a typical result to find. Usually it is countered by setting an upper limit of an investment into one asset. For example, Krokhmal, Palmquist and Uryasev (2002) [11] set the maximum limit of one investment to 20%. Their numerical results show that the limit is reached in every scenario with every constraint on return (their table 1). The column 5 in the table 2 displays the diversifying impact of the uper limit too.

2.5 Effects of correlation

Before I have assumed the pairwise correlation between the times to default of all 3 obligors to be equal. In the next table (Table 3) I show how changes in the correlation matrix affect the weights of the mean-CVaR optimal portfolio. It will help understanding the results of the correlation stress test later. "Correlation" is the correlation matrix parameter of the t-copula. It translates almost directly to the correlation between the times to default, although not precisely so. For instance, in scenario 1 the off-diagonal elements of the empirical correlation matrix between the times to default turn out between 0.31 and 0.32; and in scenario 2 - between 0.01 and 0.02. ”Default correlation” is the empirical correlation between default events within one year, i.e. 1 when the time to default is less than 1 and 0 otherwise. The correlation between the default occurrences is very much lower than the correlation between the times to default. Column ”D” contains the counts of the default occurrences together: the first row is the number of simulations with at least 1 one obligor defaulting, the second - the number of scenarios with at least 2 obligors defaulting, the third - the number of scenarios with 3 obligors defaulting. The numbers sum up to the total count of the default occurences over the simulation panel. ”CVaR” and w_{opt} are the CVaR and the weights of the mean-CVaR optimum portfolio, respectively. I impose returns of 4.9% and do 100 000 simulations. The random numbers for the simulations are the same in all 10 scenarios.

Ignoring minor differences, there are two distinct patterns of the weights over 10 scenarios: one with obligors 1 and 2 sharing 90% of portfolio almost equally, and another with obligor 3 gaining more weight and obligor 2 not invested at all. This structural break occurs when obligor 3 (with the lowest risk and return) is significantly less correlated to 1 and 2 than 1 and 2 between themselves; or when the correlation overall is very high. Normally, as the demanded return level is higher than the equal weights portfolio return (4.4%), portfolio is mostly invested into the higher risk-return profile obligors 1 and 2. This typical pattern is retained with the typical correlation structures. Thus the findings confirm the intuitive insight that when correlations get higher than typical, the least ”connected” obligor is favoured, despite of its possibly lower returns.

3 Copulas

3.1 Copula of time to default distribution

We now turn to the question how to generate the joint default scenarios. The article by Li (2000) [4] transformed the understanding and modeling of joint default completely. Li’s assumption that codependence between defaults is described by a normal copula is largely discredited and even blamed for ruining the Wall Street, but otherwise, using more fitting classes of copulas, the copula approach yields extremely satisfactory results. The ease of manipulation and simulation offered by the copula approach is compelling.

First, Li (2000) [4] introduces a random variable ”time-until-default”, or survival time of the defaultable entity or financial instrument. A discrete default event within a given time frame (usually, one year) is completely defined by the time to default.

Nr	Correlation			Default correlation			D	CVaR	w_{opt}
1	1	0.3367	0.3367	1	0.1229	0.1089	3872	0.5404	0.4416
	0.3367	1	0.3367	0.1229	1	0.1176	412		0.4584
	0.3367	0.3367	1	0.1089	0.1176	1	42		0.1000
2	1	0	0	1	0.0330	0.0262	4142	0.4560	0.4392
	0	1	0	0.0330	1	0.0255	154		0.4559
	0	0	1	0.0262	0.0255	1	3		0.1048
3	1	0.6667	0.6667	1	0.2988	0.2885	3343	0.6944	0.4411
	0.6667	1	0.6667	0.2988	1	0.2937	794		0.4579
	0.6667	0.6667	1	0.2885	0.2937	1	189		0.1009
4	1	0.9999	0.9999	1	0.5812	0.5514	2584	0.9115	0.5744
	0.9999	1	0.9999	0.5812	1	0.9487	888		0.0000
	0.9999	0.9999	1	0.5514	0.9487	1	800		0.4256
5	1	0.6667	0	1	0.2988	0.0262	3766	0.5875	0.5784
	0.6667	1	0	0.2988	1	0.0263	532		0.0000
	0	0	1	0.0262	0.0263	1	14		0.4216
6	1	0.3367	0	1	0.1229	0.0262	4028	0.5230	0.4414
	0.3367	1	0	0.1229	1	0.0239	283		0.4582
	0	0	1	0.0262	0.0239	1	6		0.1004
7	1	0	0.3367	1	0.0330	0.1149	4033	0.4702	0.4397
	0	1	0	0.0330	1	0.0262	279		0.4565
	0.3367	0	1	0.1149	0.0262	1	6		0.1038
8	1	0	0	1	0.0330	0.0251	4052	0.4617	0.4384
	0	1	0.3367	0.0330	1	0.1055	209		0.4551
	0	0.3367	1	0.0251	0.1055	1	7		0.1066
9	1	0.6667	0.6667	1	0.2988	0.2727	3415	0.6830	0.4403
	0.6667	1	0	0.2988	1	0.0280	849		0.4571
	0.6667	0	1	0.2727	0.0280	1	31		0.1026
10	1	0.6667	0	1	0.2988	0.0321	3533	0.5864	0.5742
	0.6667	1	0.6667	0.2988	1	0.3209	686		0.0000
	0	0.6667	1	0.0321	0.3209	1	64		0.4258

Table 3: 3 obligor problem: correlation impact on mean-CVaR optimal weights. ”Correlation” is the correlation matrix parameter of the t-copula. ”Default correlation” is the empirical correlation between default events within one year. ”D” is the count of simulations with at least 1, 2 and 3 defaults. ”CVaR” and w_{opt} are the CVaR and the weights of the mean-CVaR optimum portfolio, respectively.

Each time until default separately can be described in terms of its distribution function.

$$F(t) = Pr(T \leq t), \quad t \geq 0 \tag{17}$$

and a corresponding density function $f(x)$. Further Li (2000) [4] specifies a hazard rate function

$$h(x) = \frac{f(x)}{1 - F(x)}, \tag{18}$$

which has a conditional probability density interpretation: it gives the value of the conditional probability density function of T at exact age x, given survival to that time. Li (2000) [4] shows that

$$F(t) = 1 - e^{-\int_0^t h(s)ds}, \quad t \geq 0 \tag{19}$$

A typical assumption is that the hazard rate is a constant, h, over certain period, such as $[x, x+1]$. In this case, the distribution function is

$$F(t) = 1 - e^{-ht} \tag{20}$$

which shows that the survival time follows an exponential distribution with the parameter h. Under this assumption, the survival probability over the time interval $[x, x+t]$ for $0 < t \leq 1$ is e^{-ht}.

Modelling a default process is equivalent to modelling a hazard function. Li (2000) [4] lists several reasons why modelling the hazard rate function is a good idea. I take this proposal too. A simple way to model the hazard rate is to assume it is fixed and derive it from the known one year probabilities of default (PD). As PD $= F(1)$, we see from (20) that $h = -ln(1 - \text{PD})$.

The default correlation between two entities A and B can be defined with respect to their survival times (=times until default), T_A and T_B:

$$\rho_{AB} = \frac{Cov(T_A, T_B)}{\sqrt{Var(T_A)Var(T_B)}} \tag{21}$$

Li (2000) [4] calls this definition of the default correlation ”survival time correlation” and shows that it is a much more general concept than the discrete default correlation - correlation between default events within given time frame. We use this default correlation definition too, only replace the wording ”survival time” to ”time to default”.

The core issue in modelling defaults is to determine the dependence between the default occurrences. Generally speaking, given marginal distributions for every obligor and a default correlation matrix, we still cannot work out the joint default distribution function. Li (2000) [4] popularized simple and convenient approach to the problem - copulas. I give a brief overview of copulas here.

A copula function is a function that connects univariate marginals to their full multivariate distribution. A copula allows us to model marginal distribution functions of each time to default and default codependence separately. A copula is a mapping $C : [0, 1]^n \rightarrow [0, 1]$ with the following characteristics:

1. $C(u_1, \ldots, u_n)$ is non-decreasing in each component. It is a multivariate cumulative distribution.
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i \quad \forall i$. The copulas respects the given marginals.
3. $C(u_1, \ldots, u_i = 0, \ldots, u_n) = 0 \quad \forall i$. It is a multivariate cumulative distribution.
4. $\forall [a_1, \ldots, a_n], [b_1, \ldots, b_n] \in [0, 1]^n$, where $a_i \leq b_i$, the following holds: $\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\ldots+i_n} C(u_{1,i_1}, \ldots, u_{n,i_n}) \geq 0$, where $u_{j,1} = a_j$ and $u_{j,2} = b_j$. This implies a positive "area" under the copula function.

Copula functions can be used to link marginal distributions with a joint distribution. For a given univariate marginal distribution functions $F_1(x_1), F_2(x_2), \ldots, F_m(x_m)$, the function $C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)) = F(x_1, x_2, \ldots, x_m)$, which is defined using a copula function C, results in a multivariate distribution function with an univariate marginal distributions as specified by $F_1(x_1), F_2(x_2), \ldots, F_m(x_m)$.

The converse also holds, established by Sklar (1959). Any multivariate distribution function F can be written in the form of a copula function. He proved the following: If $F(x_1, x_2, \ldots x_m)$ is a joint multivariate distribution function with univariate marginal distribution functions $F_1(x_1), F_2(x_2), \ldots, F_m(x_m)$, then there exists a copula function $C(u_1, u_2, \ldots, u_m)$ such that $F(x_1, x_2, \ldots, x_m) = C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m))$.

We also use the property that copulas are invariant under strictly increasing transformation of the marginals.

One rather seldom sees copula function as such, per definition above. More often its density is used in applications, primarily to simulate random variates with a desired codependence. Figure 3 is a poster picture of normal and t-copulas with two variables, though what is in the picture should more strictly be called a simulation from a copula, revealing the copula density.

There is a wide family of known and popular copula functions. Li (2000) [4] had good reasons to use normal copula to model defaults, however, just like with equity returns, normality assumption does not hold very well. Mashal and Naldi (2002) [5] give extensive explanation with examples how t-copula fits better. For example, if one uses the t-distribution with 10 degrees of freedom and 30% correlation, the probability of having a joint tail event, given that a marginal tail event occurred, is about 3%; this is very different from the 0% that the normality entails. Figure 3 illustrates a higher probability of joint tail events in a t-copula as well.

The distribution of marginals do not have to be equal, and marginal and joint distributions do not have to come from the same family. In fact, while our marginals representing the distribution of the times to default are exponential, there is no such

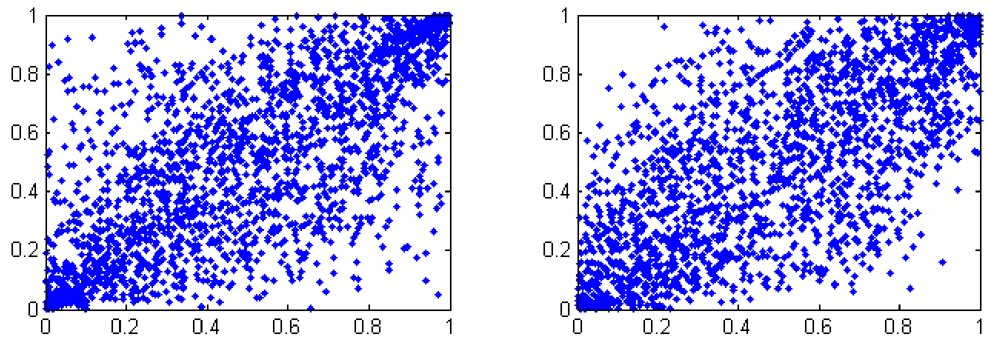


Figure 3: Simulation from a copula: t-copula(left) and normal copula(right)

thing as a multivariate exponential distribution - and this is another argument for using a t-copula.

3.2 t-Copula correlation parameter estimation

We have already established that the normal copula to represent the default correlation is not the prime choice, and that the t-copula, following the analogy with equity returns, is the next thing to try.

Mashal, Naldi and Zeevi (2003) [6] argue that the dependence between the times to default has the same structure as the dependence between asset returns. If we think of defaults as being generated by asset values falling below a given boundary, than the probabilities of joint defaults over a specific horizon must follow the joint dynamics of asset values. They extend the argument further and check the presumption that asset return dependence can be substituted by equity return dependence. The assumption has long been used by practitioners, because equity returns are observable and asset returns are not. It is also criticized because of the different leverage between assets and equity. Mashal, Naldi and Zeevi (2003) [6] find the similarities between the joint tail dependence (as measured by the DOF) of asset and equity returns quite striking and conclude that the empirical evidence strongly supports the widespread assumption of substitution.

There is a rich body of research on which distribution fits equity returns best, with t-distribution being one of the reasonably simple options that fits well. Another necessary assumption is that in the risk-neutral world the dependence stays the same as in the risky world.

Next we will extract t-copula parameters - the correlation matrix and the degrees of freedom - from the historical series of equity returns, following Meucci (2008) [3]. Mashal and Naldi (2002) [5] provide similar procedure, but Meucci (2008) [3] goes further by imposing structure on correlations.

Let us have the market $\mathbf{X}$, represented by N risk factors. Our market $\mathbf{X}$ is such that $\mathbf{Y} = \Phi_\nu^{-1}(\mathbf{X})$, where Φ_ν is a cdf of the univariate t distribution. $\mathbf{Y} \sim t(\nu, \mathbf{0}, \mathbf{C})$: multivariate t distribution with ν degrees of freedom, zero means and correlation matrix $\mathbf{C}$. Assume the marginals are known or estimated empirically and presented as cdf's: $\hat{F}_1, \ldots, \hat{F}_n$. If the number of degrees of freedom ν is known, we can transform the data to form $T \times N$ panel $\mathbf{Y}$, defined entry-wise as follows:

$$Y_{t,n} = \Phi_\nu^{-1}(\hat{F}_n(X_{t,n})) \tag{22}$$

$\mathbf{Y}$ is a time series of independent identically distributed joint multivariate t variables. Meucci (2008) [3] shows that for given degrees of freedom ν, Maximum Likelihood estimator of the correlation parameter $\mathbf{C}$ is

$$\hat{\mathbf{C}} = \frac{1}{T} \sum_{t=1}^{T} w_t \mathbf{y}_t \mathbf{y}_t^{'} \tag{23}$$

where $\mathbf{y}_t$ denotes transpose of the t-th row in panel Y and the weights w_t are defined as follows:

$$w_t = \frac{\nu + N}{\nu + \mathbf{y}_t^{'} \hat{\mathbf{C}}^{-1} \mathbf{y}_t} \tag{24}$$

This definition cannot be solved analytically, except in the normal case ($\nu = \infty$). Meucci (2008) [3] gives following recursive numerical estimation procedure:

Step 0. Set $u = 0$ and initialize $\hat{\mathbf{C}}_{(u)}$ as the sample correlation.
Step 1. Compute the weights

$$w_t^{(u)} = \frac{\nu + N}{\nu + \mathbf{y}_t^{'} \hat{\mathbf{C}}_{(u)}^{-1} \mathbf{y}_t} \tag{25}$$

Step 2. Compute the scatter matrix

$$\hat{\Sigma}_{(u+1)} = \frac{1}{T} \sum_{t=1}^{T} w_t^{(u)} \mathbf{y}_t \mathbf{y}_t^{'} \tag{26}$$

Step 3. Extract the correlation from the following relation

$$\hat{\Sigma}_{(u+1)} = diag(\hat{\sigma}) \hat{\mathbf{C}}_{(u+1)} diag(\hat{\sigma}) \tag{27}$$

Step 4. Check for convergence

$$d = \sqrt{\frac{1}{N} tr(\hat{\mathbf{C}}_{(u+1)} - \hat{\mathbf{C}}_{(u)})^2} \tag{28}$$

If d is less than desired threshold, stop. Otherwise, increment $u = u + 1$ and continue from step 1.

The estimate $\hat{\mathbf{C}}(\nu)$ is a function of degrees of freedom ν that we assumed in the beginning. To estimate ν, loop over candidates (say 2 to 50) to find the value maximizing the log-likelihood

$$\frac{1}{T}\sum_{t=1}^{T} ln(f_{\nu,\mathbf{0},\mathbf{C}}(\mathbf{y}_t)) \tag{29}$$

where $f_{\nu,\mu,\mathbf{\Sigma}}$ is the pdf of the multivariate t distribution:

$$f_{\nu,\mu,\mathbf{\Sigma}}(x) = \frac{1}{|\Sigma|} g_{N,\nu}((x-\mu)'\Sigma^{-1}(x-\mu)) \tag{30}$$

where $g_{N,\nu}$ is the generator

$$g_{N,\nu}(z) = \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{N}{2}}}(1+\frac{z}{\nu})^{-\frac{\nu+N}{2}} \tag{31}$$

Further, Meucci (2008) [3] imposes structure on the correlation matrix, because in many financial applications the number of simulations T is many times larger than number of risk factors N. The routine above is supplemented by several steps:

Step 2b. Perform the principal component decomposition of the estimated correlation:

$$\hat{\Sigma}_{(u+1)} = \hat{E}\hat{\Lambda}\hat{E}' \tag{32}$$

Step 2c. Redefine the eigenvalues by setting the trailing ones to equal value $\underline{\lambda} > 0$, assuming isotropy:

$$\hat{\Lambda} = diag(\hat{\lambda_1}, \ldots, \hat{\lambda_K}, \underline{\lambda}, \ldots, \underline{\lambda}) \tag{33}$$

where

$$\underline{\lambda} = \frac{1}{N-K+1}\sum_{k=K+1}^{N} \hat{\lambda_k} \tag{34}$$

Step2d. Recompose the scatter matrix:

$$\hat{\Sigma}_{(u+1)} = \hat{E}\hat{\Lambda}\hat{E}' \tag{35}$$

4 Subjective views

The optimal portfolio composition depends greatly on the values of the input parameters. In general, it is a well known issue in asset allocation that small changes in inputs can lead to completely different optimal weights. There are two main streams of research dedicated to this issue: (a) robust optimization and (b) methods involving subjective views of analysts. In the latter family the first widely famous and still number one in popularity with practitioners is the Black-Litterman model (1992). In this section I explore how subjective view techniques may benefit to loan portfolio optimization. I start with briefly explaining Black-Litterman and why it does not apply, move through the Copula Opinion Pooling which applies but is too weak, and finally describe in detail the Entropy Pooling.

4.1 Black-Litterman

I describe the Black-Litterman model following Walters (2009) [8].

The Black-Litterman model makes two significant contributions to the problem of asset allocation. First, it provides an intuitive prior, the CAPM equilibrium market portfolio, as a starting point for the estimation of asset returns. Second, it shows a clear way to specify investor's views and to blend investor's views with the market implied prior information. The investor's views can span partial, arbitrary or overlapping set of assets. The model estimates expected excess returns and covariances which can be used as an input to an optimizer. When used as a part of an asset allocation process, the Black-Litterman model leads to more stable and more diversified portfolios than plain mean-variance optimization. I seek the same advantages for loan portfolio optimization.

The Black-Litterman model has several representations and a few more interpretations. I start with the most usual and then proceed to the one used by Meucci (2006) [22].

The reference model for Black-Litterman expected return is

$$E(r) \sim N(\pi, \Sigma_r) \tag{36}$$

where r are the returns, π is the mean of the mean return (μ, where $r \sim N(\mu, \Sigma)$), which itself is a random variable. $\mu \sim N(\pi, \Sigma_\pi)$. $\pi = \mu + \epsilon$, $\epsilon \sim N(0, \Sigma_\pi)$.

The formula to calculate Σ_r is

$$\Sigma_r = \Sigma - \Sigma_\pi \tag{37}$$

We need Σ_π. Black and Litterman (1992) made the simplifying assumption that the structure of the covariance matrix is proportional to the covariance of returns Σ:

$$\Sigma_\pi = \tau\Sigma \tag{38}$$

Then, the prior distribution is

$$P(returns) \sim N(\Pi, \tau\Sigma) \tag{39}$$

And Π here are the CAPM excess equilibrium returns (details of derivation can be found in Walters(2009) [8]). I also discuss the starting point of returns in more detail in section 5.

The investor's views are defined as a conditional distribution. First, we require that the views are independent - it makes the views covariance matrix diagonal. Second, we require the views to be fully invested - weights summing up to 1 for absolute view and 0 for relative view. We represent the views by 3 matrices:

1. P, a $k \times n$ matrix with the weights of each asset within each view.
2. Q, a $k \times 1$ matrix of returns of each view.
3. Ω, a $k \times k$ matrix of view covariances.

Given this specification of the views, we formulate the conditional distribution mean and variance in view space:

$$P(views|returns) \sim N(Q, \Omega) \tag{40}$$

and in asset space

$$P(views|returns) \sim N(P^{-1}Q, [P^T\Omega^{-1}P]^{-1}) \tag{41}$$

With the help of an intervention from Theil's mixed estimation theory (1974) or Bayes theory (1763) we obtain that

$$P(returns|views) \sim N([(\tau\Sigma)^{-1}\Pi + P^T\Sigma^{-1}Q][(\tau\Sigma)^{-1}\Pi + P^T\Sigma^{-1}P]^{-1}, [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1}) \tag{42}$$

It is up to investor to calculate Ω, and it is here that various interpretations start to differ. He and Litterman (2000) sets $\Omega = diag(P(\tau\Sigma)P^T)$. Idzorek (2004) uses confidence of the views. Meucci(2006) [22] skips the diagonalization and sets $\Omega = \frac{1}{c}P\Sigma P^T)$, c representing overall confidence in the views, and the obvious choice for the relationship between c and τ is $c = \tau^{-1}$.

4.2 Beyond Black-Litterman: Copula Opinion Pooling

With this specification of τ and Ω, Black and Litterman's posterior can be a little simplified, namely τ gets eliminated due to proportionality, but in essence stays as above. Meucci (2006) [2] gives an alternative method of mixing the prior with the conditional views distribution:

$$\tilde{V} \sim \tilde{f}^c_{\mathbf{X}} = (1-c) f_{\mathbf{X}} + c\tilde{f}_{\mathbf{X}} \tag{43}$$

This idea is best revealed in the picture of cdf's (Figure 4).

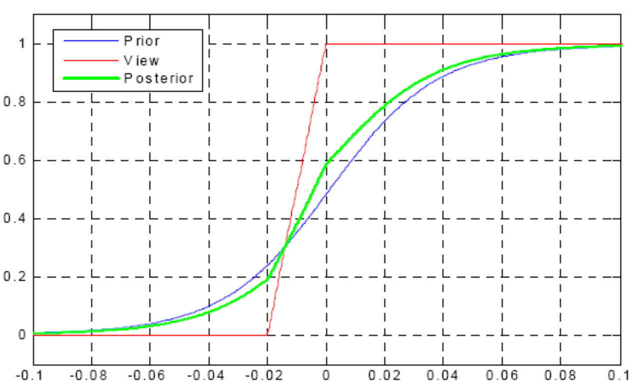


Figure 4: Copula Opinion Pooling: cdf of prior, view, and posterior (source: Meucci (2006) [2])

This mixing of cdf's is quite intuitive, at least more so than Black-Litterman, and this is the essence of Meucci's Copula Opinion Pooling method. If the conditional distribution of the views is normal too, resulting posterior is very little different from Black-Litterman posterior.

But besides intuitiveness, the other advantage of the method is that it does not require neither views nor prior distributions to be normal. Similar to the Black-Litterman retaining the covariance between prior and posterior return distribution (with changes, dependent on τ), COP retains the codependence through copula - this is the reason why a copula is in the name.

COP must implemented numerically. We start with a prior market panel $\mathbf{M}$, multiply it by the views, extract the empirical copula, get the empirical cdfs, combine with view cdfs to obtain posterior cdfs, and finally construct a posterior market panel $\tilde{\mathbf{M}}$ from the empirical copula and the posterior cdfs.

Having two panels $\mathbf{M}$ and $\tilde{\mathbf{M}}$ we can easily compare the parameters of the distributions and observe the effects of the views. The covariance matrix between prior and posterior changes in different ways, comparing to the Black-Litterman model. For the assets where no views are expressed, covariances remain unchanged. When a view is expressed, if the view distribution is tighter around its mean than the prior, the posterior covariances are lower. Otherwise, higher. The Black-Litterman model in the same situation changes the whole matrix, whether or not the view was expressed. This aspect adds to COP's intuitiveness.

Now that we see that the normality is not required, we could try applying COP for the default distribution. And it works indeed, only, unfortunately, reveals very little power. The mixing of continuous distributions is not straightforward. It is only possible to implement analytically if both distributions are normal, as demonstrated by Black and Litterman. If the distributions are not normal, numerical methods such as Copula Opinion Pooling are employed. But mixing of discrete distributions, like default distributions, does not require any effort. Say the PD (the mean of the prior) is 2%, and we have the view that the PD will be 5% instead, with confidence 20%. We immediately calculate that the PD of the posterior is 2.6% (from the mixing formula), hence the simulation from the posterior is not at all harder than the simulation from the prior. Figure 5 shows the mixing of discrete default distributions graphically.

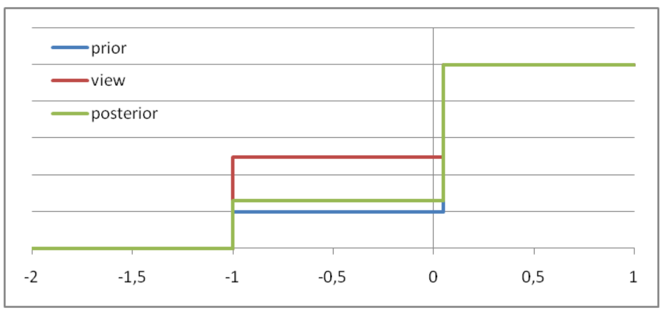


Figure 5: Blending of discrete default distribution prior and the view into the posterior

4.3 Entropy Pooling

Much more interesting task would be to specify views on the dependence between defaults, and we can do that because views on nonlinear properties of the distribution is exactly where Meucci ventured next.

Meucci (2008) [1] fully generalizes the COP and other techniques of scenario analysis. The assumption of normality of the prior distribution (risk factors) or the views is no longer present. Also, views can be formulated about any aspect of the prior, not only means: we may have an opinion or restriction on ranking, volatility, correlation, other features of codependence and so on. The input is the simulation of the market: $J \times K$ matrix of J simulations and K risk factors, and fully general views, or stress tests, on the market, expressed as constraints. The output is a distribution, which we call "posterior", expressed as a vector of probabilities, together with the input market matrix. The posterior can be used for risk management and portfolio optimization purposes.

The idea of the method is to interpret the views as statements that distort the prior distribution by changing the probabilities of relevant market outcome scenarios. For example, the returns of the portfolio according to formula (14) change if p_j change. The CVaR of the portfolio changes with respect to the probabilities as per formula (9). If we have a simple average over all scenarios in the formula, we can always rewrite it as a probability weighted average, where the probabilities are equal (=1/number of scenarios). The direction of the distortion with respect to the views is such that the least possible amount of spurious structure is induced. The natural index of structure in a distribution is the entropy. Therefore we define the posterior as such a distribution which minimizes the entropy relative to the prior.

Entropy Pooling works with non-normal markets, making it useful for a default distribution. The method handles views on non-linear combinations of risk factors which affect returns directly or indirectly through codependence, also making it particularly useful for default distribution. Specifically we can investigate what happens when the dependence between defaults increases, throughout the whole distribution or just the low end of it. We will see how increased default tendency to occur together affects banks' profit and loss.

Another property of Entropy Pooling is the reusing of just one market simulation to simulate both prior and posterior. Probabilities change, but both sets of probabilities affect the same simulated market. Thus the most complex securities can be handled without costly repricing.

Now we describe the method in detail, following Meucci(2008) [1].

We consider a book defined by a N-dimensional vector of risk factors $\mathbf{X}$. This framework is completely general, for instance, in a book of options $\mathbf{X}$ can represent the changes in the underlyings and the implied volatilities. Also, $\mathbf{X}$ can represent a set of risk factors behind a computationally intensive full Monte-Carlo pricing function, such as interest rate values at different monitoring times for mortgage derivatives. In any case, $\mathbf{X}$ can be, but is by no means restricted by, the returns of securities.

We assume a model for the joint distribution of risk factors:

$$\mathbf{X} \sim f_{\mathbf{X}} \tag{44}$$

In the Black-Litterman model this is the prior distribution (there, multivariate normal). More general, this is the model that risk managers use to perform their analysis, and on the other hand, portfolio managers use to optimize portfolios. The reference model can be estimated from historical analysis, or calibrated to current market observables.

The views. In the most general case, the user expresses the views on generic functions of the market $g_1(\mathbf{X}), \ldots, g_k(\mathbf{X})$. These functions constitute a K-dimensional random variable whose joint distribution is implied by the reference model (44):

$$\mathbf{V} \equiv \mathbf{g}(\mathbf{X}) \sim f_{\mathbf{V}} \tag{45}$$

In the Black-Litterman model the functions are linear. As a special case, we can express the views on securities values, in which case $\mathbf{V} = \mathbf{X}$. We find the indicator function to be helpful in dealing with the default distribution: it allows us to translate the joint distribution of the times to default (which we can parametrize as discussed in section 3.2) to the joint distribution of discrete default events within one year (which we express the views on).

$$g(X_k) = \mathbf{1}(X_k < 1) \tag{46}$$

where $\mathbf{1}(z)$ is the indicator function.

The views are statements on the distribution, possibly conflicting with the prior distribution. The most complete possible view specification is the complete distribution, however we usually would choose a certain aspect of it. For instance, a view on the mean return would be expressed as

$$\tilde{m}(V_k) = m_k \tag{47}$$

where m_k is the subjective opinion of the analyst what the average of the risk factor could or should be. Views are constraints on the posterior distribution, expressed through the probabilities of the scenarios.

The posterior. The posterior distribution should satisfy the views without adding additional structure and should be as close as possible to the reference model (44). The relative entropy $\Upsilon(\tilde{f}_{\mathbf{X}}, f_{\mathbf{X}})$ between a generic distribution $\tilde{f}_{\mathbf{X}}$ and a reference distribution $f_{\mathbf{X}}$ is a natural measure of the amount of structure in $\tilde{f}_{\mathbf{X}}$:

$$\Upsilon(\tilde{f}_{\mathbf{X}}, f_{\mathbf{X}}) = \int \tilde{f}_{\mathbf{X}}(\mathbf{x}) \left[ln(\tilde{f}_{\mathbf{X}}) - ln(f_{\mathbf{X}}) \right] d\mathbf{x} \tag{48}$$

Furthermore, it also measures how distorted $\tilde{f}_{\mathbf{X}}$ is with respect to $f_{\mathbf{X}}$. If two distributions coincide, the relative entropy is zero. Therefore, we define the posterior market distribution as

$$\tilde{f}_{\mathbf{X}} \equiv argmin_{f \in V}(\Upsilon(\tilde{f}_{\mathbf{X}}, f_{\mathbf{X}})) \tag{49}$$

Finally, the model can encompass more than one view. We combine views and take into account all analysts' confidence if we express the opinion-pooled posterior distribution the same way we did in Copula Opinion Pooling:

$$\tilde{f}_{\mathbf{X}}^{c} = c_0 f_{\mathbf{X}} + \sum_{s=1}^{S} c_s \tilde{f}_{\mathbf{X}}^{(s)} \tag{50}$$

where c_0 is the belief in the prior and $c_1 \dots c_s, c_i \in [0;1], i = 0, \dots, s$ - confidence levels/significance weights in the views, summing up to 1. In the extreme case where the confidence is total, the full confidence posterior is recovered; on the other hand, in the absence of confidence, the reference risk model is recovered.

Apart for the special case of Black-Litterman (assuming normality of both the prior and the views, and specifying the views on means and covariances of a linear combination of assets), Entropy Pooling cannot be implemented analytically. Meucci (2008) [1] presents a computationally efficient numerical implementation of the method in full generality.

First, we represent the reference distribution of the market $\mathbf{X}$ in terms of a $J \times N$ panel $\mathbf{M}$ of simulations: j-th row of $\mathbf{M}$ represents a scenario of joint realizations of the N variables $\mathbf{X}$, and n-th column represents marginal distribution of n-th risk factor X_n. With the scenarios we associate a vector of probabilities $\mathbf{p}$, which typically, but not necessarily, equal to $1/J$.

The user expresses views on generic non-linear functions on the market, resulting in a $J \times K$ panel $\mathbf{V}$:

$$V_{jk} = g_k(X_1, \dots, X_n) \tag{51}$$

To represent the posterior we use the same scenarios with different probabilities $\tilde{\mathbf{p}}$. Then we write down the views as constraints on the new probabilities:

$$\mathbf{A}\tilde{\mathbf{p}} \leq \mathbf{b} \tag{52}$$

where $\mathbf{A}$ and $\mathbf{b}$ are simple expressions on panel $\mathbf{V}$. For instance, in the view on means (47) $\mathbf{A} = \mathbf{V}'$ and $\mathbf{b} = (m_1, \dots, m_k)$.

Relative entropy becomes its discrete counterpart:

$$\Upsilon(\tilde{\mathbf{p}}, \mathbf{p}) = \sum_{j=1}^{J} |ln(\tilde{p}_j) - ln(p_j)| \tag{53}$$

Therefore, the full-confidence posterior is represented by

$$\tilde{\mathbf{p}} = argmin_{\mathbf{A}\tilde{\mathbf{p}} \leq \mathbf{b}}(\Upsilon(\tilde{\mathbf{p}}, \mathbf{p})) \tag{54}$$

and the opinion-pooled, confidence-weighted posterior as

$$\tilde{\mathbf{p}}_c = (1-c)\mathbf{p} + c\tilde{\mathbf{p}} \tag{55}$$

Proofs and details can be found in Meucci (2008) [1]. Especially relevant topics are:

1. Multiple users, multiple views - how to combine all confidence parameters to form a posterior.
2. Numerical entropy minimization.
3. Examples of views about different aspects of the distribution.

For our purposes three specifications of the views will be relevant. One is the view on mean (47). Here, the Entropy Pooling entropy minimization algorithm returns the set of $\tilde{p}_j$ consistent with the constraint in such a way that it is possible to replicate by tweaking the parameters of the original times to default distribution and simulating it again, because only the marginals are affected. The change in the mean of discrete default events within one year translates directly into the change in the hazard rate, which in turn is a parameter of the marginal time to default distribution. Thus we can trace back what change must be introduced into the distribution of the marginal time to default in order to achieve the desired change in the one year PD, and simulate a new panel of the times to default. Still, the advantage of Entropy Pooling remains the possibility to express all the views at once without potentially costly resimulation.

Next is the view on correlation. We impose

$$\sum_{j=1}^{J} \tilde{p}_j V_{jk} V_{jl} = \hat{m}_k \hat{m}_l + \hat{\sigma}_k \hat{\sigma}_l \tilde{Z}_{kl} \tag{56}$$

where $\hat{m}_k$ is the sample mean and $\hat{\sigma}_k$ is the sample standard deviation of the k-th column of the panel $\mathbf{V}$. Matrix $\tilde{\mathbf{Z}}$ is pre-defined, for example as the homogenous shrinkage

$$\tilde{\mathbf{Z}} = \rho_1 \mathbf{I} + \rho_2 \mathbf{Z} + \rho_3 \mathbf{1}\mathbf{1}' \tag{57}$$

where $\mathbf{Z}$ is the correlation matrix of $\mathbf{V}$, $0 \leq \rho_1, \rho_2, \rho_3 \leq 1$, $\rho_1 + \rho_2 + \rho_3 = 1$, $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is a vector of ones.

The view on correlation between discrete default events, unlike the views on means, cannot be translated straightforwardly into a change in the correlation matrix as a copula parameter of the joint times to default distribution. The $\tilde{p}_j$ are not trivial.

Finally, I write down the specification of the view on tail codependence. First we extract the copula of the panel $\mathbf{V}$: $\tilde{C}_V(u)$. In case $\mathbf{V} = \mathbf{X}$, we most likely have the copula already as a by-product of simulations to obtain $\mathbf{X}$. For example, when $\mathbf{X}$ is the joint distribution of the times to default, we generate it by first simulating univariates from the copula with given parameters and then applying the inverse marginals on the simulated univariates. Thus we know the copula (in the form of a panel of univariates simulated from it) before we have $\mathbf{X}$. If $\mathbf{V}$ does not coincide with $\mathbf{X}$, as in our case where $\mathbf{X}$ is the joint times to default and $\mathbf{V}$ are the joint discrete default events, we extract the empirical copula as in Meucci (2006) [2] from $\mathbf{V}$ by converting the members of the panel $\mathbf{V}$ into their respective normalized ranks in the respective columns. In both cases we obtain a panel $\mathbf{U}$, every row of which represents a simulation from the copula of $\mathbf{V}$.

Still we need to know the copula function $\tilde{C}_V(u)$ itself, not only the simulations. For example, when $\mathbf{X}$ is the joint distribution of the times to default where we have assumed the t-copula for the dependence structure, the copula function is the multivariate t cdf:

$$C_{R,\nu} = t(t^{-1}(\mathbf{u}_1), \dots, t^{-1}(\mathbf{u}_n)) \tag{58}$$

thus for every row of the simulation panel we obtain

$$C^j_{R,\nu} = t(t^{-1}(u^j_1), \dots, t^{-1}(u^j_n)) \tag{59}$$

In case of the empirical copula - when $\mathbf{V} \neq \mathbf{X}$ - we have

$$C^j_{emp}(\mathbf{u}) = [\text{number of rows that lie jointly below } \mathbf{u}]/J \tag{60}$$

when the scenario probabilities are equal, or

$$C^j_{emp}(\mathbf{u}) = \sum_{j=1}^{J} p_j \mathbf{1}[\text{row lies jointly below } \mathbf{u}] \tag{61}$$

where p_j are the probabilities of scenarios and the rows that lie jointly below $\mathbf{u}$ are $\forall \mathbf{v} : v_1 \leq u_1, \dots, v_n \leq u_n$ and $\mathbf{1}[z]$ is the indicator function. If we denote by $\mathbf{I_U}$ the set of scenarios in $\mathbf{U}$ that lie jointly below $\mathbf{u}$, we can write the empirical copula (61) as

$$C^j_{emp}(\mathbf{u}) = \sum_{j \in \mathbf{I_U}} \tilde{p}_j \tag{62}$$

Stress-testing the tail codependence means picking out the rows in the tail end of the copula where bad events happen together and increasing their occurence probabilities:

$$\tilde{C}_V(u) \leq \tilde{C} \tag{63}$$

where $\tilde{C}$ can be set exogenously. A convenient specification yields

$$\tilde{C} = \kappa C_V(u) \tag{64}$$

where $C_V(u)$ is the reference copula induced by $\mathbf{V}$ and κ is a stress factor, for instance, 2.

This translates into

$$\sum_{j \in \mathbf{I_U}} \tilde{p_j} \leq \kappa \sum_{j \in \mathbf{I_U}} p_j \tag{65}$$

Practically, we choose the cutoff level $C_V(u)$, find the corresponding set $\mathbf{I_U}$ as the rows that have smaller $C_V(u)$ than this predefined level, and impose the sum of the posterior probabilities of the scenarios in $\mathbf{I_U}$ to be, say, two times higher than the sum of the prior probabilities of the scenarios in $\mathbf{I_U}$.

Indeed, this procedure is not different from simply including the bad default scenarios two times in the simulation panel. The entropy minimization algorithm with the constraint (65) on $\tilde{p_j}$ returns a set of $\tilde{p_j}$ exactly like this: for the bad scenarios all $\tilde{p_j}$ are equal and the value is κp_j, and for the rest of the scenarios, $\tilde{p_j}$ are equal too and scaled so that the sum of all $\tilde{p_j}$ is 1. But again, the advantage of Entropy Pooling remains the possibility to express all the views at once with assigned confidences within a common framework.

5 Market quilibrium returns and inverse optimization

Although I have mentioned in the beginning of the section 4 that the Black-Litterman model brought two significant contributions to asset allocation theory, the entire section concentrated on the second contribution, the views. This section will try to explore the first contribution, the intuitive prior, in the context of the mean-CVaR optimization instead of the mean-variance optimization.

Here we start with the Markowitz mean-variance framework and derive the equations for 'reverse optimization' starting from the quadratic utility function.

$$U = w^T \Pi - \frac{\lambda}{2} w^T \Sigma w \tag{66}$$

where U is the investors utility which is the function we maximize during portfolio optimization. w is the vector of weights, Π - excess returns, Σ - the covariance matrix of the assets and λ - the risk aversion parameter of the market. U is a concave function thus it has single global maximum. We find it by taking the first derivative by w of U and setting it to 0:

$$\frac{dU}{dw} = \Pi - \lambda \Sigma w = 0 \tag{67}$$

Solving for the excess returns Π we get

$$\Pi = \lambda \Sigma w \tag{68}$$

If we feed the market capitalization weights w_{cap} into (68) and assume that the market is in equilibrium, we obtain equilibrium excess returns. Solving back for weights:

$$w = (\lambda \Sigma)^{-1} \Pi \tag{69}$$

If we use historical equity returns in (69) rather than the equilibrium excess returns, the results will be very sensitive to the changes in Π. With the Black-Litterman model, the weight vector is less sensitive to the reverse optimized Π vector. The stability of the optimization process is one of the strengths of the model.

If we change the risk measure from variance to CVaR, we can try to recover the returns implied by market equilibrium in return-CVaR sense. Or, could mean-variance optimum implied market equilibrium returns still be used as a starting point? Most likely not, as there is some research pointing out that mean-variance (M-V) and mean-CVaR optimal portfolios can be quite different. Fabozzi (2007) [20] notes that in return/CVaR coordinates, as expected, then mean-CVaR efficient frontier lies above the mean-variance efficient frontier. With the inclusion of derivative assets such as options and credit derivatives, the frontiers are no longer close to each other. Agarwal and Naik (2004) [19] find that "tail risk is significantly underestimated using the M-V approach, the range of underestimation being 12% to 54% for confidence level ranging from 90% to 99%". The reason for underestimation is the high non-normality of return distributions of analyzed portfolios. When distribution of returns is normal, mean-variance and mean-CVaR efficient frontiers actually match (Fabozzi (2007) [20]).

Since the tractable analytical expressions of M-V optimization are convenient, a lot of research goes in the direction of dealing with non-normality in the form of including higher moments of the distribution into optimization objective. In particular, coskewness and cokurtosis, to account for skewness of distributions, which is usually not favoured by agents.

Another stream of research expands the family of returns distribution having analytical expressions to stable distributions. Fabozzi et.al. (2007) [12] presents the following formula for the CVaR equilibrium returns under the assumption that the returns distribution is stable:

$$\Pi = \frac{\lambda}{2}(\text{CVaR}_{\beta}\frac{\Sigma w}{\sqrt{w^T \Sigma w}} - E(r)) \tag{70}$$

where $E(r)$ is the expected return. λ is the market risk aversion parameter. We calculate λ as follows:

$$\lambda = \frac{w^T r}{\sqrt{w^T \Sigma w}} \tag{71}$$

The α-stable distributions describe a general class of distribution functions. The α-stable distribution is identified by four parameters: the index of stability $\alpha \in (0, 2]$ which is the parameter of the kurtosis, the skewness parameter $\beta \in [-1, 1]$; $\mu \in \mathbf{R}$, and $\gamma \in \mathbf{R}^+$ which are, respectively, the location and the dispersion parameter. The stable distribution is normal, when $\alpha = 2$ and it is leptokurtotic when $\alpha < 2$. A positive skewness ($\beta > 0$) identifies distributions with right fat tails, while a negative skewness ($\beta < 0$) typically characterizes distributions with left fat tails. Therefore, the stable density functions synthesize the distributional forms empirically observed in the real data. Unfortunately the density of stable distributions cannot be express in closed form. Thus, in order to value the density function, it is necessary to invert the characteristic function.

It is hard to say whether the assumption of a stable distribution is still too strong for our loan returns distribution, and more likely than not it is too strong. It remains to be answered what exactly is the effect of the distribution assumption; perhaps this result could still be partially utilized. Applying the formula (70) regardless of the possibly wrong assumption produces strange negative returns Π. The returns are still strange if in the formula (70) the expected returns are not deducted. Table 4 summarizes the results on the 3 obligor example from the section 2.4.

w_{opt}	r	λ	Π	Π^*	r_{opt}
0.4303	9.00%	5.5	22.81%	5.88%	9.00%
0.4302	5.00%		8.88%	-2.13%	5.00%
0.1396	4.00%		3.72%	-5.05%	5.00%

Table 4: Market equilibrium returns according to (70). w_{opt} are the true optimum weights attained with returns r. λ as in (71), Π are returns from the formula (70) and Π^* are returns from the formula (70) if $E(r)$ are not deducted. r_{opt} is the numerical result from the appendix B.

As w_{opt} is the true optimum attained with returns r, we would expect the Π to be close to the original r, but this is not the case.

The final option is to use numerical methods, and this is the way I go. Instead of the analytical expression (68) implied by the mean-variance equilibrium (which is the same thing as to say that our utility function is quadratic, or that the returns distribution is normal, or that the higher moments are irrelevant in the sense that odd moments are 0 and even moments are subordinated by a second moment), I try to obtain Π, knowing w, assuming that the market is in the mean-CVaR equilibrium. I do not assume normality or a certain shape of the utility function.

Thus we need to investigate what sort of problem is to find returns, given which a known vector of weights represents the market equilibrium. The weights should be optimal across all other *weights*. It is tempting to simply change the decision variables and exchange the objective function with the constraints in the original problem: instead of finding the optimal weights across all other weights when returns are known, look for the optimal returns across all returns when the weights are known. Obviously this transposition is not answering the core question: the optimal returns across all returns are definitely the 100% returns, or whatever the maximum limit for the returns we set, because objective would be to maximize the return. Even if we would write down the alternative formulation and minimize risk, given fixed level of returns, the optimal returns would be the maximum allowed: there is not much freedom to change the risk when weights are fixed. The problem of finding the objective function with which the given weights vector is the optimum in weights space belongs to a class of problems referred to as inverse optimization.

The chapter "Inverse Optimization" in the appendix B tackles this issue. We see that the problem is not insurmountable but usable results remain beyond the reach of this paper. Here is the list of problematic issues.

1. It is necessary to separate risk from returns in the problem definition. If we put returns in the optimization objective, as in (13), defined by the market outcome scenarios $\mathbf{M}$, we cannot have return impacting also the optimization constraints, i.e. fixed risk. We can only fix the risk in the form of "pure loss CVaR": a matrix $\mathbf{M}^p$, where in case of a default event we find -1 and in case of no default we find 0 or another fixed value instead of a loan margin.

2. We start with given returns r_0 and chosen starting weights (say, market capitalization weights) x^0. If the point is infeasible with respect to (13), we find a nearest feasible x^1. Then we solve the dual problem to find r_{opt}, the market equilibrium returns. The goodness (acceptance) criteria for r_{opt} is whether (13) with r_{opt} instead of r_0 optimizes to x^*, which is sufficiently close to $x^0(x^1)$. And there is clearly some problem with my results as with reasonably sized problem the criteria does not always hold well, although theoretically it should. The problem is linear and on small data sizes the theory seems to work well. I am not able to find the error.

3. In addition to this problem, there are two interesting aspects of the returns sensitivity in this area. First, the sensitivity of optimal returns to the initial returns provided. We set the distance between optimum and provided returns to be minimal. This restriction places a very high importance on what returns we provide as a benchmark, because the returned returns will not deviate much. They will deviate as little as possible, given the restraint that weights will still be optimal. Perhaps, if returns provided are historical, this behaviour is desirable.

4. Second, we move the dual solution returns by an ϵ in order to get into the inside of the area where the primal problem has unique optimum. It would be interesting to know what exactly are the limits of the movement of r_{opt} so that the same optimum is returned. In a sense an answer to this problem would help answering the first as we would see how far r_{opt} can deviate from r_0.

The first issue is methodological and not critical. All three remaining issues in the inverse optimization process carry more technical character and require more expertise in linear programming than is relevant to the overall scope of the paper, hence unfortunately I will defer them to the further research.

We have shown that given sensible initial returns, we can find the returns as close as possible to the initial, with which the ”market capitalization” weights are optimal in the mean-CVaR sense, as long as the market capitalization weights satisfy the CVaR constraint initially. If they do not satisfy the CVaR constraint, we move the weights towards the CVaR satisfaction. For every pair of initial weights and returns, then, we discover a closest pair of weights and returns which make mean-CVaR equilibrium.

However, it is feasible and relevant to start analyzing what could be the effects of the subjective views on optimization even if we take as prior given returns instead of the market equilibrium returns. I am going to investigate the significance of the starting point of returns too, in order to see whether the optimal portfolio in the mean-CVaR framework is as sensitive to changes in returns as the mean-variance optimal portfolio.

6 Data

I illustrate the use of subjective views in loan portfolio optimization. I first optimize CVaR, then stress test the distribution in 3 ways, reoptimize CVaR after each stress test and compare the differences. As inputs I take the data that banks usually have collected for various other purposes. It means that the methods described in this thesis do not require a specific data collection or estimation effort and therefore are directly applicable in practice.

1. Probabilities of default(PD). We assume that every obligor has its one year PD assigned, either through rating, or by internal models. The Basel II accord spec-

ifies that the internally specified PD should be average over all phases of economy.

2. Loss given default (LGD), equal to 100% minus the recovery rate. Internal assessment or, for example, Basel IRBF specified, which then are downturn LGD.

3. Historical equity returns. We will use historical equity returns to determine the default correlation matrix (meaning correlation between the times to default), as described in section 3.2. In practical situations a time series of returns of each individual obligor would usually be condensed into a time series of returns on risk factors, most likely the average returns of sectors of economy, due to data unavailability and practical difficulties of working with too big datasets.

4. Loan margins.

We have a portfolio of 10 randomly selected large multinational corporate customers with known credit ratings and historical equity returns. Thus the obligors and their risk profiles are real but the lender, the portfolio and the loan margins are fictional. I simulate the loan margins to be roughly inversely proportional to PDs, ensuring a good trade-off between risk and return.

The source of the ratings is Standard & Poors(http://www.standardandpoors.com), date: August 20, 2009. Associated PDs are of year 2008, taken from S&P report, publicly available on the web on http://www.standardandpoors.com. Note that 2008 was a crisis year with several large highly rated companies defaulting, so PDs of 2008 are greater than average. It also happened that a lower rating class (BBB+) had less defaults than a higher rating class (A-). The Loss Given Default for every obligor is 100%, for simplicity. Table 5 displays the obligors and their ratings.

Nr	Name	DScode	Country	Industry	S&P
1	BHP BILLITON	912398	United Kingdom	Basic Materials	A+
2	BRITISH AMERICAN TOBACCO	905478	United Kingdom	Consumer Goods	BBB+
3	DIAGEO	275791	United Kingdom	Consumer Goods	A-
4	ENI	923306	Italy	Oil & Gas	AA-
5	HSBC HDG.	531865	United Kingdom	Financials	AA
6	IBERDROLA	929420	Spain	Utilities	A-
7	NESTLE 'R'	870533	Switzerland	Consumer Goods	AA
8	ROCHE HOLDINGS GSH.	866056	Switzerland	Health Care	AA-
9	TELEFONICA	923374	Spain	Telecommunications	A-
10	TESCO	929534	United Kingdom	Consumer Services	A-

Table 5: Obligors and ratings

We have the time series of weekly log equity returns of this portfolio for the period 26 October 2001 - 17 October 2008 (364 observations). We assume a joint multivariate t distribution for this dataset and extract the t-copula parameters: degrees of freedom and correlation matrix, as outlined in section 3.2. The estimated correlation matrix is given in table 6.

1.0000	0.1056	0.2087	0.4258	0.3415	0.3288	0.2393	0.1800	0.3414	0.2031
0.1056	1.0000	0.3974	0.1893	0.2968	0.2900	0.3257	0.1938	0.2109	0.3629
0.2087	0.3974	1.0000	0.2744	0.3548	0.3666	0.3555	0.1984	0.2628	0.4028
0.4258	0.1893	0.2744	1.0000	0.3962	0.3439	0.3248	0.2654	0.3875	0.2684
0.3415	0.2968	0.3548	0.3962	1.0000	0.3498	0.4073	0.3382	0.4067	0.3443
0.3288	0.2900	0.3666	0.3439	0.3498	1.0000	0.3010	0.1634	0.2849	0.3354
0.2393	0.3257	0.3555	0.3248	0.4073	0.3010	1.0000	0.3579	0.3755	0.3460
0.1800	0.1938	0.1984	0.2654	0.3382	0.1634	0.3579	1.0000	0.3598	0.2145
0.3414	0.2109	0.2628	0.3875	0.4067	0.2849	0.3755	0.3598	1.0000	0.2677
0.2031	0.3629	0.4028	0.2684	0.3443	0.3354	0.3460	0.2145	0.2677	1.0000

Table 6: The original correlation matrix (t-copula parameter **C**) extracted from the historical equity returns time series

The t-copula parameter degrees of freedom ν is estimated as 10.

I simulate 100 000 time to default scenarios using the t-copula with parameters **C** and ν and exponential marginals for the times to default with hazard rate parameter h, derived from the PDs, according to this procedure:

1. Get multivariate t variable with df ν, mean 0 and correlation C:

   (a) find $\mathbf{A} = Cholesky(\mathbf{C})$

   (b) generate n standard normals $z_1, \dots, z_n \sim N(0,1)$

   (c) $y = \mathbf{A}z$

   (d) generate s, independent of z: $s \sim \chi^2_\nu$

   (e) $x = \frac{\sqrt{\nu}}{\sqrt{s}} y$

   (by multiplying by standard deviation and adding means we would get the original stock returns back - if assumption on the original distribution holds. It is a good check, especially for covariances)

2. Transform to univariates: $u = t_\nu(x)$, where t_ν is cdf of t distribution with ν df. The resultant multivariate vector u is what is called a simulation from a copula. It retains the dependence structure from the equity returns.

3. Transform u to exponential marginals of times to default: $T = F^{-1}(u)$, where F is the exponential distribution function of time to default:

$$F(t) = 1 - e^{-ht} \tag{72}$$

($\rightarrow T = -ln(1-u)/h$) assuming constant hazard rates h. Hazard rate itself is calculated from PD, as one year PD is $F(1)$: $h = -ln(1 - \text{PD})$.

Default events within one year are the cases when the time to default is less than 1. We check afterwards whether the average of default events matches back to the original input PDs.

The table 7 summarizes the information about the risk/return profile of the obligors.

Nr	PD	Margin	h($\times 10^{-2}$)	$\bar{T}$	$\hat{\text{PD}}$	$\bar{r}$	Profile
1	0.31%	5.96%	0.3105	321.5	0.309%	5.63%	L
2	0.18%	5.16%	0.1802	555.2	0.158%	5.00%	L
3	0.58%	6.90%	0.5817	171.9	0.554%	6.30%	H
4	0.40%	5.03%	0.4008	248.5	0.388%	4.62%	L
5	0.43%	6.31%	0.4309	231.9	0.432%	5.85%	L
6	0.58%	7.57%	0.5817	172.2	0.539%	6.99%	H
7	0.43%	7.87%	0.4309	231.6	0.428%	7.41%	H
8	0.40%	5.70%	0.4008	249.6	0.428%	5.24%	L
9	0.58%	7.05%	0.5817	171.2	0.595%	6.41%	H
10	0.58%	7.73%	0.5817	172.0	0.557%	7.13%	H

Table 7: Risk and return data of the obligors. The first column is the PD corresponding to the rating. The loan margins are generated inversely proportional to PD (high risk-high return) plus a random factor. $\bar{T}$ is the average time to default. $\hat{\text{PD}}$ is the empirical one year PD, as the average of the default events (where $T < 1$). $\bar{r}$ is the risk adjusted loan returns.

Based on this information obligors fall into categories: High(H) or Low(L) risk profile. Obligor 7 stands out as having the most favourable risk/return ratio - the return is highest of all obligors, and the PD is lower than some; and obligor 4 as the most unfavourable.

The panel of simulations of the times to default is translated to the panel of discrete default events within one year by applying the indicator function (1 if the time to default is less than 1; 0 otherwise). The empirical correlation matrix of the panel of discrete default events within one year is given in the table 8.

The mean time to default in the stressed posterior distribution is calculated according to the formula

$$\bar{T}_i = \sum_{j=1}^{J} \tilde{p}_j T_{ij} \tag{73}$$

where $\tilde{p}_j$ are the posterior probabilities and T_{ij} are the elements of the $\mathbf{T}$ - $J \times K$ panel of the simulations of the times to default.

1.000	0.034	0.047	0.121	0.084	0.087	0.035	0.046	0.115	0.054
0.034	1.000	0.112	0.066	0.090	0.073	0.067	0.063	0.062	0.065
0.047	0.112	1.000	0.091	0.112	0.088	0.100	0.053	0.064	0.121
0.121	0.066	0.091	1.000	0.119	0.085	0.092	0.067	0.093	0.077
0.084	0.090	0.112	0.119	1.000	0.093	0.131	0.110	0.132	0.102
0.087	0.073	0.088	0.085	0.093	1.000	0.073	0.047	0.096	0.092
0.035	0.067	0.100	0.092	0.131	0.073	1.000	0.127	0.128	0.098
0.046	0.063	0.053	0.067	0.110	0.047	0.127	1.000	0.116	0.073
0.115	0.062	0.064	0.093	0.132	0.096	0.128	0.116	1.000	0.066
0.054	0.065	0.121	0.077	0.102	0.092	0.098	0.073	0.066	1.000

Table 8: The empirical correlation matrix between discrete defaults

The posterior PDs are calculated according to the formula

$$\bar{PD}_i = \sum_{j=1}^{J} \tilde{p}_j D_{ij} \tag{74}$$

where D_{ij} are the elements of $\mathbf{D}$ - $J \times K$ panel of the default events within one year (1 if time to default T is less than 1 and 0 otherwise).

Another statistics that defines the posterior joint default distribution is the count of scenarios where at least x obligors default together. In the prior distribution we can simply count the events, and in posterior the number of defaults is a posterior probability weighted average of scenarios where default number is greater than x (of course one can always also express a simple count as an average weighted by equal probabilities):

$$\bar{\mathbf{D}} = J \sum_{j=1}^{J} \tilde{p}_j \mathbf{1}[\sum_{i=1}^{K} D_{ij} \geq x] \tag{75}$$

where $\mathbf{1}[z]$ is the indicator function.

7 Results

7.1 Mean-CVaR optimal portfolio before stress testing

The default event panel (multiplied by minus LGD) is the loss part of the loan return distribution. The profit part of the loan return distribution consists of the loan margins when default does not occur - i.e. returns before they are risk adjusted. The complete market outcome matrix to be used for CVaR optimization looks in essence like this:

$$\begin{bmatrix} 0.09 & -1 & 0.04 & \dots \\ 0.09 & 0.05 & 0.04 & \dots \\ 0.09 & 0.05 & -1 & \dots \\ \dots & \dots & \dots & \dots \\ 0.09 & 0.05 & 0.04 & \dots \\ -1 & -1 & -1 & \dots \end{bmatrix} \tag{76}$$

We start with the equal weights (=1/10) portfolio. The mean risk adjusted return of this portfolio is 6.06%.

Typically, the 99% quantile of VaR and CVaR is used and reported, based on habit and convenience. In our 100 000 simulation panel 1% of the worst outcomes is 1 000 scenarios, and with these low default rates as we have (mean: 0.447%), 1 000 cases would contain a very significant part of scenarios where *anybody* defaults. If all the cases where anybody defaults would fall under the CVaR quantile, not much sense is there in stressing the default correlation. If defaults are even more clumped together, they would all still fall under the threshold, and only the marginal default rate defines the weight.

Table 9 pictures how many scenarios have at least x obligors defaulting together.

≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7	≥ 8
3252	740	245	94	38	14	4	1

Table 9: Joint default events: how many scenarios have at least x obligors defaulting together.

It makes better sense to work with the 99.9% quantile than the 99%, then. The 99.9% quantile is also the stress level of Basel II accord formulas for Risk Weighted Assets in Internal Ratings Based approach.

We set the desired returns to the same level as the return of the equal weights portfolio (6.06%) and optimize 99.9% CVaR (with 99% and 95% CVaR optimum weights given for comparison) as setup in (11). Table 10 presents optimal portfolios.

Q	1	2	3	4	5	6	7	8	9	10
99.9%	0.1703	0.1716	0.0507	0.0516	0.0509	0.1174	0.1171	0.1025	0.0506	0.1173
99%	0.1061	0.2641	0.1052	0	0	0.1045	0.1042	0.1064	0.1050	0.1044
95%	0.0608	0.3513	0.0603	0	0.0606	0.0599	0.2262	0.0609	0.0602	0.0598

Table 10: Mean-CVaR optimal portfolios before the stress testing: 99.9%, 99% and 95% quantiles.

99.9% CVaR optimum portfolio looks normal, favoring obligors with lower risk (1 and 2). 95% CVaR, with all defaults under the threshold, chooses obligor 2 predominantly as the most invested, because it has the lowest PD overall. Obligor 7, with largest return overall, is also favoured.

Next table 11 compares the CVaR and VaR of the equal weights and optimal portfolios.

Q	VaR eq	CVaR eq	VaR opt	CVaR opt
99.9%	0.2572	0.4158	0.2961	0.3889
99%	0.0425	0.1629	0.0476	0.1575
95%	-0.0653	0.0283	-0.0644	0.0121

Table 11: CVaR and VaR of the equal weights and optimal portfolios.

We start with the panel of simulation of times to default and apply Entropy Pooling procedure to obtain posterior probabilities. We express the views on the discrete default events within one year rather than the times to default, using the function g as in (46) for the transition. Then we take the same market scenarios as prior and optimize CVaR using new probabilities. We investigate three scenarios:

1. Increased mean (i.e, PD)
2. Increased default correlation throughout the whole distribution
3. Increased lower tail codependence

The views are already relative to the original distribution, thus we will apply 100% confidence.

7.2 Stress testing: means

As the mean stress testing exercise we impose a view on PDs of 5 obligors. The view states that PDs of obligors 1,2,3,5 and 10 is actually higher by 0.1 percentage point than S&P rating class average (Table 12).

PD	1	2	3	4	5	6	7	8	9	10
Original	0.31	0.18	0.58	0.40	0.43	0.58	0.43	0.40	0.58	0.58
Stressed	0.41	0.28	0.68	0.40	0.53	0.58	0.43	0.40	0.58	0.68

Table 12: Stressed PDs

Although this kind of stress-testing does not require sophisticated methods of distribution blending, as discussed before, entropy pooling can accommodate it as well.

The stressed scenario probabilities ($\tilde{p_j}$) reflect the stress test. When no default occurs (100 000 - 3 252 = 96 478 cases) the probability is practically unchanged (9.97×10^{-6}).

The scenarios where others default but obligors 1,2,3,5 and 10 do not default get lower $\tilde{p_j}$'s, and the scenarios where obligors 1,2,3,5 and 10 default at least partly together, get higher $\tilde{p_j}$'s.

Table 13 shows the average times to default and average one year PD.

Obligor	1	2	3	4	5	6	7	8	9	10
T	321.2	554.3	171.6	248.2	231.5	172.0	231.3	249.5	171.1	171.7
PD	0.409%	0.258%	0.654%	0.388%	0.532%	0.539%	0.428%	0.428%	0.595%	0.657%

Table 13: Time to default (T) and one year PD in the mean stress test

The times to default change little, compared to the table 7, but PDs change accordingly to the specification of the stress test. Table 14 pictures how many scenarios have at least x obligors defaulting together in this stress test.

≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7	≥ 8	≥ 9	Total
3507	867	306	124	54	22	7	2	0	4888

Table 14: Joint default events in the mean stress test: how many scenarios have at least x obligors defaulting together.

The mean stress test increases overall default count, as expected.

The 99.9%-CVaR optimal portfolio is presented in the table 15.

1	2	3	4	5	6	7	8	9	10
0.1692	0.1513	0.0466	0.0758	0.0000	0.1203	0.1200	0.1224	0.0743	0.1201

Table 15: Mean-CVaR optimal portfolio after the mean stress test

For the discussion of the effects of the stress tests on the optimal portfolio the reader is referred to the section 7.5.

7.3 Stress testing: correlation

Here I take the correlation matrix between discrete default events (table 8) and use the homogenous shrinkage (56) with parameters in (57): $\rho_1 = 0.3, \quad \rho_2 = 0.5 \quad, \rho_3 = 0.2$. The covariance matrix this way increase element-by-element by roughly the same amount. The correlation matrix shown in table 16.

The stressed scenario probabilities ($\tilde{p_j}$) sensibly reflect the stress test again. When no default occurs (100 000 - 3 252 = 96 478 cases) the probability is practically unchanged (1.01×10^{-5}). The default scenarios get $\tilde{p_j}$'s dependent on the number of obligors defaulting together: the more obligors default together, the higher the $\tilde{p}$, but $\tilde{p_j}$ is smaller than the neutral (1.01×10^{-5}) when there is just 1 default event in the scenario.

1.000	0.217	0.223	0.261	0.242	0.243	0.217	0.223	0.258	0.227
0.217	1.000	0.256	0.233	0.245	0.236	0.233	0.231	0.231	0.232
0.223	0.256	1.000	0.245	0.256	0.244	0.250	0.226	0.232	0.261
0.261	0.233	0.245	1.000	0.259	0.243	0.246	0.234	0.247	0.239
0.242	0.245	0.256	0.259	1.000	0.247	0.266	0.255	0.266	0.251
0.243	0.236	0.244	0.243	0.247	1.000	0.236	0.224	0.248	0.246
0.217	0.233	0.250	0.246	0.266	0.236	1.000	0.264	0.264	0.249
0.223	0.231	0.226	0.234	0.255	0.224	0.264	1.000	0.258	0.237
0.258	0.231	0.232	0.247	0.266	0.248	0.264	0.258	1.000	0.233
0.227	0.232	0.261	0.239	0.251	0.246	0.249	0.237	0.233	1.000

Table 16: Stressed discrete default correlation matrix, correlation stress test

The test favours the cases with many default events over the cases with just one default event.

Table 17 shows the average times to default and average one year PD.

Obligor	1	2	3	4	5	6	7	8	9	10
T	322.5	556.4	172.4	249.3	232.7	172.7	232.4	250.1	171.8	172.5
PD	0.309%	0.158%	0.554%	0.388%	0.432%	0.539%	0.428%	0.428%	0.595%	0.557%

Table 17: Time to default (T) and one year PD in the mean stress test

The times to default again change little, compared to the table 7, and PDs this time remain unchanged too. Table 18 pictures how many scenarios have at least x obligors defaulting together in this stress test.

≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7	≥ 8	≥ 9	Total
2350	793	495	333	223	117	58	19	0	4388

Table 18: Joint default events in the correlation stress test: how many scenarios have at least x obligors defaulting together.

This time the overall default count stays exactly the same as before stress testing (table 9), showing that indeed we do not change the overall default frequency. But we clearly see that the defaults are much more likely to occur together: there are less scenarios with at least one default occuring but more scenarios with multiple default. If one obligor defaults, others are more likely to default too.

The 99.9%-CVaR optimal portfolio is presented in the table 19.

1	2	3	4	5	6	7	8	9	10
0.2551	0.1440	0.1264	0.0566	0.0000	0.1105	0.0551	0.0717	0.0000	0.1806

Table 19: Mean-CVaR optimal portfolio after the correlation stress test

For the discussion of the effects of the stress tests on the optimal portfolio the reader is referred to the section 7.5.

7.4 Stress testing: tail dependence

Here I increase the tendency of default events to happen together. First we extract the empirical copula of the discrete default events panel V (45). The empirical copula function is the proportion of rows lying jointly below the given row, as in (61). Then I take the low end of the copula (to make the low end the bad case, we invert the sign of the default indicator and represent the default by -1): cases where the empirical copula function is less than 10^{-5}. There are 73 such scenarios out of 100 000. Let us increase the dependence between the low end of the distribution by a factor of 4.41. I constrain the sum of probabilities for these 73 bad cases to 4.41 times the sum of the prior probabilities (=$73 * 4.41 \times 10^{-5} = 0.00322$). Basically I set the κ in the formula (65) to 4.41.

The resulting correlation matrix (over all distribution) is quite different from the original empirical one, but not as homogeneously as in the correlation stress test case. This indicates that we have indeed affected only the low end of the distribution, table 20:

1.000	0.076	0.124	0.200	0.167	0.150	0.112	0.107	0.182	0.111
0.076	1.000	0.197	0.161	0.172	0.162	0.141	0.149	0.151	0.105
0.124	0.197	1.000	0.178	0.218	0.156	0.180	0.109	0.129	0.198
0.200	0.161	0.178	1.000	0.234	0.153	0.173	0.137	0.168	0.153
0.167	0.172	0.218	0.234	1.000	0.166	0.230	0.198	0.210	0.184
0.150	0.162	0.156	0.153	0.166	1.000	0.136	0.106	0.171	0.155
0.112	0.141	0.180	0.173	0.230	0.136	1.000	0.203	0.207	0.151
0.107	0.149	0.109	0.137	0.198	0.106	0.203	1.000	0.166	0.128
0.182	0.151	0.129	0.168	0.210	0.171	0.207	0.166	1.000	0.117
0.111	0.105	0.198	0.153	0.184	0.155	0.151	0.128	0.117	1.000

Table 20: Stressed discrete default correlation matrix, tail dependence stress test

The stressed scenario probabilities ($\tilde{p_j}$) in this case are different from the other stress tests - only the 73 "bad" scenarios get a significantly different $\tilde{p_j}$ from the rest. It is equal for all 73 scenarios (=$73 * 4.41 \times 10^{-5}$). The rest 99 927 scenarios all get 9.98×10^{-6}, regardless of whether any defaults occur or not.

Table 21 shows the average times to default and average one year PD.

The times to default again change little, compared to the table 7, but PDs increase all. Table 22 pictures how many scenarios have at least x obligors defaulting together in this stress test.

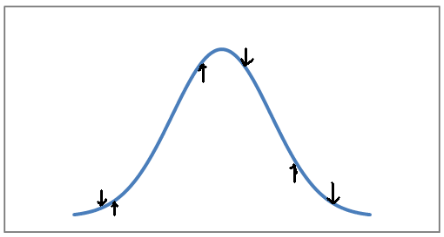
Now we observe the overal default incidence increasing. Figure 6 helps to explain why the correlation and the tail dependence stress tests behave differently.

Obligor	1	2	3	4	5	6	7	8	9	10
T	320.9	554.1	171.5	248.0	231.4	171.9	231.1	249.2	170.9	252.1
PD	0.418%	0.243%	0.686%	0.513%	0.568%	0.664%	0.540%	0.526%	0.716%	0.672%

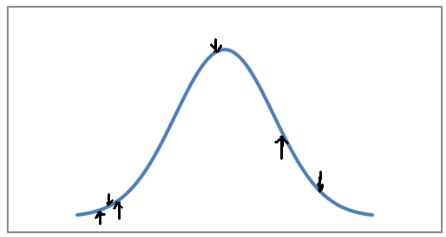
Table 21: Time to default (T) and one year PD in the mean stress test

≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7	≥ 8	≥ 9	Total
3490	990	490	320	170	62	18	4	0	5544

Table 22: Joint default events in the tail dependence stress test: how many scenarios have at least x obligors defaulting together.



(a) Obligor 1



(b) Obligor 2

Figure 6: Schematic illustration of default events: correlation stress test as downwards pointing arrows and tail dependence stress test as upwards pointing arrows

Let's assume a joint distribution of a variable, marginals of which are in the pictures, say, obligor 1 is on the left and obligor 2 on the right (the distribution of the time to default in reality is, of course, exponential; I show as if it was normal only for simplicity). Without the stress testing, the joint tail events are likely to occur with a certain frequency: if obligor 1 defaults, with a certain probability obligor 2 will default too. The downwards pointing arrows display the correlation stress test: if a certain value is obtained by obligor 1, the value of obligor 2 will be close, because the correlation is high. It means that if a tail event occurs in obligor 1, tail event will likely occur in obligor 2 as well, but, if a normal event occurs in obligor 1, tail event is not likely to occur in obligor 2. It means that besides strenghtening the ties between default events, the correlation stress test is strenghtening the relationship between other events as well, driving PD both up and down. The tail dependence stress test, pictured as the upwards pointing arrows, increases the chance of tail event in obligor 2 only in case when tail event occured in obligor 1. If a normal event happened with obligor 1, a tail event still happens with obligor 2 with the same probability as before. So there is no pressing down of PD, only up - and overall PD increases.

The 99.9%-CVaR optimal portfolio is presented in the table 23.

For the discussion of the effects of the stress tests on the optimal portfolio the reader is referred to the section 7.5.

1	2	3	4	5	6	7	8	9	10
0.1699	0.1926	0.0702	0.0500	0.0000	0.0976	0.0973	0.0993	0.0770	0.1461

Table 23: Mean-CVaR optimal portfolio after the tail dependence stress test

7.5 Comparison of optimal portfolios

For the reference, I include the table 24 which displays the default count in all 4 scenarios: original and the 3 stress tests.

Scenario	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7	≥ 8	≥ 9	Total
Original	3252	740	245	94	38	14	4	1	0	4388
Mean	3507	867	306	124	54	22	7	2	0	4888
Correlation	2350	793	495	333	223	117	58	19	0	4388
Tail	3490	990	490	320	170	62	18	4	0	5544

Table 24: Joint default events in 4 scenarios: how many simulations have more than x default occuring together

The ratios of each count of default in each stress test to original are presented in table 25 (NA=not available).

Scenario	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7	≥ 8	≥ 9	Total
Mean	1.08	1.17	1.25	1.31	1.43	1.58	1.75	1.51	NA	1.11
Correlation	0.72	1.07	2.02	3.54	5.86	8.35	14.61	19.48	NA	1.00
Tail	1.07	1.34	2.00	3.40	4.47	4.41	4.41	4.41	NA	1.26

Table 25: Joint default events in 4 scenarios as ratios to the original distribution joint default events

The correlation stress test has the ratio growing with the x: the more default events occur together, the higher the difference between the stressed and original distributions. The ratio does not grow so pronounced in the tail stress test. The number 4.41 in the tail test row represents the ratio of stressed probability of a scenario ($\tilde{p}$) and the original probability, p (= $1/J$). All the scenarios with 6 defaults and more got the same $\tilde{p}$.

The table 26 shows jopint default occurences in the shape of conditional probabilities: how likely additional defaults are to occur, if any default occurs.

Again we see how likely the defaults are to occur together, with correlation stress test standing out.

All 4 scenario's optimal portfolio weights are presented together in table 27 and CVaR, VaR and returns in table 28 - essentially all the portfolios are equal on these three main characteristics.

The figure 7 shows the composition of all 4 optimal portfolios together.

The main observations:

	Joint defaults							
Scenario	1	2	3	4	5	6	7	
Original	0.772	0.152	0.046	0.017	0.007	0.003	0.001	0.000
Mean	0.753	0.160	0.052	0.020	0.009	0.004	0.002	0.000
Correlation	0.663	0.127	0.069	0.047	0.045	0.025	0.017	0.008
Tail	0.716	0.143	0.049	0.043	0.031	0.013	0.004	0.001

Table 26: Joint default events in 4 scenarios as conditional probabilities of exactly x defaults occuring given that at least 1 occurs

Scenario	1	2	3	4	5	6	7	8	9	10
Original	0.1703	0.1716	0.0507	0.0516	0.0509	0.1174	0.1171	0.1025	0.0506	0.1173
Mean	0.1692	0.1513	0.0466	0.0758	0.0000	0.1203	0.1200	0.1224	0.0743	0.1201
Correlation	0.2551	0.1440	0.1264	0.0566	0.0000	0.1105	0.0551	0.0717	0.0000	0.1806
Tail	0.1699	0.1926	0.0702	0.0500	0.0000	0.0976	0.0973	0.0993	0.0770	0.1461

Table 27: Optimal portfolios in 4 scenarios: original and 3 stress tests

Scenario	CVaR	VaR	Return*
Original	0.3889	0.2961	-0.0606
Mean	0.3953	0.3032	-0.0601
Correlation	0.4318	0.3242	-0.0606
Tail	0.3921	0.3025	-0.0594

Table 28: CVaR, VaR and returns of the optimal portfolios in 4 scenarios: original and 3 stress tests

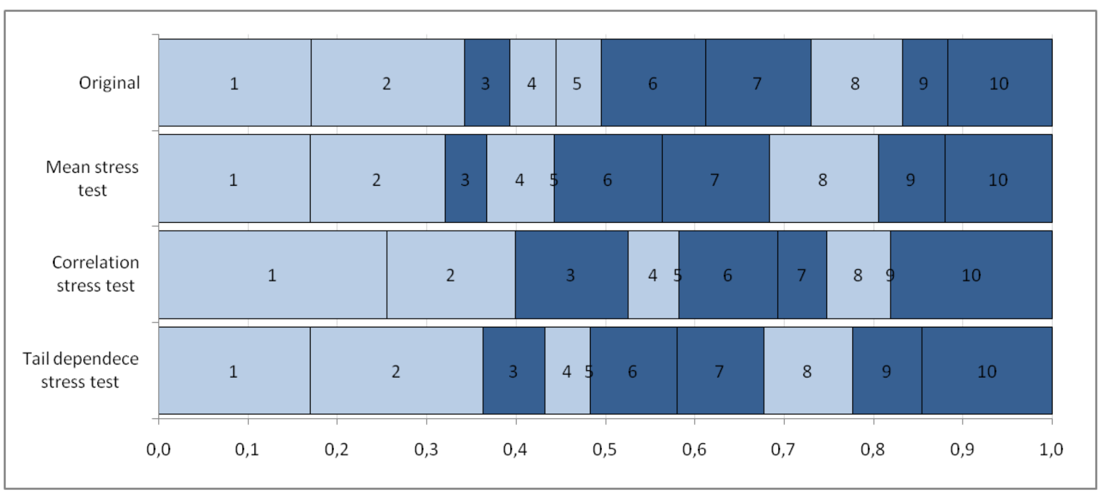


Figure 7: Composition of the original and stress-tested portfolios as per table 27. Low risk obligors displayed in light color and high risk obligors in dark color.

- The mean stress test results are consistent with the increases in PD: the weight of raised PD obligor shrinks and vice versa, except for the obligor 10, which requires separate explanation.

- The share of obligor 10 grows consistently, although 10 is a high risk/return obligor. In the mean test it consumes a part of the share of obligor 5. In correlation and especially tail dependence cases, another factor is involved: obligor 10 is comparatively little ”connected” to others. Default correlation matrix does not reveal this fact but we can observe other measures of dependence. Table 29 displays one possible measure: the ratio of number of defaults of obligor N to the number of defaults of all other obligors in scenarios where N defaults. It shows how likely any of other obligors are to default too when obligor N defaults.

Obligor	Defaults	Defaults other	Ratio
1	309	248	1.246
2	158	180	0.878
3	554	396	1.399
4	388	348	1.115
5	432	438	0.986
6	539	370	1.457
7	428	390	1.097
8	428	321	1.333
9	595	453	1.313
10	557	386	1.443

Table 29: Ratio of defaults of obligor N to the count of default occurences of other obligor in the same simulation row. For instance, obligor 1 defaults 309 times. In these 309 simulation scenarios, all other obligors default in total 248 times.

By this measure, obligor 10 is less connected to others than for example obligor 9. Obligor 6 is comparable, but the lower connectivity does not outweigh lower returns.

- Table 29 also explains why, although having the highest returns overall, obligor 7 is not favoured much when the dependence is stressed: it has one of the lowest dependence ratios. If obligor 7 is selected with a high weight, it contaminates the portfolio through correlations.

- The correlation stress test favors very strongly obligor 1. It has the second lowest PD, after obligor 2, and also has very much lower connectivity than obligor 2, which would clearly be a favourite otherwise (and obligor 2 is a favourite in the tail dependence stress test).

7.6 Significance of starting returns

As mentioned in section 5, in addition to the views, the Black-Litterman model introduces reverse optimized market equilibrium weights as the prior. Thus besides stress-testing (corresponding to the former contribution), it would be interesting to investigate how sensitive the mean-CVaR optimal portfolio weights actually are to the starting point of returns. Does mean-CVaR optimization is as dependent on small changes in returns as Markowitz mean-variance optimization and whether small changes in returns induce large swings in the optimal portfolio? For the purpose of this thesis I only check a few scenarios and leave the deeper analysis for further research.

Table 30 presents the findings. A and B are scenarios of different returns (before risk adjustment) when everything else stays the same. For reference I also present the optimal portfolio weights under the tail dependence stress test (as in subsection 7.4) with the new returns.

	Original			A			B		
Nr	Returns	Prior	Posterior	Returns	Prior	Posterior	Returns	Prior	Posterior
1	5.96%	0.1703	0.1699	5.27%	0.1346	0.1667	6.51%	0.1669	0.1714
2	5.16%	0.1716	0.1926	6.10%	0.2003	0.1777	5.04%	0.1692	0.1982
3	6.90%	0.0507	0.0702	7.81%	0.0657	0.0623	6.99%	0.0554	0.0728
4	5.03%	0.0516	0.0500	5.00%	0.0391	0.0516	5.59%	0.0561	0.0495
5	6.31%	0.0509	0.0000	6.91%	0.0331	0.0000	6.78%	0.0555	0.0000
6	7.57%	0.1174	0.0976	6.85%	0.0994	0.1014	7.85%	0.1099	0.0960
7	7.87%	0.1171	0.0973	7.72%	0.0986	0.1006	7.30%	0.1089	0.0965
8	5.70%	0.1025	0.0993	6.53%	0.1330	0.1017	6.22%	0.1131	0.0985
9	7.05%	0.0506	0.0770	7.63%	0.0658	0.0885	6.09%	0.0543	0.0735
10	7.73%	0.1173	0.1461	8.65%	0.1304	0.1495	8.58%	0.1106	0.1436
Tot	r	CVaR	CVaR	r	CVaR	CVaR	r	CVaR	CVaR
	6.06%	0.4158	0.3921	6.38%	0.3878	0.3923	6.23%	0.3857	0.3905

Table 30: Two scenarios of changes in returns: mean-CVaR optimal portfolios unstressed and under tail dependence stress test with different starting returns

From this small sample I conclude that the optimal portfolios in the original setup before stress testing indeed change with respect to the changed returns. Changes are not extreme and consistent with the movements in returns. Market equilibrium returns may indeed introduce some more stability. However, stressed optimal portfolio depends on the initial returns very little.

8 Conclusions, limitations and further research suggestions

I have shown that subjective views and stress-testing methods created originally for stock portfolios can successfully be applied for portfolios of loans.

The loan return distribution is characterized by the joint distribution of the times to default. We assume an exponential distribution for each individual obligor and a t-copula for the dependence structure of the joint distribution of the times to default. The parameters for the t-copula are estimated using historical equity returns, based on assumption that dependence between times to default is the same as dependence between equity returns. The assumption about default distribution enables us to simulate the loan ”market” outcomes. Subsequently we apply numerical methods of CVaR optimization.

Even more importantly, different aspects of the market can be stressed. In addition to stressing the PDs, we have stressed the dependence between default events in two ways: if correlation of joint default distribution is increased and if we only increase the probability of events in lower tails of distribution to occur together, that is, the low values of times to default to occur together. Both stress tests were done using Meucci (2008) Entropy Pooling methodology, which applies for loan portfolios because it does not assume a specific distribution for the risk factors. We have exploited a feature of Entropy Pooling that views can be expressed on a *function* of risk factors. It is very useful for working with default distributions to choose the function as the one year default indicator. We start working with the joint distribution of times to default, because we can parametrize the times to default distribution using equity returns, but express the views on the discrete default distribution, because it is much more intuitive to do so.

The mean stress test (increasing PDs) could be implemented in a different way, by resimulating the joint times to default distribution with the different hazard rate parameters for the marginals, because a change in a PD resolves directly into a change in the hazard rate which in turn affect the time to default in a predictable way. Also, the tail dependence stress test could be done by including the cases with many defaults repeatedly into the simulation panel. (There is no such a clear cut way to stress test the correlation.) Still, Entropy Pooling is convenient because it is a generalization: it ties different stress tests together in one framework. Many different views can be expressed with different confidence at once. Besides, one could express views on any other parameter of the joint default distribution that I did not go into: ranking and relative views to name a few. The method does not require resimulating the panel and is very fast computationally.

The stressed default distribution can be used in many different ways for risk management and optimization purposes. We have optimized it again with CVaR as risk measure and compared the weights. The difference in weights is subtle, with preference for lower risk obligors and least connected obligors when defaults are more likely to occur together.

Entropy Pooling is shown to work on loan portfolios and may be directly applied in practice on real data. Further research is proposed mainly for insufficiently explored auxiliary problems:

1. Well-known tendency of Markowitz MV optimization to produce extreme, undiversified and sensitive to small changes in returns is countered by reverse optimized market equilibrium returns prior by Black-Litterman. It is unclear whether the weights are just as sensitive to small changes in returns if the risk measure is CVaR.

2. If so, numerical methods (inverse optimization) may be used to find the mean-CVaR equilibrium prior. Our efforts (section 5 and appendix B) were limited and inconclusive.

3. In real life instead of individual obligors we would work with risk factors, generally sectors of economy.

References

[1] Meucci, A. 2008. *Fully flexible views: theory and practice*, Risk Magazine, October 2008, p.100-106.

[2] Meucci, A. 2006. *Beyond Black-Litterman in practice: a five-step recipe to input views on non-normal markets*, Risk Magazine, February 2006, p.87-92.

[3] Meucci, A., 2008a. *Estimation of Structured t-Copulas* Working paper, SSRN.

[4] Li, D.X., 2000. *On default correlation: a copula function approach*. Journal of Fixed Income, vol. 9, p.43-54.

[5] Mashal, R., Naldi, M., 2002. *Pricing multiname credit derivatives: heavy tailed hybrid approach*. Working paper, Columbia University.

[6] Mashal, R., Naldi, M., Zeevi, A., 2003. *On the dependence of equity and asset returns*. Risk Magazine, October 2003, p.83-87.

[7] DiClemente, A., Romano, C., 2004. *Measuring and optimizing portfolio credit risk: a copula-based approach*. Economic Notes, Volume 33, Number 3, November 2004, p.325-357.

[8] Walters, J., 2009. *The Black-Litterman model in detail*. Working paper, www.blacklitterman.org.

[9] Rockafellar, R.T., Uryasev, S., 2000. *Optimization of conditional Value-at-Risk*. Journal of Risk, vol.2, p.2141.

[10] Ahuja, R.K., Orlin, J.B., 2001. *Inverse optimization*. Operations Research, Vol. 49, No. 5, SeptemberOctober 2001, p.771783.

[11] Krokhmal,P. , Palmquist, J., Uryasev, S., 2002. *Portfolio optimization with conditional Value-at-Risk objective and constraints*. Risk Magazine, April 2002, p.11-27.

[12] Giacometti, R., Bertocchi, M., Rachev, S.T., Fabozzi, F.J., 2007. *Stable distributions in the Black-Litterman approach to asset allocation* Quantitative Finance, Volume 7 (2007), Issue 4, p.423-433.

[13] Iyengar, G., Kang, W., 2005. *Inverse conic programming with applications*. Operations Research Letters 33 (2005) p.319330.

[14] Tantawy, S.F., 2007. *A new method to find all alternative extreme optimal points for linear programming problem*. Australian Journal of Basic and Applied Sciences, 1(1), p. 38-44.

[15] Paulraj, S., Chellappan, C., Natesan, T.R., 2006. *A heuristic approach for identification of redundant constraints in linear programming models*. International Journal of Computer Mathematics, Vol. 83, Nos. 89, AugustSeptember 2006, p.675683.

[16] Athayde, G.M. de, Flôres, R.G., 1997. *Introducing higher moments in the CAPM: some basic ideas* Working paper, EPGE / Fundação Getúlio Vargas, Rio de Janeiro.

[17] Romero-Meza, R., Laengle, S., 2007. *Implementation of conditional Value-at-Risk (CVaR): the case of Chilean pension funds*. Working paper, Universidad de Chile.

[18] Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1998. *Coherent measures of risk*. Mathematical Finance, 1998, 9(3), p.203-228.

[19] Agarwal, V., Naik, N.Y., 2004. *Risks and portfolio decisions involving hedge funds*. The Review of Financial Studies, Vol. 17, No. 1 (Spring, 2004), pp. 63-98.

[20] Fabozzi, F., Kolm, P. N., Pachamanova, D., Focardi, S. M., 2007. *Robust Portfolio Optimization and Management* (Wiley).

[21] Rardin, R.L., 1997. *Optimization in operations research* (Prentice Hall).

[22] Meucci, A. 2006. *Beyond Black-Litterman: Views on Non-Normal Markets*, Working paper, SSRN.

[23] Meucci, A., 2009. *Enhancing the Black-Litterman and related approaches: views and stess-tests on risk factors*, 2008. Journal of Asset Management, Volume 10, Number 2, June 2009 , pp. 89-96(8).

[24] Meucci, A., 2005a. *Risk and asset allocation* (Springer).

A Linear CVaR optimization problem

Here I show how to write down (16) in order to solve it with an optimization package. In matrix form and converted to equality shape (for use with SeDuMi package all constraints should be binding - we achieve that by adding slack variables) (16) becomes

$$\begin{aligned} Minimize \quad & c^T\mathbf{x} \\ s.t \quad & \mathbf{Ax} = b \\ & \mathbf{x} \geq 0 \end{aligned} \tag{77}$$

$$\mathbf{A} = \begin{pmatrix} -\sum_j p_j m_{1j} & \dots & -\sum_j p_j m_{Kj} & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ -1 & \dots & -1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ -m_{11} & \dots & -m_{1K} & -1 & -1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\ -m_{J1} & \dots & -m_{JK} & -1 & 0 & \dots & -1 & 0 & 0 & \dots & 1 \end{pmatrix} \tag{78}$$

$$b = (-r \quad -1 \quad 0 \quad \dots \quad 0)' \tag{79}$$

$$c = (0 \quad \dots \quad 0 \quad 1 \quad \frac{1}{1-\beta}p_1 \quad \dots \quad \frac{1}{1-\beta}p_J \quad 0 \quad 0 \quad \dots \quad 0)' \tag{80}$$

In concise notation:

$$\mathbf{A} = \begin{pmatrix} -pM & 0 & \mathbf{0}_{1\times J} & 1 & \mathbf{0}_{1\times J} \\ \mathbf{1}_{1\times S} & 0 & \mathbf{0}_{1\times J} & 0 & \mathbf{0}_{1\times J} \\ -M & -\mathbf{1}_{J\times 1} & -\mathbf{I}_J & \mathbf{0}_{J\times 1} & \mathbf{I}_J \end{pmatrix} \tag{81}$$

$$b = (-r \quad 1 \quad \mathbf{0}_{1\times J})' \tag{82}$$

$$c = (\mathbf{0}_{1\times K} \quad 1 \quad \frac{1}{1-\beta} * \mathbf{1}_{1\times J})' \tag{83}$$

The size of the decision variable now is K+1+J+1+J. K is the number of obligors, so first K of the final weight vector will be the true portfolio holdings. Next, K+1'th entry is VaR, referred to as ζ. Next J are the z's, the dummies resulting from CVaR function linearization. Final 1+J are dummies required for converting inequalities to equalities. In total there are K+1+J+1+J variables and 2+J constraints in this problem. (J was the number of market outcome scenarios). The first K entries of the solution $\mathbf{x}^* = (x_1, \dots, x_K)$ is the vector of weights that define mean-CVaR optimal portfolio.

B Inverse optimization

B.1 Inverse linear problem with a two-dimensional example

A typical optimization problem is a forward problem because it identifies the values of observable parameters (optimal decision variables), given the values of the model parameters (cost coefficients, right-hand side vector, and the constraint matrix). An inverse optimization problem consists of inferring the values of the model parameters (cost coefficients, right-hand side vector, and the constraint matrix), given the values of observable parameters (optimal decision variables). Recently inverse optimization gained interest in areas as diverse as geophysical science, medical imaging, transportation flow analysis, railroad scheduling and others, among which also portfolio optimization is found. For instance, Iyengar and Kang(2004) [13] describe a case of finding optimal portfolio in Sharpe equilibrium with transactions costs. In principle, their problem is similar to ours. We assume that market is in equilibrium, but equilibrium is of non-trivial kind - we cannot inverse the optimum easily to get back to optimum weights. Iyengar and Kang's case is complicated by transaction costs, ours by equilibrium defined by CVaR rather than variance as risk measure and nonnormal returns.

We gain insight into inverse optimization by studying a two-dimensional example and looking at its graphical representation first. Called "Two Crude petroleum model", our example problem is defined as follows.

$$
\begin{array}{lrrll}
min & 20x_1 & +15x_2 & & \text{(total cost; coefficients:} \\
 & & & & \text{ costs of Saudi and Venezuelan oil per gallon)} \\
s.t & 0.3x_1 & +0.4x_2 & \geq 2.0 & \text{(1 gasoline requirement)} \\
 & 0.4x_1 & +0.2x_2 & \geq 1.5 & \text{(2 jet fuel requirement)} \\
 & 0.2x_1 & +0.3x_2 & \geq 0.5 & \text{(3 lubricant requirement)} \\
 & x_1 & & \leq 9 & \text{(4 Saudi availability)} \\
 & & x_2 & \leq 6 & \text{(5 Venezuelan availability)} \\
 & x_1, & x_2 & \geq 0 & \text{(6,7 nonnegativity)}
\end{array}
\tag{84}
$$

As the problem has only two dimensions we can plot the constraints and the objective graphically.

Points A,B,C,D,E where constraints intersect define the feasible region of the problem. Thin lines are lines where the objective function has the same value. It is easy to solve the problem just by looking at the picture: optimum is the point B, where feasible region is bordered by constraints 1 and 2 (i.e., constraints 1 and 2 are binding). Solve linear equation $0.3x_1 + 0.4x_2 = 2.0$; $0.4x_1 + 0.2x_2 = 1.5$ and we obtain the point [2.0;3.5]. Plugging it into objective function yields 92.5.

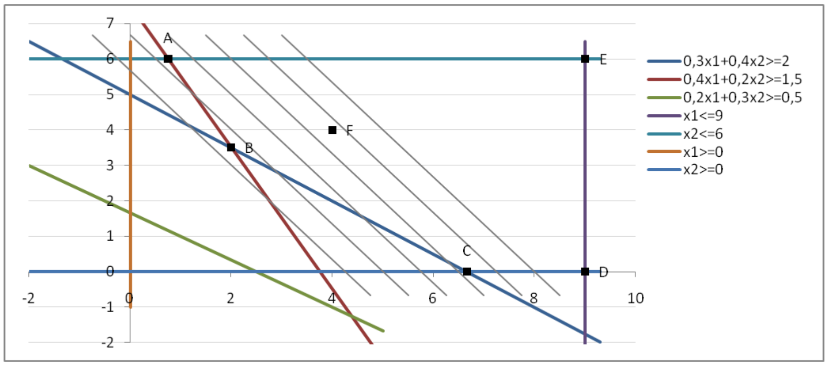


Figure 8: Feasible region and objective of two crude problem

However, having solved the forward problem and knowing the optimum buying quantities we can still ask many questions. For instance, what if oil prices change? Consider we start production and buy specified amounts of two types of oil. When prices move with respect to each other, we cannot easily change the process, so there is no guarantee that the point will still be optimal. What are the ranges of oil prices where the point B is still optimal? When we should move to another point?

Inverse problem is stated as follows. Given the current quantities, what are the prices such that the current point is optimal across all quantities. It is not a coincidence that we are reminded of the dual formulation here. The dual is the subsidiary optimization model, defined over the same input parameters as the primal but characterizing the sensitivity of primal results to changes in inputs. Dual variables reflect the rate of change in primal optimal value per unit increase from the given right-hand-side value of the corresponding constraint. They provide implicit prices for the marginal unit of the resource modeled by each constraint as its right-hand-side limit is encountered.

If the primal problem is

$$\begin{array}{rll} Minimize & \sum_{j\in J} c_j x_j & \\ s.t. & \sum_{j\in J} a_{ij} x_j \geq b_i & \forall i \in I \\ & x_j \geq 0 & \forall j \in J \end{array} \tag{85}$$

,

the dual problem is

$$\begin{array}{rll} Maximize & \sum_{i\in I} b_i v_i & \\ s.t. & \sum_{i\in I} a_{ij} v_i \geq c_j & \forall j \in J \\ & v_i \geq 0 & \forall i \in I \end{array} \tag{86}$$

.

The dual of the Two Crude petroleum model is

$$
\begin{array}{llllllll}
max & 2v_1 & +1.5v_2 & +0.5v_3 & +9v_4 & +6v_5 & & \\
s.t & 0.3v_1 & +0.4v_2 & +0.2v_3 & +1v_4 & & \leq 20 & \\
 & 0.4v_1 & +0.2v_2 & +0.3v_3 & + & +1v_5 & \leq 15 & \\
 & v_1, & v_2, & v_3 & & & \geq 0 & \\
 & v_4, & v_5 & & & & \leq 0 &
\end{array} \tag{87}
$$

The optimal solution is [20;35;0;0;0] with the dual objective function value 92.5. It tells us about marginal unit prices. If the first activity constraint is modified by 1 unit, the objective function value will change by 20. Besides, we notice that constraints 1 and 2 of the primal were active (binding) (from the principle of primal complementary slackness).

The main dual constraints $\sum_{i \in I} a_{ij} v_i \leq c_j$ keep activity prices below the true cost. It may seem reasonable to ask for more, i.e., demand that activity prices match exactly corresponding c_j. But more importantly, we want the dual variables to measure resource value at optimality. The only (nonnegative) primal variables involved in an optimal solution are those with optimal $x_j > 0$. Limiting perfect valuation to this more limited list of activities produces our final set of requirements - *dual complementary slackness* conditions. Either a nonnegative primal variable has optimal value $x_j = 0$ or the corresponding dual prices v_i must make the jth dual constraint $\sum_{i \in I} a_{ij} v_i \leq c_j$ active. Dual complimentary slackness thus binds dual constraints with the primal optimum.

So far we have determined that the primal solution $\mathbf{x}^*$ and the dual solution $\mathbf{v}^*$ are optimal for their respective problems if x^* is feasible for equations (85), $\mathbf{v}^*$ is feasible for equations (86), and together they satisfy the dual complementary slackness conditions. Solving the dual, then, we automatically get the optimal solution of the primal. Having dual problem with satisfied dual complimentary slackness conditions (derived from the primal optimum) and the primal optimum itself, we can reconstruct the primal problem. Reconstructing the primal is exactly what we are aiming for in inverse optimization.

Let us build the dual complimentary slackness conditions from known primal solution. We started with the primal problem (85). We want now to perturb the cost vector $\mathbf{c}$ to $\mathbf{d}$ in such a way that given $\mathbf{x^0}$ (which must be a feasible point, but not necessarily optimal) would be the optimal solution.

Let $\mathcal{B}$ denote the set of active constraints in the primal formulation (if all were active, $\mathcal{B}$=$\mathcal{I}$), $\mathcal{L}$ the set of $\mathbf{x}$ at lower bound(0) (if all $\mathbf{x}$ were at lower bound, $\mathcal{L}$=J), and $\mathcal{F}$ the set of $\mathbf{x}$ above lower bound ($\mathcal{L}$+$\mathcal{F}$=J). Following the principle of dual complimentary slackness, the dual formulation may be rewritten as follows:

$$
\begin{array}{rll}
Maximize & \sum_{i \in B} b_i v_i & \\
s.t. & \sum_{i \in B} a_{ij} v_i \geq c_j & \forall j \in L \\
s.t. & \sum_{i \in B} a_{ij} v_i = c_j & \forall j \in F \\
 & v_i \geq 0 & \forall i \in B
\end{array}
\tag{88}
$$

We have a given solution $\mathbf{x^0}$ and are looking for a cost vector $\mathbf{d}$ (preferably in some measure close to $\mathbf{c}$) which would make it optimal. $\mathcal{B}$, $\mathcal{L}$ and $\mathcal{F}$ are sets corresponding to the solution $\mathbf{x^0}$ now. $\mathcal{B}$ consists of all constraints where $\sum_{j \in J} a_{ij} x_j = b_i$. $\mathcal{L}$ consists of all j where $x_j^0 = 0$. $\mathcal{F}$=J-$\mathcal{L}$.

if $\mathbf{v}^*$ is the optimal solution of (88), then $\sum_{i \in B} a_{ij} v_i^*$ would give us the answer: the coefficients of the primal objective d_j, making the $\mathbf{x^0}$ optimal.

However, there generally are many $\mathbf{d}$'s which make $\mathbf{x^0}$ optimal. Returning back to Two Crude example, we notice that the optimal point B does not change if the coefficients c_j=[20;15] move a little. The angle of objective contour has space to move between the angle of constraint 1 and constraint 2.

We can even build a table to illustrate how far c_2 can move, if we fix c_1 to existing value 20. By analogy similar table can be built for c_2 (table 31).

Optimum	c_1	c_2 from	c_2 to	Optimum	c_2	c_1 from	c_1 to
A	20	$-\infty$	10	A	10	$-\infty$	11.25
B	20	10	80/3	B	10	11.25	20
C	20	80/3	∞	C	10	20	∞

Table 31: Values of c_1 and c_2 defining different optimums. Left: c_1 fixed, Right: c_2 fixed.

That is, any problem (with the same constraints) with objectives from $20x_1 + 10x_2$ to $20x_1 + 80/3x_2$ (coefficient of x_2 is moving), including our original $20x_1 + 15x_2$ optimizes to the same point [2.0;3.5]. There is a catch that marginal problems $20x_1 + 10x_2$ to $20x_1 + 80/3x_2$ both have continuum solutions. We will come back to this later.

Notice that it is the ratio that matters, as the problem is linear. We can cover the complete space $[c_1, c_2]$ as displayed in the figure 9.

Subsequently, entering $\mathbf{x^0}$=[2.0;3.5] into the dual and solving it, we may obtain $[d_1, d_2]$ anywhere in the area $\mathbb{B}$. Of multiple possible $\mathbf{d}$'s, how do we select one? One way is to select the $\mathbf{d}$, which is closest to the original $\mathbf{c}$ in some way. Norms L_1, L_p and L_∞ are proposed, resulting in optimization problems

$\min \sum_j w_j |d_j - c_j|$,
$\min \sum_j w_j \|d_j - c_j\|_p$, and

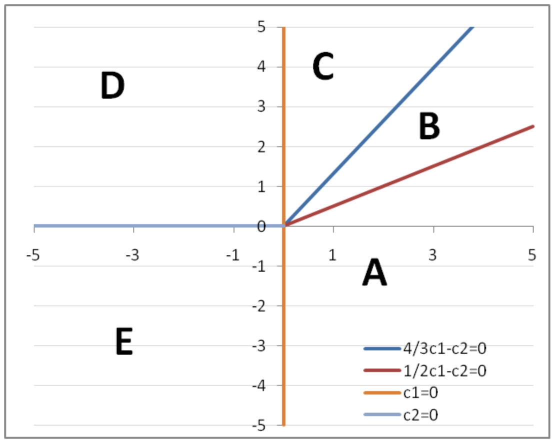


Figure 9: Optimal points $\mathbf{x}^0$ in space $[c_1, c_2]$

$\min \max_j w_j|d_j - c_j|$,
respectively.

For linear problems L_1 and L_∞ are appropriate, as we stay in linear realm after the minimum norm objective is included into the formulation. Let us choose L_1 and reformulate our inverse programming problem as

$$\begin{array}{rll} Minimize & \sum_j w_j|d_j - c_j| & \\ s.t. & \sum_{i\in B} a_{ij}v_i \geq c_j & \forall j \in L \\ s.t. & \sum_{i\in B} a_{ij}v_i = c_j & \forall j \in F \\ & v_i \geq 0 & \forall i \in B \end{array} \tag{89}$$

It is well known that minimizing $\sum_j w_j|d_j - c_j|$ is equivalent to minimizing $\alpha_j + \beta_j$, subject to $d_j - c_j = \alpha_j - \beta_j$, $\alpha_j \geq 0$, $\beta_j \geq 0$.

Restating the problem to get rid of the absolute value and introducing dummy variables λ_j to make all constraints equalities:

$$\begin{array}{rl} Minimize & \sum_j w_j\alpha_j + \sum_j w_j\beta_j \\ s.t. & \sum_{i\in B} a_{ij}v_i - \alpha_j + \beta_j + \lambda_j = c_j \quad \forall j \in L \\ s.t. & \sum_{i\in B} a_{ij}v_i - \alpha_j + \beta_j = c_j \quad \forall j \in F \\ & v_i \geq 0 \quad \forall i \in B, \quad \alpha_j \geq 0, \quad \beta_j \geq 0 \quad \forall j \in J, \quad \lambda_j \geq 0 \quad \forall j \in L \end{array} \tag{90}$$

If optimal solution of this problem is $\mathbf{v}^*$, α^*, β^*, λ^*, the optimal $\mathbf{d}$ is

$$d_j = \sum_{i\in B} a_{ij}v_i^* \tag{91}$$

However, norm L_1 still cannot ensure unique $\mathbf{d}$. That is normal. Worse, $\mathbf{d}$ is always a marginal solution - in figure 9, on the border between $\mathbb{A}$ and $\mathbb{B}$ - which when fed back to the original primal problem, does not have a unique optimum. We would expect to get the original starting point $\mathbf{x^0}$ as the solution of the primal with the coefficients of the objective $\mathbf{d}$ instead of $\mathbf{c}$. But we get $\mathbf{x^0}$ only if we move $\mathbf{d}$ by at least a very small amount from the border. Again returning to the Two Crude example, if $\mathbf{x^0} = [0.75; 6]$, in the intersection of constraint 2 and 5, the optimal $\mathbf{x^0}$ is [20;10]. The primal with this objective has a continuum of optimums, one among which is $\mathbf{x^0}$, but clearly this is not satisfactory. Only if $\mathbf{d}$ is moved by an ϵ into $\mathbb{A}$ territory - to $[20, 10 - \epsilon]$ - optimum $\mathbf{x^0}$ is unique.

As the problem is linear, it is straightforward to find the direction of this movement: simply continue the same direction as to get from $\mathbf{c}$ to $\mathbf{d}$ by a small amount more. This direction is $\mathbf{d}$-$\mathbf{c}$. The new point $\mathbf{d}^*$ which guarantees unique optimum then is

$$\mathbf{d}^* = \mathbf{d} + \epsilon(\mathbf{d} - \mathbf{c}). \tag{92}$$

As we have seen, not all the starting points $\mathbf{x^0}$ will find us a cost vector which with the point $\mathbf{x^0}$ would be optimal. Choices for the location of the point $\mathbf{x^0}$ are as follows:

1. $\mathbf{x^0}$ is a feasible point inside the feasible region (point F). No constraint is binding, no coordinates are at lower limit. There is no cost vector with which the solution is optimal to the primal. $\mathcal{B}$ is empty; $\mathcal{L}$ is empty; $\mathcal{F}$ is empty $\rightarrow$ inverse problem cannot even be written down. This should not happen with CVaR equilibrium problem, as it is very hard, if not impossible, to hit the point not activating at least one constraint, because CVaR equilibrium problem has very many constraints compared to the number of decision variables. Perhaps it is even possible to prove this conjecture.

2. $\mathbf{x^0}$ is on a border of the feasible region, activating less constraints than necessary to be an extreme point (point G). The cost vector with which the point is optimal should be parallel to the limiting facet of the feasible region. In the figure 10 the contours of the optimal cost vector are dashed lines. In two dimensions the cost vector is unique, but with higher dimensions the border is much more likely a facet rather that an edge, and the solution is not unique in this case.

3. $\mathbf{x^0}$ is an infeasible point (point H). Naturally, there is no cost vector with which the solution is optimal.

4. $\mathbf{x^0}$ is an extreme point (point B). In this case the set of cost vectors with which solution is optimal contains more than one point. Among possible choices we have our original cost vector of Two Crude problem.

In the case of 2 and 4, then, we have obtained the optimal cost vector without much difficulty. If we are especially lucky, the vector is unique already; otherwise we choose

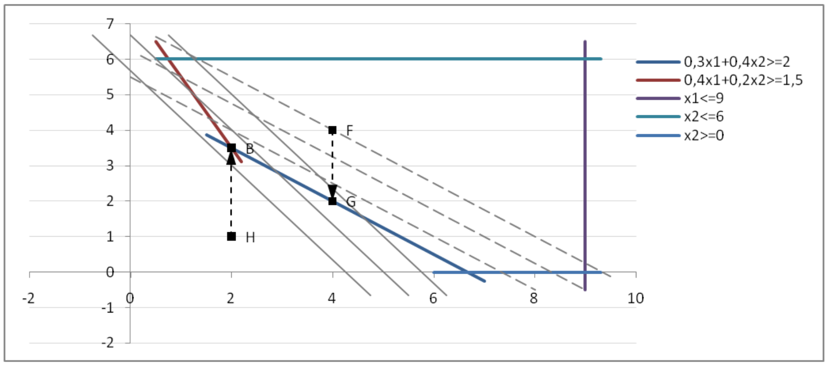


Figure 10: Options for $\mathbf{x^0}$

from alternatives, for instance, move the result by a small amount to the same direction as **d-c**.

1 and 3 remains. What to do? Mathematically, in both cases, the answer is to find a point on the border of the feasible region as close to $\mathbf{x^0}$ as possible. It remains to be discussed how appropriate this answer is in business application. The measure of the distance may become very important, best if it is also intuitive.

3, infeasible, is a little easier. Taken norm L_1 as a measure of distance, all that is necessary is to solve a LP with the same constraints as the original primal LP and distance measure as objective.

$$\begin{array}{rll} Minimize & \sum_j w_j |x_j^0 - x_j| & \\ s.t. & \sum_{j \in J} a_{ij} x_j \geq b_i & \forall i \in I \\ & x_j \geq 0 & \forall j \in J \end{array} \tag{93}$$

We have already seen how to convert the distance measure in L_1 into a linear problem.

$$\begin{array}{rll} Minimize & \sum_j w_j \alpha_j + \sum_j w_j \beta_j & \\ s.t. & \sum_{j \in J} a_{ij} x_j \geq b_i & \forall i \in I \\ & x_j^0 - x_j - \alpha_j + \beta_j = 0 & \forall j \in J \\ & x_j \geq 0, \alpha_j \geq 0, \beta_j \geq 0 & \forall j \in J \end{array} \tag{94}$$

The optimal $\mathbf{x}^*$ may be an extreme point or a point on the facet.

With case 1, the procedure is more complex. First, notice that the problem to move F to the border of the feasible region would be the same as case 3 problem if we could invert the feasible region. We move infeasible point F as close as possible to the feasible region, if we redefine the feasible region as the *outside* of the original feasible region. Now, how to invert the original feasible region? A first answer that comes to mind is to take all the constraints and simply invert the sign: turn all "$\leq$" to "$\geq$" and vice versa (and keep equalities as they are). But we notice soon that there are constraints which do not define the feasible region at all, for instance, $x_1 \geq 0$ in Two Crude example. Inverting this constraint would move the outside further away than we wish. (As a side note, it may not be a very bad thing. For sure if we just solve the distance minimizer problem with naively inverted constraint set, we are almost surely getting the point infeasible to the original problem. Then we are back to already solved case 3. Problem is, the optimum with naive inverse feasible region may be very far from the original feasible region). First we need to detect the redundant constraint of the original problem - in Two Crude example these would be constraint 3 and lower bound of x_1.

There are several known algorithms for detecting redundant constraints of LP. One example is explored in the Paulraj, Cheppalan and Natesan (2006) [15]. Once the set of essential (not redundant) constraints is identified, we invert the inequalities and solve the distance minimizer problem to move from F to G. Again the point on the border of the feasible set may or may not be extreme.

I suspect that stability problems arise if the identified border point is not extreme. Especially when the border point is on the edge and the optimizer cost vector is unique. If we optimize the primal with this unique cost vector, the resulting optimum depends heavily on the optimizing algorithm. We may or may not - and more likely not - will get back our starting border point as the optimum. If the objective contour is parallel to the constraint, the choice between multiple alternative optima depends on many things. The simplex method will always return an extreme point. With the interior point method surely the optimum will be on the edge, but not certainly the same. Moreover, move the coordinates of the unique cost vector by a little bit and the optimum will surely move to an adjacent extreme point. Thus it would be desirable to find a nearest extreme point of the feasible region instead of finding just a nearest point in the feasible region if the starting point $\mathbf{x^0}$ is either infeasible or not extreme (vast majority of cases will fall into these two categories in practice).

One solution would be to enumerate all adjacent optimal extreme points next to the border point and choose the nearest. The same kind of problem would be to find an extreme point of the primal, nearest to infeasible $\mathbf{x^0}$ from the start. Both formulations are difficult subproblems on their own.

Another solution - instead of finding the nearest feasible point on the facet of feasible region, we would ”extend” the constraints in the convex way to form new feasible region with an extreme point in our infeasible point.

It is straightforward to identify active constraints but not so clear how to make them pass through the originally infeasible point H. Convex linear combination of all adjacent extreme points would most likely do it, again returning us to the problem of enumerating all optimal extreme points. Tantawy (2008) [14] discusses this issue.

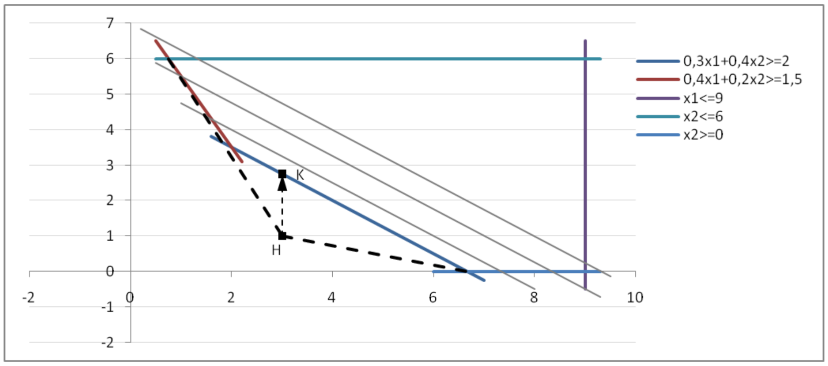


Figure 11: Extend constraints to form new extreme point

B.2 Inverse of linearized CVaR minimization

Now that we have learned basic machinery of inverse optimization, let us return to the CVaR minimization problem (11).

The problem is linear, so we should be close to inverting it: given portfolio weights, what are the returns, with which the given weights are optimal? Given weights will be then interpreted as market capitalization weights and optimizing returns - "market mean-CVaR equilibrium returns".

There are two hurdles to overcome before we can answer this question. First, returns should switch places with CVaR and get up to the objective while CVaR moves down to constraints. It has been proven that efficient mean-CVaR frontier stays the same. In its return maximizing shape the problem is (13).

Second, returns still participate in the constraints through the relationship $z_j + \zeta + f(\mathbf{x}, y_j) \geq 0$. If $f(\mathbf{x}, y_j)$ are truly market outcomes, in case of default the entry is -lgd and in case of non-default the entry is the return. For instance, continuing three obligor problem, if returns are [9%; 5%; 4%] and recovery rate is 0%, the matrix $\mathbf{M}$ looks more or less like (76), with 3 columns and 10 rows.

We want it exactly in this shape in the objective, but in the constraint we wish not to see the returns and risk combined into the risk-adjusted returns. There is no (easy) solution to the problem of finding returns with which given weights are optimal if we cannot isolate the movable parts to objective alone, at least the theory above does not apply.

Several alternatives can be proposed:

1. Leave the original returns, r^0, in the dual optimization constraint fixed. Vary only the returns in the objective.

2. Set the returns to some other sensible value.

3. Set the returns to 0. In this case we implicitly change our risk measure. It is no longer the CVaR as usual, the average of losses above certain quantile, but the average of ”pure” losses above certain quantile.

It is hard to say which option is the most sensible. At least all seem to be viable. For the numerical trials in the following section I assumed the third option: substitute returns with 0 and this way measure risk by ”pure loss” CVaR, or CVaR^p. Preliminary numerical tests show that CVaR and CVaR^p are not very different.

CVaR minimize problem now becomes

$$
\begin{array}{rl}
Maximize & \sum_i \sum_j p_j m_{ij} x_i \\
s.t. & \zeta + \frac{1}{1-\beta} \sum_{j=1}^{J} p_j z_j \leq \omega \\
 & z_j + \zeta + \sum_i m_{ij}^p x_i \geq 0 \quad \forall j \in J \\
 & \sum_i x_i = 1 \\
 & z_j \geq 0 \quad \forall j \in J, \quad x_i \geq 0 \quad \forall i \in I
\end{array} \tag{95}
$$

where m_{ij} are the elements of full market outcome matrix and m_{ij}^p are the elements of pure loss matrix.

Let us write down the inverse problem now: find the returns with which given weights minimize the CVaR across all weights. If the forward problem is (95), concisely

$$
\begin{array}{rl}
Minimize & c^T x \\
s.t & \mathbf{A}x \leq b \\
 & x \geq 0
\end{array} \tag{96}
$$

its inverse is

$$
\begin{array}{rl}
Minimize & c^{*T} v \\
s.t & \mathbf{A}^* v = b^* \\
 & v \geq 0
\end{array} \tag{97}
$$

$\mathcal{B}$: theset of active constraints in (96) $i : \sum_j a_{ij} x_j = b_i$; $\mathcal{N}$: set of inactive constraints. $\mathcal{L}$: set of x where $x_j = 0$; $\mathcal{F}$: set of $x_j > 0$. m is the size of $\mathcal{B}$; l is the size of $\mathcal{L}$. We need to sort $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{x}$ into the order by binding/not binding (binding constraints first) and at lower limit/not at lower limit (x_j =0 first) for writing the inverse. For instance, $\mathbf{A_{bL}}$ is the submatrix in $\mathbf{A}$ where the constraints are active and $x_j = 0$.

$$
\mathbf{A} = \begin{pmatrix} \mathbf{A_{bL}} & \mathbf{A_{bF}} \\ \mathbf{A_{nL}} & \mathbf{A_{nF}} \end{pmatrix} \tag{98}
$$

$$
b = (b_b \quad b_n)', \qquad c = (c_L \quad c_F)', \qquad x = (x_L \quad x_F)' \tag{99}
$$

Then inverse:

$$
\mathbf{A}^* = \begin{pmatrix} -\mathbf{A_{bL}^T} & -\mathbf{I}_L & 0_{L\times F} & \mathbf{I}_L & 0_{L\times F} & \mathbf{I}_L \\ -\mathbf{A_{bF}^T} & 0_{F\times L} & -\mathbf{I}_F & 0_{F\times L} & \mathbf{I}_F & 0_{F\times L} \end{pmatrix} \tag{100}
$$

$$b^* = c, \qquad c^* = (\mathbf{0}_{1\times m} \quad w_{1\times J} \quad w_{1\times J} \quad \mathbf{0}_{1\times L})' \tag{101}$$

The solution: $v^* = (v_1, \ldots, v_m)'$. In order to get the cost vector d, we multiply the optimum by $-A_{bL}^T$:

$$\mathbf{d} = -\mathbf{A}_{bL}^T \mathbf{v}^* \tag{102}$$

and the result need to be sorted back to the original order.

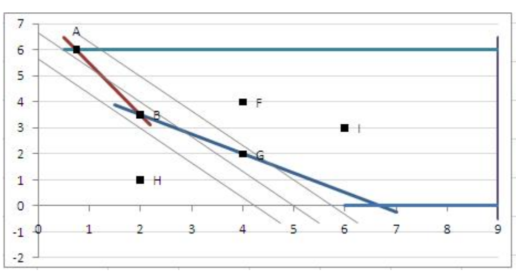
Concerning weights w_j in $\mathbf{c}^*$ which define how far $\mathbf{d}$ can be from $\mathbf{c}$ in norm L_1. In the original primal formulation the components of $\mathbf{c}$ are all zeros, starting from J, i.e. all linearization dummies get cost 0; and we would like $\mathbf{d}$ to retain this property. Thus we set first $S+1$ w_j to 1 and the other J w_j to 1000.

B.3 Some numerical results

Let us begin with a 10 scenarios example. Portfolio returns r_0 are as before: (9%, 5%, 4%), LGD=100%. Out of 10 scenarios, 1 contains all 3 obligors defaulting and additional 3 - one different obligor defaulting each. Individual PDs thus are 20%, and default dependence structure is such that in one case all 3 default. We want to maximize return, given that 80% pure loss CVaR should not be higher than 0.80.

Optimal weights with this specification - solution of the primal (95) - are [0.6;0.4;0].

Our goal is to determine market equilibrium returns given specific portfolio weights as an input. We solve (97) with different values $\mathbf{x}$. As noted above, $\mathbf{x}$ does not participate in the (97) expression directly but only through defining sets of binding constraints $\mathcal{B}$ (and its complement nonbinding $\mathcal{N}$) and variables at lower limit $\mathcal{L}$ (and its complement $\mathcal{F}$).



B	Optimum
A	Suboptimal extreme point
G	Point on edge
H	Infeasible point
I	Feasible point inside
F	Feasible point inside equal weights

Figure 12: The points illustrating the results of inverse optimization

We illustrate the market equilibrium returns resulting from different kinds of input $\mathbf{x}$. Different kinds of points in two dimensional Two Crude example are displayed in figure 12. Our example has many more dimensions, but idea remains the same.

Table 32 displays the results. Epsilon=0.01. x^0 is the starting point of weights; x^1 - closest feasible point in case x^0 is infeasible; r_{opt} - the optimum of inverse problem, the returns with which x^0 (x^1) is optimal; r^{ϵ}_{opt} - solution r_{opt} moved by an ϵ as described above, because r_{opt} is usually the border solution; x^* - test solution of the primal where original returns are replaced by r^{ϵ}_{opt}.

The criteria of goodness of the results is how close x^0 (x^1) is to x^*. Perfect match is not likely, neither for feasible nor for infeasible points, because the x^* will always be a corner. If an initial point is feasible, is very likely not a corner. If it is infeasible, then the closest feasible point x^1 is more likely a point on edge than a corner.

	x^0	x^1	r_{opt}	r^{ϵ}_{opt}	x^*
	0.6		0.09	0.09	0.6
B	0.4		0.05	0.05	0.4
	0		0.05	0.05001	0
	0.6		0.09000	0.09000	0.6
A	0		0.04315	0.04308	0
	0.4		0.04315	0.04318	0.4
	0.6		0.09000	0.09000	0.6
G	0.2		0.043761	0.043698	0
	0.2		0.043761	0.043798	0.4
	0.2	0.27483	0.042932	0.042461	0
H	0.8	0.60000	0.050000	0.050000	0.6
	0	0.12517	0.042932	0.042961	0.4
	0.59		0.09000	0.09000	0.6
F	0.2		0.043761	0.043698	0
	0.21		0.043761	0.043798	0.4
	1/3		0.09	0.09	0.6
I	1/3		0.05	0.05	0.4
	1/3		0.04	0.04	0

Table 32: Inverse optimization results on 10 simulations

Feeding the true optimum B as x^0 to the inverse, equilibrium returns come back as r_{opt}=[0.09; 0.05; 0.05]. As explained in the previous section, this is a border solution. Solving the primal (95) with these returns does not return [0.6;0.4;0] as unique optimum. In order to get it back as optimum, we have to move the returns by ϵ further to r^{ϵ}_{opt}=[0.09; 0.05; 0.05001] does return x^*=[0.6;0.4;0] back.

Suboptimal corner A generates a border solution too, which moved by ϵ generates returns which optimize back to the same corner of weights.

Point on edge G. Optimizing back with r_{opt} instead of r^{ϵ}_{opt} would bring back original point on edge, but the result is highly unstable. It depends on the implementation of the interior point algorithm whether the optimum is the same original point or not. r^{ϵ}_{opt} optimizes to the suboptimal corner.

Starting with infeasible point H we first get to point on edge x^1, then as above; r^{ϵ}_{opt} optimizes to the suboptimal corner.

Feasible point F generates r^{ϵ}_{opt}, which optimize to a true optimum - a good result because F was very close to it.

Another feasible point I, not as close to true optimum, also finds it back.

Overall results on this very little example where corners and edges are identifiable easily are very satisfactory. Let us increase the problem to a realistic size. We will not be able to say whether the point is a corner, is on the edge or inside the feasible region. Counting the active constraints is helpful only up to certain level, due to degeneracy of the problem.

Now the problem is built with 100 000 simulations. We maximize returns with the constraint 99%-CVaR $\leq$ 0.55, epsilon=0.01. Table 33 displays the results.

	x^0	x^1	r_{opt}	r^{ϵ}_{opt}	x^*
	0.43026		0.09000	0.09000	0.43026
B	0.43018		0.04338	0.04328	0.43018
	0.13956		0.04000	0.04000	0.13956
A					
G					
	1	0.44635	0.06786	0.06764	0.08041
H	0	0.10731	0.05000	0.05000	0.45593
	0	0.44635	0.04817	0.04825	0.46365
	0.6	0.44179	0.09000	0.09000	0.44517
H’	0.2	0.2	0.04329	0.04322	0.12631
	0.2	0.35820	0.04000	0.04000	0.42851
	0.4		0.09000	0.09000	0.43026
F	0.4		0.04338	0.04328	0.43018
	0.2		0.04000	0.04000	0.13956
	1/3		0.09	0.09	0.43026
I	1/3		0.05	0.05	0.43018
	1/3		0.04	0.04	0.13956

Table 33: Inverse optimization results on 100 000 simulations

Optimum B, just like in the small example, generates a border solution, which moved by ϵ generates returns which optimize back to the original optimum B.

Suboptimal extreme point and point on the edge are hard to identify. Count of active constraints can be inconclusive. Most reliable way to get one of these points is to start with infeasible one and use distance minimization procedure. Other than suboptimal extreme point and point on the edge it can only be the true optimum, which is a good case too.

Starting with infeasible point H we do get on the border of feasible region. Here we have some problem, as border point x^1 is not very close to x^*, the point to which r^{ϵ}_{opt}-problem optimizes. Another infeasible point H' looks better, as x^1 is closer to x^*.

Feasible points F and I generate r^{ϵ}_{opt} which optimize to true optimum.

There is clearly some problem with finding the optimal cost vector for nearest feasible points to infeasible points, as the goodness criteria that $x^0(x^1)$ would be close to x^* does not always hold well (H).

The sensitivity issues listed in section 5, were not investigated, except for the observation that on 3 obligors, 10 scenarios example we had $\mathbf{x^0} = [0.6; 0.4; 0]$. The same optimum is achieved back as long as the first return is the highest; the second is in the middle, and the third is the last.