Recent approaches to portfolio optimisation: Part 1

"Near the end of his reign in 14 AD, the Roman emperor Augustus could boast that he had found Rome a city of brick and left it a city of marble. Markowitz can boast that he found the field of finance awash in the imprecision of English and left it with the scientific precision and insight made possible only by mathematics."
Mark Rubinstein (2002)

The year 2012 will mark sixty years since Harry Markowitz published his groundbreaking work "Portfolio Selection".[1] We are taking advantage of the occasion to begin a series on portfolio optimisation. We will present the approach itself and its progression over time, but also look at more recent portfolio construction methods. All of them are based on Markowitz's risk-return paradigm. Due to certain factors, not the least of which being the market turbulence of the past decade, minimum variance portfolios have experienced a renaissance.

Markowitz's novel realisation was that the risk-return profile of individual assets cannot be viewed separately, but must be considered in their portfolio context.[2] According to Markowitz, a portfolio is efficient if it exhibits either the lowest possible risk for a given expected return or the highest expected return at a given level of portfolio risk. Although the two ways of approaching the issue are equivalent in their end result, the optimisation procedures for determining portfolio weightings are substantially different. The former presents a quadratic optimisation problem subject to linear constraints; the latter a linear objective function with quadratic constraints.

The mathematics of Markowitz

Here, we assume N infinitely divisible assets, the returns of which are multivariate normally distributed. The portfolio return $\bar{r}$ is then the dot product of the ($N \times 1$) weight and return vectors ω and μ. The portfolio risk is defined by the portfolio variance $\sigma_W^2 = \omega'\Sigma\omega$, where Σ stands for the positive semi-definite variance-covariance matrix.

The optimisation method for minimum variance portfolios with a given portfolio return $\bar{r}$ is then:

$$(1) \quad \arg\min_{\omega} \sigma_W^2 = \omega'\Sigma\omega \quad \text{where } \omega'\mu = \bar{r}, \quad \omega' i = 1$$

in which i stands for the ($N \times 1$) one vector.

Already in the same year when Markowitz first published his article, Roy (1952) demonstrated how the approach could be used to derive an efficient portfolio, though the work of Merton (1972) is often cited in this respect. Accordingly, the weight vector for a minimum variance portfolio with a given return is:[3]

$$(2) \quad \omega^* = \bar{r}\omega_0^* + \omega_1^* \quad \text{where } \omega_0^* = \tfrac{1}{d}\left(c\Sigma^{-1}\mu - b\Sigma^{-1}i\right), \quad \omega_1^* = \tfrac{1}{d}\left(b\Sigma^{-1}\mu - a\Sigma^{-1}i\right)$$

Portfolio risk can be determined using

$$(3) \quad \sigma = \sqrt{\tfrac{1}{d}\left(c\bar{r}^2 - 2b\bar{r} + a\right)}$$

with $a = \mu'\Sigma^{-1}\mu,\ b = \mu'\Sigma^{-1}i,\ c = i'\Sigma^{-1}i$ and $d = ac - b^2$

As per equation (2), the portfolio weights constitute a linear function of the expected returns. It can also be shown that each efficient portfolio can be expressed as a linear combination of two efficient portfolios – and thus as a linear combination of a global minimum variance (GMV) portfolio and another random efficient portfolio. The covariance between the two equals the variance of the GMV portfolio. The only restriction applicable to the portfolio weightings is that their sum must be one. Negative weights (short positions) and weights greater than one (leveraged positions) are thus permissible.

Equation (3) describes the so-called efficiency frontier (figure 1), which is enclosed by the asymptotes $\bar{r} = b/c \pm \sqrt{d/c}\,\sigma$. At the vertex of this hyperbola is the GMV portfolio with the weight vector $\omega_{GMV}^* = \Sigma^{-1}i / i'\Sigma^{-1}i$, which (unlike the remaining mean variance efficient portfolios) is independent of the expected returns of the assets.

All mean variance efficient portfolios are situated at the upper branch. The marginal risk contributions of the individual assets in unrestricted minimum variance portfolios are all equally large, and their weights equal their contributions to overall risk. This results in Pareto optimal efficiency, which is intuitively plausible: If there were different marginal risk contributions, one could reduce total risk. This, however, would violate the assumption of variance minimality. The risk-return points within the hyperbola are also possible, though suboptimal. After all, there are portfolios with higher returns at the same level of risk or equivalent returns with lower risk, which are of more utility to investors – namely those on the upper branch of the hyperbola.

Expansion of the Markowitz model

Up to this point, the assumption has been of an investor with 100% of the portfolio dedicated to risk securities. We will now depart from the Markowitz model itself and introduce a risk-free investment with a return of r_f, the optimum proportion of which in the portfolio is determined by the investor's risk aversion. No matter the risk appetite, every investor aims to maximise wealth (i.e. the expected utility to the investor) at the end of the period, for investment decisions made at the start of the period.[4]

$$(4) \quad \max \mathrm{E}\left[U\left(W_{t+1}\right)\right]$$

Here, E stands for the expected value operator. The utility function can be approximated using a Taylor series, and it is assumed to be twice differentiable. Adding a neutral term $U(W_{t+1}) = U(W_{t+1} + \mathrm{E}[W_{t+1}] - \mathrm{E}[W_{t+1}])$, the utility function to be maximised can be stated as:

$$(5)\quad \mathrm{E}\left[U(W_{t+1})\right] = U\left(\mathrm{E}[W_{t+1}]\right) + \frac{U'\left(\mathrm{E}[W_{t+1}]\right)}{1!}\mathrm{E}\left[W_{t+1} - \mathrm{E}[W_{t+1}]\right] + \frac{U''\left(\mathrm{E}[W_{t+1}]\right)}{2!}\mathrm{E}\left[W_{t+1} - \mathrm{E}[W_{t+1}]\right]^2 + \sum_{i=3}^{\infty}\frac{U^{(i)}\left(\mathrm{E}[W_{t+1}]\right)}{i!}\mathrm{E}\left[W_{t+1} - \mathrm{E}[W_{t+1}]\right]^i$$

This does not imply any assumption with respect to the distribution function, so that equation (5) is defined to apply generally. Utility, however, is also dependent on the higher order moments of the portfolio. If the final wealth is normally distributed, the formula simplifies to:

$$(6)\quad \mathrm{E}\left[U(W_{t+1})\right] = U\left(\mathrm{E}[W_{t+1}]\right) + \frac{U'\left(\mathrm{E}[W_{t+1}]\right)}{1!}\mathrm{E}\left[W_{t+1} - \mathrm{E}[W_{t+1}]\right] + \frac{U''\left(\mathrm{E}[W_{t+1}]\right)}{2!}\mathrm{E}\left[W_{t+1} - \mathrm{E}[W_{t+1}]\right]^2$$

The utility to the investor then depends on the first two moments only. In the case of a quadratic utility function of $U(W) = W - \frac{\lambda}{2}W^2$, the weight vector is $\omega_U = (1/\lambda)\Sigma^{-1}\mu$. The higher the risk aversion, the smaller the weight sum.[5] The expected total return R from the risk-free and risk assets is then:

$$(7)\quad \mathrm{E}[R] = (1-\gamma)r_f + \gamma\bar{r} = r_f + \gamma(\bar{r} - r_f)$$

Here, $\gamma = i'\omega$ is the total proportion of the risk assets in the portfolio. The standard deviation of the portfolio is $\sigma(R) = \gamma\sigma_W$. This results in the capital market line – a linear function in the (μ,σ)-level:

$$(8)\quad \mathrm{E}[R] = r_f + \frac{\bar{r} - r_f}{\sigma_W}\sigma(R)$$

The optimum portfolio lies at the tangent point between these lines and the efficiency frontier, precisely where the slope – and thus the Sharpe ratio – is highest. This point on the efficiency frontier is thus also referred to as the maximum Sharpe ratio (MSR) portfolio. With the MSR, the investor holds only risk securities, and the marginal contributions of the individual assets to the Sharpe ratio are all equal in size. The degree of investment is reflected in the tangent point between the utility function and the capital market line. This lies south-west of the MSR portfolio and moves closer to the y-axis as risk aversion increases.

Estimation error: What now?

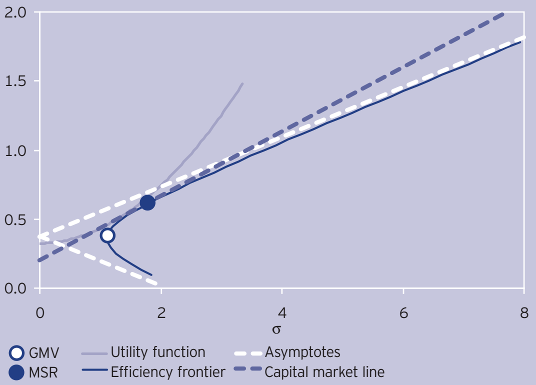
Unfortunately, these theoretical portfolio concepts can only be implemented to a certain degree in practice, as they are based on population moments that are not known. These moments must be expressed using estimates, which means that the achievable portfolios are situated below the efficiency frontier.

At first glance, the sample mean and unbiased variance estimator would appear to provide the best estimation function for the variance-covariance matrix of the returns. It must be noted, however, that estimation errors for these two moments will have a direct influence on the portfolio weightings. As a rule, this will preclude portfolio efficiency and/or minimum variance.

Ultimately, estimation errors result in higher portfolio risks. All other factors being equal, estimation of both the expected value of returns and the variance-covariance matrix is likely to result in more significant errors than in the case of approaches requiring estimates of return dispersion only.[6] Consequently, greater estimation errors can be expected for mean variance portfolios than for minimum variance portfolios.

Simulations demonstrate a strong divergence and erratic development of the portfolio weightings of mean variance portfolios, which is naturally undesirable. This is less distinct with minimum variance portfolios, making this approach more favourable from the standpoint of estimation accuracy.[7]

Figure 1: Efficiency frontier, GMV and MSR portfolio



Source: Invesco. For illustrative purposes only.

The strong reaction of the optimised portfolio to small changes in expected return should not be viewed as a flaw in the Markowitz approach, but rather a consequence of quadratic optimisation.

Monte Carlo simulations ...

One possible solution would be heuristic treatment of the estimation errors for expected value and the variance-covariance matrix using a Monte Carlo simulation.[8] This would involve first calculating the estimated values $\hat{\mu}_0$ and $\hat{\Sigma}_0$ for the theoretical moments μ and Σ, for a given sample size T, along with the efficiency frontier for m points. Next, K random data sets are generated for the dimension ($T \times N$) and the sample moments determined for these randomised data sets, creating K value pairs ($\hat{\mu}_i$, $\hat{\Sigma}_i$) with $i = 1, ..., K$. These are then used to determine m efficiency points. The simulated efficiency lines all lie below the efficiency frontier for $\hat{\mu}_0$ and $\hat{\Sigma}_0$.

The above procedure can be repeated to evaluate the estimation error for the expected values, using the value pairs ($\hat{\mu}_i$, $\hat{\Sigma}_0$) with $i = 1, ..., K$ – thus limiting estimation errors to the expected values. In the same way, ($\hat{\mu}_0$, $\hat{\Sigma}_i$) with $i = 1, ..., K$ can be applied to determine the estimation error for the variance-covariance matrix.

By averaging the m and K weights, randomised efficient portfolios can be determined. This approach, however, has its problems. The main argument against it is that the estimation errors are propagated – because the initial estimates of ($\hat{\mu}_0$, $\hat{\Sigma}_0$), on which the Monte Carlo simulations are based, are also subject to uncertainty. This type of randomised portfolio also lacks intuitive solutions for constrained optimisation.[9]

Moreover, the assumption of multivariate normally distributed returns is not usually valid in practice. The estimation error is thus compounded by a modelling error, as a non-stationary return process is being modelled by a stationary method. Consequently, a decision must be made between the two error sources: when using a larger sample, the estimation error is likely to be less significant, but coupled with a more substantial modelling error, and vice versa.

... or portfolio weighting restrictions?

Another conceivable solution would be to limit the effects of estimation errors by restricting portfolio weightings, as seen in the following example based on two independent asset investments A and B with equal expected returns of 3% and volatilities of 10%: An investor aiming to maximise returns would then be indifferent to any particular linear combination of the two assets. An estimation error of even one basis point, however, would have a substantial influence on portfolio allocation. For estimation values $\hat{\mu}_A = 3.01$ and $\hat{\mu}_A = 2.99$, this implies an infinite long position for investment A, financed by an equivalent short position of investment B. The solution would be completely different, however, in the case of long-only restrictions or limits on maximum portfolio weightings. The result would be improved out-of-sample performance or lower portfolio risk.[10]

Both can be explained by the lower implicit estimation error in the restricted case. Binding restrictions result in portfolio solutions that lie below the efficiency frontier for non-restricted portfolios – the greater the deviation of the restrictions from the optimum unrestricted weights, the farther below they lie. An investor, however, generally desires a portfolio situated as near as possible to the efficiency frontier. And, this is not the only reason that long-only restrictions are undesirable. Hedge funds and hedging strategies, for instance, generally require short positions. Restrictions are not always a reliable solution to avoid the pitfalls in practical application of the Markowitz approach.

Summary and outlook

In this first instalment of our series on recent methods of portfolio optimisation, we looked at the Markowitz approach to the construction of mean variance portfolios and highlighted the difficulties that can be experienced in practical implementation. These range from excessive portfolio concentration to significant fluctuations in portfolio weightings.

Next time, we will examine further recent optimisation methods and Bayesian approaches to these problems. Alternative estimation methods for portfolio risk will be discussed, along with an alternative optimisation method, which will lead us to the issue of risk-based portfolio optimisation. Further topics will be hedging strategies, core-satellite portfolios, the Black-Litterman model (with copula-opinion pooling) as well as Liability Driven Investments (LDI) and Asset Liability Management (ALM).

Bernhard Pfaff, Portfolio Manager,
Invesco Quantitative Strategies

Notes:

1 Markowitz (1952).
2 Comprehensive textbook descriptions of portfolio and construction are, e.g., in Ingersoll (1987), Huang and Litzenberger (1988), Markowitz (1991), Elton et al. (2007) and Scherer (2010).
3 Equation (2) can be derived using a Lagrange multiplier under the constraints of a given porfolio return and a weight sum of one. A detailed explanation is provided by Merton (1972).

4 This explanation is founded on Huang and Litzenberger (1988).
5 Also in the case of non-normality, the optimisation approach can be derived via the quadratic utility function. The parameter λ is then a measure of investor risk aversion.
6 cf. Merton (1980), Chopra and Ziemba (1993).
7 cf. DeMiguel et al. (2007), Frahm (2008), Jagannathan and Ma (2003), Kempf and Memmel (2006), Ledoit and Wolf (2003)
8 This resampling approach was proposed by Michaud (1998). A comprehensive explanation and critique of this approach is contained in Scherer (2010), chapter 4.
9 cf. Scherer (2010), chapter 4.
10 Frost and Savarino (1988) or Gupta and Eichhorn (1998) as well as Jagannathan and Ma (2003).

Bibliography

- Chopra, V. and W. Ziemba (1993): The effect of errors in means, variances, and covariances on optimal portfolio choice, Journal of Portfolio Management 19, 6-11.
- DeMiguel, V., L. Garlappi, and R. Uppal (2007): Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, Review of Financial Studies 22(5), 1915-1953.
- Elton, E., M. Gruber, S. Brown, and W. Goetzmann (2007): Modern Portfolio Theory and Investment Analysis (7th ed.), New York, NY (John Wiley & Sons).
- Frahm, G. (2008): Linear statistical inference for global and local minimum variance portfolios, Statistical Papers 51(4), 789-812.
- Frost, P. and J. Savarino (1988), For better performance: Constrain portfolio weights, Journal of Portfolio Management 15 (1), 29-34.
- Gupta, F. and D. Eichhorn (1998), Handbook of Portfolio Management, Chapter Mean-Variance Optimization for Practioners of Asset Allocation, pp. 57-74. John Wiley & Sons.
- Huang, C. and R. Litzenberger (1988): Foundations for Financial Economics, Amsterdam (Elsevier Science Publishing).
- Ingersoll, J. (1987): Theory of Financial Decision Making, Savage, Maryland (Rowman & Littlefield).
- Jagannathan, R. and T. Ma (2003): Risk reduction in large portfolios: Why imposing wrong constraints helps, Journal of Finance 58, 1651-1683.
- Kempf, A. and C. Memmel (2006): Estimating the global minimum variance portfolio, Schmalenbach Business Review 58, 332-348.
- Ledoit, O. and M. Wolf (2003): Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, Journal of Empirical Finance 10, 603-621.
- Ledoit, O. and M. Wolf (2008): Robust preformance hypothesis testing with the sharpe ratio, Working Paper Series 320, University of Zürich, Institute for Empirical Research in Economics.
- Markowitz, H. (1952): Portfolio selection, The Journal of Finance 7(1), 77-91.
- Markowitz, H. (1991): Portfolio Selection: Efficient Diversification of Investments (2nd ed.), Cambridge, MA (Basil Blackwell).
- Merton, R. (1972): An analytical derivation of the efficient portfolio frontier, Journal of Financial and Quantitative Analysis 7, 1851-1872.
- Merton, R. (1980): On estimating the expected return on the market: An exploratory investigation, Journal of Financial Economics 8, 323-361.
- Michaud, R. (1998): Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation, New York (Oxford University Press).
- Roy, A. (1952): Safety first and the holding of assets, Econometrica 20, 431-449.
- Rubinstein, M. (2002): Markowitz's "portfolio selection": A fifty-year retrospective, The Journal of Finance 57(3), 1041-1045.
- Scherer, B. (2010): Portfolio Construction and Risk Budgeting (4th ed.), London (Risk Books).

In the first part of this series, we introduced readers to the Markowitz approach, pointing out potential problems with implementation. These include excessive portfolio risk concentrations and considerable volatility in portfolio weightings, which are largely attributable to the high sensitivity of mean and variance/covariance to outliers – i.e. their lack of robustness. In this article, we look at various robust estimators for the location and dispersion of samples, examine their usefulness under a Monte Carlo simulation and apply them in backtesting. Next month's article will then cover robust portfolio optimisation methods.

The arithmetic mean is sensitive to extreme values, and consequently often conveys a distorted view of most sample observations. Estimation of covariance between two random variables can also be substantially affected by a single extreme value pair. Therefore, methods are desirable which are relatively immune to outliers while delivering representative estimates at the same time. This is the domain of "robust" estimation methods.[1]

The outlier problem could be pragmatically countered by trimming (elimination of extreme data items) or Winsorisation[2] (setting all outliers to a specified quantile). Although both of these methods provide robust results, the underlying classification of sample values as outliers is, by its nature, subjective. It could also make sense to estimate the location parameter based on the median and dispersion derived from the mean absolute deviation.

It must also be noted that classification of a data point as an outlier or extreme value is always relative and depends on the underlying model and/or distribution assumptions. Assuming standard normal distribution, a value of 5 could easily be seen as an outlier. But, would this be the case if you used a Student's t-distribution with four degrees of freedom or a Cauchy distribution?

Selected robust estimators[3]

Breakdown point and relative efficiency

The most useful measure of an estimator's robustness is its breakdown point (BP). This indicates the maximum proportion of outliers that can be tolerated by an estimator – in other words: the percentage of outliers in the sample as of which the estimator functions can take on an arbitrarily large value ("breaks down"). By definition, the BP can have a value between 0 and 0.5.

The arithmetic mean has a BP=0, since the replacement of a single value in the sample allows the estimator to take on an arbitrary value. For the median, on the other hand, BP=0.5. The maximum value of the BP can be explained by the fact that an outlier share of more than 50% will distort the sample in such a way that conclusions can no longer be drawn as to the properties of the data pool.

A further estimator quality criterion is its relative efficiency. Here, the (asymptotic) variance is divided by the variance of an optimal estimator derived using a suitable estimation principle, strictly observing the distribution assumptions.

Below, we introduce various estimators for location and dispersion of samples, which according to these criteria are more robust than the less sophisticated sample estimators: mean and variance/covariance.

Robust M and MM estimators

As early as 1964, Huber introduced the class of M estimators. The name is loosely derived from the maximum likelihood method (ML). For the sake of clarity, the explanation is limited to univariate samples.

Please note: according to the ML principle, the parameter vector $\tilde{\theta}$ of an independent and identically distributed (iid) random sample $\{x_1, \ldots, x_n\}$ is determined in a manner that the sample most likely derives from the distribution $F(x, \theta)$ with the density function $f(\cdot)$. As a result of the assumed iid, the joint distribution is equal to the product of the marginal distributions to be maximised:

$$(1) \quad \tilde{\theta} = \arg\max\left(\prod_{i=1}^{n} f\left(x_i, \theta\right)\right)$$

As a rule, since the logarithm is a strict monotonic transformation, the negative log likelihood is minimised.

$$(2) \quad \tilde{\theta} = \arg\max\left(-\sum_{i=1}^{n} \log\left(f\left(x_i, \theta\right)\right)\right)$$

By analogy, the M estimators are defined as the minimum sum of a function $\rho(x, \theta\, \cdot)$:

$$(3) \quad \tilde{\theta} = \arg\max\left(\sum_{i=1}^{n} \rho\left(x_i, \theta\right)\right)$$

M estimators, in addition to the ML principle ($\rho(\cdot) = -\log f(x, \theta)$), also include the method of least squares. In this case, the function $\rho(\cdot)$ is the squared deviation. The function $\rho(\cdot)$ must be symmetrical and positive definite, as well as exhibiting a global minimum of zero. Moreover, it should deliver useful parameter estimates, irrespective of the accuracy of the model assumptions.

The difference between robust M estimators and the simple least squares or ML principle lies in the specification $\rho(\cdot)$. In the context of a robust least squares estimate, extreme observations receive a lower weighting and thus have less impact on the result. Two specifications for $\rho(\cdot)$ are frequently applied in practice: the Huber function and the bisquare function.

The class of a Huber function is determined as:

$$\rho_k(x) = \begin{cases} x^2 & \text{if } |x| \le k \\ 2k|x| - k^2 & \text{if } |x| > k \end{cases}$$

This function is squared in a central area surrounding k, but linear outside of this area. If the following is set: $k \to \infty$, $k \to 0$ the M estimators are identical with the arithmetic mean and the median.

Tukey's proposed bisquare function is defined as:

$$\rho_k(x) = \begin{cases} 1 - \left[1 - (x/k)^2\right]^3 & \text{if } |x| \le k \\ 1 & \text{if } |x| > k \end{cases}$$

The function is bounded for large absolute values of x. In the case of symmetrical distributions with excess kurtosis, it is preferable to the Huber function, due to the ability to completely eliminate the influence of outliers.

Instead of directly optimising equation (3), a solution is often achieved by setting the gradients $\psi(x) = \delta\rho(x, \theta)/\delta\theta$ to zero.

The MM estimator was introduced by Yohai (1987) and Yohai et al. (1991) as a way to implement robust regressions. Lopuhaã (1991, 1992) later adapted the estimator for multivariate data analysis. With this method, the dispersion is first estimated with an M estimator, based on which the location is estimated with another M estimator.

Estimators based on robust scaling
Next, we will introduce three estimators, derived from robust scaling: the minimum volume ellipsoid estimator (MVE), the minimum covariance determinant (MCD) and the S estimator (S).

First, we will look at a p-dimensional random variable $\mathbf{x} = (x_1, ..., x_p)' \in \Re^p$. Provided the random variable $\mathbf{x}$ is assumed to be jointly normally distributed, ($\mathbf{x}$ $N(\mu, \Sigma)$), the location vector is $\mu = E(\mathbf{x}) = (E(x_1), ..., E(x_p))'$ and the dispersion $\Sigma = \mathrm{Var}(\mathbf{x}) = E((\mathbf{x} - \mu)(\mathbf{x} - \mu)')$.

Distance is measured as the squared Mahalanobis distance between the sample and the location and dispersion parameters: $d(\mathbf{x}, \mu, \Sigma) = (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu)$. A trivial solution of zero can be achieved if the lowest eigenvalue of an estimate $\tilde{\Sigma}$ approaches zero. In order to rule out this trivial solution, $|\tilde{\Sigma}| = 1$ is necessary, where $|\cdot|$ represents the determinants of a matrix.

The three estimators above for $\tilde{\mu}$ and $\tilde{\Sigma}$ are based on optimisation of:

$$(4) \quad \arg\min \tilde{\sigma}\left(\mathbf{d}\left(X, \tilde{\mu}, \tilde{\Sigma}\right)\right)$$

here, X stands for the overall sample, $\mathbf{d}$ for the vector of the distance measure $\mathbf{d}(\mathbf{x}_i, \tilde{\mu}, \tilde{\Sigma})$ for $i = p+1, ..., N$ and $\tilde{\sigma}$ for a robust metric.

The MVE and MCD estimators come from Rousseeuw (1985) along with Rousseeuw and Leroy (1987). For the MVE estimator, the smallest point cloud is identified containing at least half of the observations. This is accomplished using the median for $\tilde{\sigma}$. The convergence rate of this estimator is only $N^{-1/3}$. For the MCD estimator, on the other hand, the robust data set $h>N/2$ is solved such that the determinant of the variance-covariance matrix is minimal. The location vector is then equal to the p arithmetic mean of the selected data, and the estimator for covariance is expanded by correction factors, so that the resulting estimator $\tilde{\Sigma}$ corresponds to the normal distribution model. The BP is highest for $h = \mathrm{int}[(N+p+1)/2]$, but can be selected for every value in the interval $[(N+p+1)/2, N]$.

Davies (1987) introduced the class of S estimators. The 'S' stands for scale. In the above-described optimisation approach, an M estimator is then used in line with

$$1/N \sum_{i=1}^{N} \rho(d_i) = \delta$$

for the dispersion $\tilde{\sigma}$, where $\rho(\cdot)$ must be a limited function and $\delta \in (0, 1)$.

The class of SD estimators
SD estimators (SDE) go back to Stahel (1981) and Donoho (1982). Under this method, the outlier ratio is determined using a projection of data matrix X and a vector $\mathbf{a} \in \Re^p$. For a set of observations $\mathbf{x}$ and robust location and dispersions estimator $\tilde{\mu}$ and $\tilde{\Sigma}$ relating to data matrix X is defined as:

$$(5) \quad t(\mathbf{x}, \mathbf{a}) = \frac{\mathbf{x}'\mathbf{a} - \tilde{\mu}(X\mathbf{a})}{\tilde{\sigma}(X\mathbf{a})}$$

The outlier ratio is the maximum of equation (5) across all vectors $\mathbf{a}$ with norm 1. This determines the weighting of X, which represent a non-incremental function of $t(\mathbf{x}_i)$.

The OGK estimator
One problem encountered with affine invariant robust estimators is non-convex optimisation. The

Table 1: Risks of mean-variance portfolios

Data generating process	Estimator	T = 60 Median	T = 60 Spread	T = 120 Median	T = 120 Spread	T = 240 Median	T = 240 Spread
Normal/Normal	COV	0.79	0.092	0.78	0.069	0.78	0.051
	MCD	0.83	0.171	0.79	0.100	0.78	0.062
	MVE	0.80	0.169	0.79	0.100	0.78	0.067
	M	0.79	0.131	0.78	0.089	0.78	0.059
	MM	0.79	0.096	0.78	0.071	0.78	0.051
	S	0.79	0.109	0.78	0.081	0.78	0.057
	SDE	0.79	0.145	0.78	0.099	0.77	0.068
	OGK	0.74	0.122	0.73	0.084	0.72	0.060
Normal/Student	COV	0.95	0.139	0.96	0.107	0.97	0.076
	MCD	0.83	0.171	0.79	0.100	0.78	0.062
	MVE	0.83	0.167	0.85	0.114	0.85	0.080
	M	0.83	0.143	0.82	0.103	0.83	0.075
	MM	0.87	0.116	0.87	0.085	0.87	0.061
	S	0.85	0.123	0.85	0.092	0.85	0.064
	SDE	0.80	0.153	0.81	0.102	0.81	0.071
	OGK	0.77	0.135	0.75	0.090	0.74	0.067
Student/Student	COV	0.95	0.160	0.97	0.125	0.98	0.098
	MCD	0.85	0.170	0.86	0.114	0.86	0.082
	MVE	0.84	0.175	0.86	0.114	0.87	0.084
	M	0.83	0.162	0.84	0.114	0.83	0.077
	MM	0.86	0.133	0.87	0.089	0.87	0.067
	S	0.86	0.147	0.87	0.098	0.87	0.067
	SDE	0.81	0.154	0.83	0.109	0.83	0.079
	OGK	0.75	0.139	0.74	0.101	0.73	0.067

Source: Invesco. For illustrative purposes only.

estimator proposed by Maronna and Zamar (2002) avoids this through robust covariance expressed as:

$$(6)\quad s_{jk} = \frac{1}{4}\left(\sigma\left[\frac{Y_j}{\sigma(Y_j)} + \frac{Y_k}{\sigma(Y_k)}\right]^2 - \sigma\left[\frac{Y_j}{\sigma(Y_j)} + \frac{Y_k}{\sigma(Y_k)}\right]^2\right)$$

This form of dispersion estimatation was originally presented by Gnanadesikan and Kettenring (1972). In that case, however, affine invariance was abandoned in favour of more comfortable calculation. If, however, equation (6) is applied directly to data matrix X, the variance-covariance matrix would no longer be positive definite. Thus, Maronna and Zamar (2002) proposed orthogonalisation of X, so that this is known as the orthogonalized Gnanadesikan/Kettenring estimator. Through application of a robust estimator σ, the paired covariances s_{jk} with $j = 1, ..., p$ and $k = 1, ..., p$ are also robust.

Empirical application

Simulation

In this section, we will begin with a simulation to examine whether the robust estimators for the location and dispersion parameters are suitable to estimate portfolio risk.[4]

In the context of a Monte Carlo study, we assumed three different data generating processes:

- a Gauss copula with normally distributed marginal distributions
- a Gauss copula with Student's-t distributed marginal distributions
- a Student's-t copula with Student's-t distributed marginal distributions

For the t distribution, five degrees of freedom are assumed.

For the first data generating process (DGP), the sample estimators for the mean and variances/covariances are the best linear unbiased estimators (BLUE). The other two models, on the other hand, take into account that distributions of real financial market returns exhibit greater kurtosis overall and are dependent on one another at the tails of the joint distribution (tail dependence).

With each of these three models, 1,000 data sets were generated for five variables with a sample size of 60, 120 and 240 observations. The dependence structure between the five variables were set as equally correlated with $\rho = 0.5$ and equal dispersion. Applying monthly data, the simulation would equate to a five, ten and twenty-year window.

The generated returns were then used for optimisation of mean-variance and minimum-variance portfolios. In addition to the non-robust sample estimators, the robust estimation methods MCD, MVE, M, MM, S, SDE and OGK were used as location and dispersion estimators.

The measure of estimator quality in this context was portfolio risk. Thus, we compared the median and dispersion of the standard deviations of the individual optimised portfolios with one another. The spread of portfolio standard deviations is defined as the difference between the upper and lower hinge, which for their part are defined as the medians of the observations above and below the median of the overall sample. Defined in this way, the spread itself is a robust measure of scale.

Next, we will look at the results of the mean-variance optimisations (table 1).

In the case of the Gauss copula with normally distributed tails, the mean portfolio risk using the sample estimators (with the exception of the robust OGK estimator) was at most equal to that of the other robust estimators. This also results in the smallest risk spread, and portfolio risk volatility sinks as the sample size increases. These results were to be expected, given that the sample estimators in this DGP have BLUE characteristics. All the more remarkable, however, is the simulation result for the OGK estimator. Even in the case of jointly normally distributed random variables, it results in a lower level of portfolio risk than the sample estimators. This reflects the sensitivity of the ML estimators to outliers, which can easily appear under the normal distribution model.

In the case of the Gauss copula with Student's-t marginal distributions, on the other hand, the level of portfolio risk increases rises substantially under

Table 2: Risks of minimum-variance portfolios

Data generating process	Estimator	T = 60		T = 120		T = 240	
		Median	Spread	Median	Spread	Median	Spread
Normal/Normal	COV	0.74	0.092	0.76	0.067	0.77	0.048
	MCD	0.75	0.154	0.76	0.094	0.77	0.060
	MVE	0.72	0.159	0.76	0.095	0.77	0.068
	M	0.73	0.126	0.75	0.085	0.77	0.059
	MM	0.73	0.094	0.76	0.066	0.76	0.051
	S	0.73	0.104	0.76	0.076	0.76	0.054
	SDE	0.70	0.137	0.74	0.096	0.76	0.063
	OGK	0.66	0.105	0.69	0.080	0.70	0.055
Normal/Student	COV	0.91	0.144	0.95	0.109	0.97	0.077
	MCD	0.75	0.154	0.76	0.094	0.77	0.060
	MVE	0.77	0.167	0.82	0.106	0.84	0.077
	M	0.76	0.149	0.80	0.102	0.81	0.072
	MM	0.82	0.120	0.85	0.090	0.87	0.058
	S	0.80	0.123	0.83	0.094	0.84	0.064
	SDE	0.72	0.136	0.78	0.101	0.80	0.071
	OGK	0.68	0.125	0.71	0.089	0.72	0.062
Student/Student	COV	0.91	0.170	0.95	0.126	0.97	0.092
	MCD	0.82	0.176	0.83	0.110	0.85	0.076
	MVE	0.79	0.171	0.83	0.124	0.85	0.085
	M	0.78	0.167	0.80	0.107	0.82	0.077
	MM	0.82	0.126	0.84	0.088	0.86	0.064
	S	0.82	0.140	0.84	0.096	0.86	0.067
	SDE	0.76	0.156	0.79	0.104	0.81	0.073
	OGK	0.67	0.132	0.69	0.086	0.71	0.063

Source: Invesco. For illustrative purposes only.

Statistic	S&P 500	FTSE 100	TOPIX	EuroStoxx
Minimum	-16.795	-12.863	-20.256	-17.735
Maximum	11.436	10.981	14.591	14.957
Median	1.295	1.048	-0.268	1.339
Standard deviation	4.322	4.196	5.434	5.263
Interquartile range	5.361	4.816	7.323	6.151
Kurtosis	1.203	0.447	0.559	1.015
Skewness	-0.673	-0.499	-0.014	-0.482

Table 3: Descriptive market return statistics

Source: Invesco. For illustrative purposes only.

the sample estimators. Here as well, the spread declines as the sample size grows. Portfolio risks are considerably lower across all methods when the robust estimators are used, and they are less volatile than with the ML estimators.

If one additionally assumes a Student's-t copula, a jump can be observed under the robust methods (with the exception of the OGK estimator). It is, however, substantially less pronounced than that which occurs with the ML estimators.

Under the OGK method, mean portfolio risk in all three cases is nearly independent from the distribution of the data pool: The differences between the Gauss copula and the Student's-t copula are negligible for all estimation functions. Any tail dependence thus has no significant impact on overall portfolio risk.

For minimum-variance optimisation (table 2), only the dispersion of the simulated random variables is taken into account, so that estimation errors for the expected values are, by definition, ruled out. As expected, the risk in this type of portfolio is lower than that of mean-variance portfolio. Qualitatively, however, the results are similar. If robust estimators are used, portfolio risks are generally no greater than when jointly normally distributed returns are assumed – and in many cases actually less. And, portfolio risks are largely independent of DGP. Risks (and their dispersion) are lowest with the OGK method. With robust estimators, the assumption of Student's-t distributed tails has substantially less impact on the results than when using empirical variance/covariance.

Backtest

Below, in order to further highlight the differences between robust estimators and sample estimators, we will show the results of a comparative portfolio simulation.

The mean-variance optimisations were conducted under the secondary conditions of non-negativity and a 100% investment rate. The monthly returns for the equity indices S&P 500, FTSE 100, EuroStoxx and TOPIX were used for this purpose. Currency risks were not taken into account. The simulation covers the period from 31 January 1991 to 29 April 2011.

Table 3 shows the selected, descriptive statistics for these return series over the entire simulation period. The returns are characterised by a negative skew and excess kurtosis. TOPIX returns exhibit both the largest spread and the largest standard deviation, as well as the most significant interquartile range (IQR). The S&P 500 returns have the greatest kurtosis. The least volatile returns are in FTSE 100. Whereas most of the returns have a positive median, that of the TOPIX is negative.

The estimates for the mean and the variance-covariance matrix were arrived at using a moving 120-month window. Then, the portfolio values were calculated using the inner products of the weighting vectors and the following month's returns. The performance indicators used for the simulated portfolios were the information ratio (IR), the Sharpe and the ES-Sharpe ratio at a 95% confidence level, as well as the maximum accumulated loss and the value of the utility function $U = (\omega'\hat{\mu} - \mu_{EQW}) - 1/2\omega'\Sigma\omega'$. For the Sharpe ratios, the returns of an equally weighted portfolio (naïve allocation) were used as a benchmark.

Table 4 displays the results. All indicators are better under the robust estimators. In particular, MVE, M, SDE und OGK estimators generate considerably better results than the sample estimators (COV).

Figure 1 shows the excess returns of the robust portfolio compared to portfolios constructed on the basis of the non-robust ML estimators; table 5 contains selected, descriptive statistics in this regard.

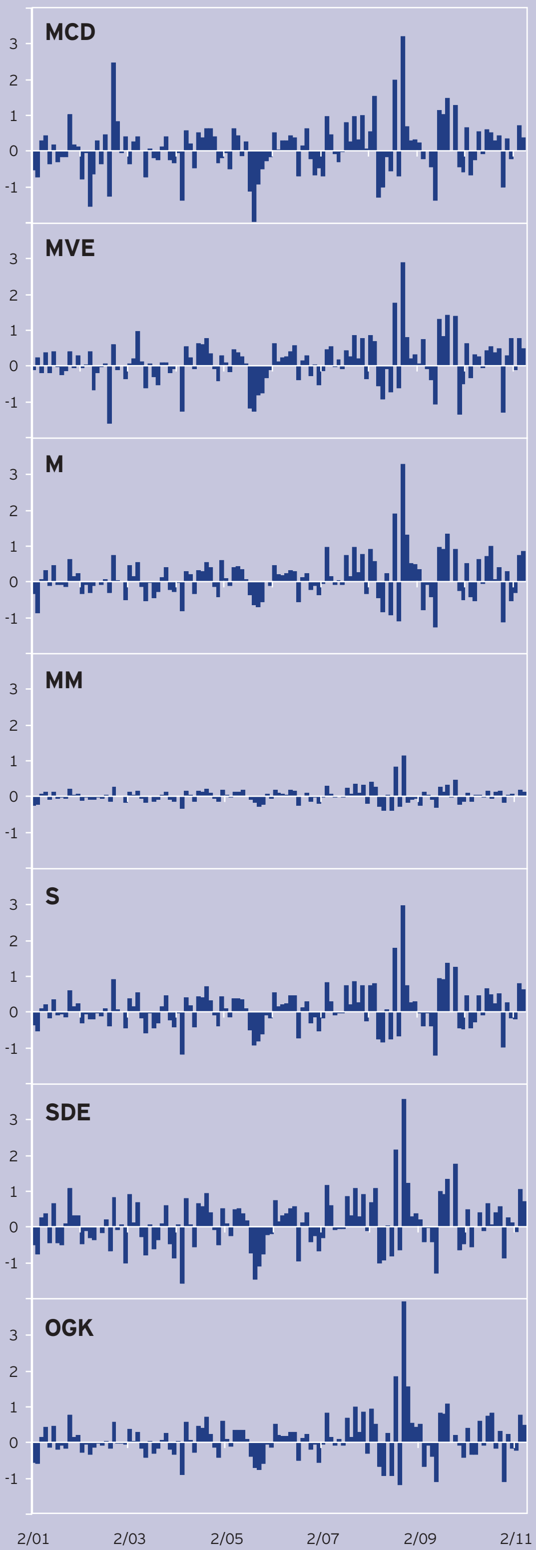
In all cases, the median is positive. Moreover, there is a significant positive skew of the excess returns.

Table 4: Simulated performance indicators

Statistic	Information ratio	Sharpe ratio	Sharpe ES	Maximum drawdown	Certainty equivalent
COV	-0.139	-0.026	0.020	46.597	-0.033
MCD	0.139	0.032	0.029	42.369	0.040
MVE	0.245	0.055	0.031	43.111	0.067
M	0.244	0.058	0.032	42.094	0.071
MM	-0.036	-0.007	0.023	45.600	-0.008
S	0.198	0.045	0.030	43.015	0.054
SDE	0.222	0.054	0.032	41.831	0.067
OGK	0.245	0.057	0.032	41.585	0.069

Based on data from 31 January 1991 to 29 April 2011.
Source: Invesco. For illustrative purposes only.

Figure 1: Simulated excess returns of robust estimators



Source: Invesco. For illustrative purposes only.

Table 5: Simulated relative performance of robust estimators

Statistic	Minimum	Lower hinge	Median	Upper hinge	Maximum
MCD	-1.964	-0.343	0.054	0.439	3.210
MVE	-1.595	-0.188	0.095	0.428	2.883
M	-1.259	-0.285	0.056	0.404	3.281
MM	-0.393	-0.079	0.009	0.122	1.148
S	-1.209	-0.260	0.058	0.410	2.990
SDE	-1.572	-0.419	0.074	0.524	3.564
OGK	-1.171	-0.193	0.037	0.383	3.954

Based on data from 31 January 1991 to 29 April 2011.
Source: Invesco. For illustrative purposes only.

The relative frequency of outperformance, however, is not significantly different from 50%.

Summary and outlook

In this article, we have examined robust estimation methods for the moments of random variables, which should continue to deliver plausible values even if the model assumptions of optimal estimators are breached, e.g. as a result of outliers/extreme observations. In our Monte Carlo study, it can be seen that robust estimators can generate better results compared to conventional sample estimators, even in conjunction with multivariate, normally distributed random variables. In the context of return distributions with excess kurtosis, the robust methods result in substantially lower levels of simulated portfolio risks. With less dispersion on average, the estimators also function largely independent of distribution assumptions. A simulation using actual market data also points to the advantages of robust estimators.

After our discussion of robust estimators in the above article, the next part in this series will look at robust optimisation methods. These are set apart from the above approach by the fact that the parameters of the optimisation methods may vary within predefined limits.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Asset Allocation

Notes:

1 Textbooks on this topic include Huber (1981), Hampel et al. (1986), Rousseeuw and Leroy (1987), Staudte and Sheather (1990) and Maronna et al. (2006).

2 Named for Charles P. Winsor.

3 This presentation of robust estimation methods and their characteristics is oriented on Maronna et al. (2006) as well as Todorov and Filzmoser (2009).

4 All calculations were conducted using the free statistical program R 2.13.0 (see R Development Core Team, 2011). Additionally, the CRAN packages Hmisc (Harrell, 2010), copula (Yan, 2007; Kojadinovic and Yan, 2010), quadprog (Turlach and Weingessel, 2010), rrcov (Todorov and Fikzmoser, 2009) and timeSeries (Wuertz and Chalabi, 2011) were used.

Bibliography

- Davies, P. (1987), Asymptotic behavior of s-estimators of multivariate location parameters and dispersion matrices, The Annals of Statistics, 15, 1269–1292.
- Donoho, D. (1982), Breakdown properties of multivariate location estimators, Technical report, Boston: Harvard University.
- Gnanadesikan, R. and J. Kettenring (1972), Robust estimates, residuals and outlier detection with multiresponse data, Biometrics, 28, 81–124.
- Hampel, F. R., E. M. Rochetti, P. J. Rousseeuw, and W. A. Stahel (1986), Robust Statistics, The Approach Based on Influence Functions, New York: John Wiley & Sons.
- Harrell, F. (2010), Hmisc: Harrell Miscellaneous, R package version 3.8-3.
- Huber, P. (1964), Robust estimation of a location parameter, Annals of Mathematical Statistics, 35, 73–101.
- Huber, P. J. (1981), Robust Statistics, New York: John Wiley & Sons.
- Kojadinovic, I. and J. Yan (2010), Modeling multivariate distributions with continuous margins using the copula r package, Journal of Statistical Software, 34(9), 1–20.
- Lopuhaã, H. (1991), Multivariate ô-estimators for location and scatter. Canadian Journal of Statistics, 19, 307–321.
- Lopuhaã, H. (1992), Highly efficient estimators of multivariate location with high breakdown point, The Annals of Statistics, 20, 398–413.
- Maronna, R., D. Martin, and V. Yohai (2006), Robust Statistics: Theory and Methods, New York: John Wiley & Sons.
- Maronna, R. and R. Zamar (2002), Robust estimates of location and dispersion of high-dimensional datasets, Technometrics, 44(4), 307–317.
- Pfaff, B. (2010), Modelling Financial Risks: Fat Tails, Volatility Clustering and Copulae, Frankfurt am Main: Frankfurt Allgemeine Buch.
- R Development Core Team (2011), R: A Language and Environment for Statistical Computing, Vienna: R Foundation for Statistical Computing.
- Rousseeuw, P. (1985), Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, and W. Wertz (Eds.), Mathematical Statistics and Applications, Volume B, Dordrecht: Reidel Publishing, 283–297.
- Rousseeuw, P. and A. Leroy (1987), Robust Regression and Outlier Detection, New York: John Wiley & Sons.
- Stahel, W. (1981), Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen, Ph. D. thesis, Swiss Federal Institute of Technology (ETH), Zurich.
- Staudte, R. G. and S. J. Sheather (1990), Robust Estimation and Testing. New York: John Wiley & Sons.
- Todorov, V. and P. Filzmoser (2009), An object oriented framework for robust multivatiate analysis. Journal of Statistical Software, 32(3), 1–47.
- Turlach, B. and A. Weingessel (2010), quadprog: Functions to solve Quadratic Programming Problems, R package version 1.5-3.
- Wuertz, D. and Chalabi (2011), timeSeries: Rmetrics – Financial Time Series Objects, R package version 2130.92.
- Yan, J. (2007), Enjoy the joy of copulas: With a package copula, Journal of Statistical Software, 21(4), 1–21.
- Yohai, V. (1987), High breakdown-point and high efficiency estimates for regression, The Annals of Statistics, 15, 642–656.
- Yohai, V., W. Stahel, and R. Zamar (1991): A procedure for robust estimation and inference in linear regression. in W. Stahel and S. Weisberg (Eds.), Directions in Robust Statistics and Diagnostics (Part II), Volume 34, The IMA Volumes in Mathematics and its Applications, New York: Springer, 365–374.

Invesco

In the first two parts of this series, we assumed precise knowledge of the parameters used for portfolio optimisation (i.e. the means and variances of individual investments and their covariances) – regardless of whether they are estimated using conventional or robust methods. Although there are advantages to the use of robust methods, they cannot solve every problem. For instance, even a minimal change in parameters can result in a completely different portfolio. This is the point where robust optimisation takes hold.

Classical optimisation determines optimum portfolio weightings given predefined values for mean, variance and covariance. For robust optimisation, we dispense with such concrete assumptions. Instead, the portfolio weighting to be determined should be optimised for a multitude (band) of parameter values.

Thus, "robustness" in this case does not refer to resistance to outliers, but that the weighting is optimum for other parameters as well. In this way, robust optimisation is able to account for uncertainties about actual parameter values.[1, 2]

Robust optimisation: the concept in detail

Below, we will discuss the concept of robust optimisation using the example of the Markowitz approach. In the first article of this series, we formalised this approach as follows:[3]

$$(1)\quad P_\lambda = \arg\min_{\omega\in\Omega}(1-\lambda)\sqrt{\omega'\Sigma\omega} - \lambda\omega'\mu$$

in which ω refers to the $n\times 1$ portfolio weighting vector and $\Omega \subset \left\{\omega \in \mathbb{R}^N \middle| \omega'\mathbf{1} = 1\right\}$ to the number of permissible solutions. The vector μ stands for the (expected) returns of the N assets. $\Sigma \in \mathbb{R}^{N\times N}$ designates the (positive definite) covariance matrix. The parameter λ represents the weighting of portfolio return and risk; it can assume all values in the interval [0,1]. In addition to the restriction of a weighting sum of one ($\omega'\mathbf{1} = 1$), conditions of non-negativity ($\omega \geq 0$) and/or interval limits ($A\omega \leq \mathbf{b}$) may be assumed for the permissible weightings.

The extreme solutions of this equation include a minimum variance portfolio (if $\lambda = 0$) and a maximum return portfolio (if $\lambda = 1$). All interim values for λ result in portfolio solutions along the efficiency curve.

In the two previous articles point estimation was used for the unknown parameters (μ, Σ). Simulation studies and backtesting confirmed that the results can be improved using robust estimators. However, one key problem remained unresolved: even minimal deviations between the estimated values and the true (unknown) parameters can result in a completely different portfolio. It is thus desirable to directly account for this parameter uncertainty when formulating the problem. Because it has been shown that the optimum weighting vector is more dependent on returns than on the dispersion matrix, we will limit this discussion to the use of robust optimisation in the context of uncertainty about returns.

What is uncertainty?

So far, the term "uncertainty" has not been precisely defined. For robust optimisation, however, this must be rectified. Thus, we will specify an uncertainty set $U_{\tilde{\mu}}$ for the possible parameter values.

Possible specifications are:

$$(2a)\quad U_{\tilde{\mu}} = \left\{\mu \in \mathbb{R}^N \middle| \tilde{\mu}_1,\ldots,\tilde{\mu}_i,\ldots,\tilde{\mu}_M, i = 1,\ldots,M\right\}$$

$$(2b)\quad U_{\tilde{\mu}} = \left\{\mu \in \mathbb{R}^N \middle| \left|\tilde{\mu}_i - \mu_i\right| \leq \delta_i, i = 1,\ldots,N\right\}$$

$$(2c)\quad U_{\tilde{\mu}} = \left\{\mu \in \mathbb{R}^N \middle| \left(\mu - \tilde{\mu}\right)' \tilde{\Sigma}^{-1}\left(\mu - \tilde{\mu}\right) \leq \delta^2/T\right\}$$

Under Specification 1 (equation 2a), there are M possible return vectors μ. These may be subjective return expectations or parameters defined using various estimation methods.

Under Specification 2 (equation 2b), the uncertainty set is defined using intervals. The interval boundaries may be subjectively fixed or derived using distribution assumptions.

If one assumes, for instance, normally distributed returns, key fluctuation intervals with a confidence level of $(1-\alpha)$ can be determined in accordance with $\delta_i = \Phi^{-1}(\alpha/2)\cdot\tilde{\sigma}_i/\sqrt{T}$, where Φ^{-1} refers to the quantile function, $\tilde{\sigma}_i$ to the standard deviation in returns of i-th asset and T to the sample size. Another possible assumption is of Student's t-distribution, which accounts for the empirical occurrence of excess kurtosis in return processes. The next step would be to determine the number of degrees of freedom using ML estimation, so that the quantile value can then be determined for a given probability of error.

Under Specification 3 (equation 2c), an ellipsoidal uncertainty set is defined. Unlike equation (2b), it is now assumed that the returns follow a joint elliptical distribution pattern. Here, the covariances between the return processes are explicitly accounted for in the uncertainty set, in contrast to the interval probabilities under equation (2b), where any such dependencies between the uncertainty margins of the return processes were not taken into account.

Note too that a combination of Specification 1 and Specification 3 is also conceivable. An example would be an ellipsoidal uncertainty set around several point estimations rather than just one – for instance, around the mean values determined using classical and robust methods:

$$(3)\quad U_{est}=\left\{\mu\in\mathbb{R}^N \,\middle|\, (\mu-\bar{\mu})'\bar{\Sigma}^{-1}(\mu-\bar{\mu})\le\bar{\delta}^2\right\}$$

where

$$(4a)\quad \bar{\mu}=\frac{1}{|M|}\sum_{m\in M}m$$

$$(4b)\quad \bar{\Sigma}=\operatorname{diag}\left(\bar{\sigma}_{11},\dots,\bar{\sigma}_{NN}\right)\text{ with}$$

$$\bar{\sigma}_{ii}=\frac{1}{|M|-1}\sum_{m\in M}\left(m_i-\bar{\mu}_i\right)^2$$

$$(4c)\quad \bar{\delta}=\arg\max_{m\in M}(m-\bar{\mu})'\bar{\Sigma}^{-1}(m-\bar{\mu})$$

Optimum solutions

In this section, we will demonstrate how optimum portfolio solutions can be determined for the three different specifications of the uncertainty set. A worst-case approach is used for this purpose, meaning that we specify an optimum weighting vector for all three variants assuming the least advantageous parameter constellation. Put loosely: expect and prepare for the worst, then at least you will not be disappointed.

Specification 1: For an uncertainty set in accordance with equation (2a), first the weighting vectors were determined for M possible return vectors. Then, those were selected whose target function displayed the greatest value. This is equivalent to a portfolio with the lowest return.

Note that the solution can be very computationally intensive, especially when there are many possible return vectors. Moreover, it is assumed that all return vectors are equally probable. Ultimately, the least favourable scenario is of principal importance. Thus, it can be generalised that this is a particularly conservative approach, since the solution for optimum portfolio weighting is influenced by individual outliers. The defined return vectors are consequently of crucial significance.

Specification 2: Next, we will look at the uncertainty set described by (2b) through a symmetrical interval around the position parameter μ. The true return of the i-th asset at a given confidence level is within the limits defined as $\mu_i\in\left[\tilde{\mu}_i-\delta_i,\tilde{\mu}_i+\delta_i\right]$. The least favourable return of a long position in the i-th asset is thus $\mu_i=\tilde{\mu}_i-\delta_i$; the least favourable return of a short position is $\mu_i=\tilde{\mu}_i+\delta_i$.

These N confidence intervals form a polyhedron, which can be expressed as a linear inequality system. At first glance, however, it is not evident whether an asset is included in the portfolio with positive or negative weighting. To solve the problem, two slack variables ω_+ and ω_- are applied in the target function, resulting in the following:

$$(5)\quad PR_\lambda=\arg\max_{\omega,\omega_+,\omega_-}\omega'\tilde{\mu}-\delta'(\omega_+-\omega_-)-(1-\lambda)\sqrt{\omega'\Sigma\omega}$$

$$\omega=\omega_+-\omega_-$$

$$\omega_+\ge 0$$

$$\omega_-\ge 0$$

Naturally, the optimisation programme in (5) can include a budget restriction $\omega'\mathbf{1}=1$. For positive weightings $\omega_i>0$, these now become $(\tilde{\mu}_i-\delta_i)$, for negative weightings $(\tilde{\mu}_i+\delta_i)$. As a result of the condition of inequality, a positive weighting of the i-th element of ω_+ is positive and the i-th element of ω_- is zero; the opposite is true for a negative weighting.

An asset with an uncertain return estimate is weighted lower in absolute terms than one with a smaller estimation interval. Uncertainty with respect to expected returns may also be limited to a subset $U\subset N$. In this case, $\delta_i\notin U$ equals zero. If the objective is long-only optimisation, the mathematical programme is reduced to equation (1), where the return factors is $\tilde{\mu}-\delta$ with the auxiliary condition $\omega\ge 0$.

Specification 3: In the context of an ellipsoidal uncertainty set for the return vector (equation 2c), we likewise apply a worst-case scenario approach:

$$(6)\quad PR_\lambda=\arg\min_{\omega\in\Omega}\ \arg\max_{\mu\in U}(1-\lambda)\sqrt{\omega'\Sigma\omega}-\lambda(\omega'\mu)$$

If the returns are jointly normally distributed, then $(\mu-\tilde{\mu})'\Sigma^{-1}(\mu-\tilde{\mu})$ is χ^2-distributed with N degrees of freedom. At a given level of confidence $(1-\alpha)$, the relevant quantile value is then the scalar δ^2. The stochastic return vector μ thus lies within the ellipsis for the given level of confidence. Now, the maximum interval between this uncertainty ellipsis and the estimated position parameters is determined, which provides the least favourable return values. From equation (6) can be concluded that the uncertainty relates solely to the returns and that its covariance matrix can be assumed as given. As such, in a first step, the maximum interval can be determined using a Lagrange method, in which $\tilde{P}=\frac{1}{T}\tilde{\Sigma}$ represents the covariance matrix of the estimated returns and γ the Lagrange multiplier:

$$(7)\quad L(\mu,\gamma)=\omega'\tilde{\mu}-\omega'\mu-\left[\frac{\gamma}{2}(\tilde{\mu}-\mu)'\tilde{P}^{-1}(\tilde{\mu}-\mu)-\delta^2\right]$$

The optimal solution is determined through partial derivatives from equation (7) with respect to μ and γ, and setting to zero. This results in a system of equations with two unknowns, which is solved for μ:

$$(8) \quad \mu = \tilde{\mu} - \left(\frac{\delta}{\sqrt{\omega' \tilde{P} \omega}} P\omega \right)$$

Left multiplication by ω' thus results in the following for the portfolio return:

$$(9) \quad \omega'\mu = \omega'\tilde{\mu} - \delta\sqrt{\omega'\tilde{P}\omega} = \omega'\tilde{\mu} - \delta\left\| \tilde{P}^{\frac{1}{2}}\omega \right\|$$

Using the above formula demonstrates that the portfolio return in the case of robust optimisation with elliptical uncertainty is reduced by the term $\delta\sqrt{\omega' P\omega}$. The root of the quantile value in this context can be interpreted as a risk aversion parameter in relation to estimation accuracy. Applying the inner solution from equation (9) to the robust optimisation method from equation (6) thus results in:

$$
\begin{aligned}
(10) \quad PR_{\lambda} &= \arg\min_{\omega\in\Omega} \arg\max_{\mu\in U} (1-\lambda)\sqrt{\omega'\Sigma\omega} - \lambda(\omega'\mu) \\
&= \arg\min_{\omega\in\Omega} (1-\lambda)\sqrt{\omega'\Sigma\omega} - \lambda(\omega'\mu) + \\
&\qquad \lambda\frac{\delta}{\sqrt{T}}\sqrt{\omega'\tilde{\Sigma}\omega} \\
&= \arg\min_{\omega\in\Omega} \left(1-\lambda+\lambda\frac{\delta}{\sqrt{T}} \right) - \lambda\omega'\tilde{\mu}
\end{aligned}
$$

Equation (10) implies two statements:

First, that the efficiency line of a portfolio in the case of robust optimisation with ellipsoidal uncertainty of returns is identical to one reduction factor to the efficiency line without uncertainty. If the risk-return parameter in equation (10) is represented by θ, the equivalent trade-off parameter λ from equation (1) is:

$$(11) \quad \lambda := \frac{\theta}{1+\theta\frac{\delta}{\sqrt{T}}}$$

The maximum interval defined for $\theta \in [0,1]$ is now restricted for λ in the equivalent classical portfolio: $\lambda \in \left[0,1/\left(1+\delta/\sqrt{T}\right) \right]$.

At the same time, the weighting vector for a minimum variance portfolio $\lambda = 0$ is identical in both cases, since the uncertainty is assumed to be restricted to the location parameters. But, also in the case of an ellipsoidal uncertainty set, the robust solution is very conservative. The portfolio weightings were implicitly determined under the assumption of the least favourable return for all assets. Problem formulation in accordance with equation (10) can be described as a second-order cone optimisation, which can be solved using interior-point methods.

Empirical application

The described optimisation methods for an ellipsoidal uncertainty quantity will now be applied to a sample portfolio comprising two assets: US equities and bonds, represented by the S&P 500 and ten-year US Treasuries. For the portfolio optimisation, 20 monthly returns were used in each case, beginning on 30 June 2009. Table 1 depicts the estimates for the vector of the position parameters and the covariance matrix.[4]

The average equity return in the observation period is roughly five times as large as the average bond return. The variance of the equity return is also clearly larger. The returns of the two assets exhibit negative correlation with one another.

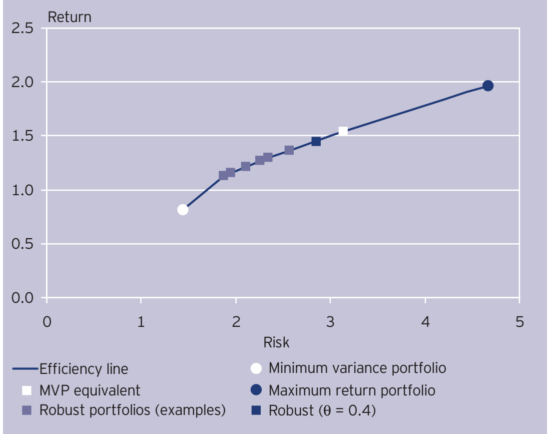
These values were used initially for a classical portfolio optimisation, in which a condition of non-negativity $I\omega \geq \mathbf{0}$ and a budget restriction $\omega'\mathbf{1} = 1$ were included. Figure 1 shows the efficiency line for this optimisation. The white circle represents the minimum variance portfolio, the dark blue circle the maximum return portfolio.

Table 1: Estimates for position and dispersion parameters

Assets	$\tilde{\mu}$	$\tilde{\Sigma}$	
Equities (S&P 500)	1.97	21.92	-5.56
Bonds (10y US Treasuries)	0.37	-5.56	5.02

Based on data from 30 June 2009 to 31 January 2011.
Source: Invesco. For illustrative purposes only.

Figure 1: Efficiency line and alternative optimum portfolios



Source: Invesco. For illustrative purposes only.

The light blue squares on the efficiency line are examples for optimum robust portfolios with different values of the risk aversion parameter θ. For the robust optimisations, a confidence level of $(1 - \alpha) = 0.9$ was assumed. The relevant quantile value is $\chi^2_{n=2,0,9} = 4.61$.

The optimum robust portfolio for $\theta = 0.4$ has been depicted as a dark blue square; the equivalent classical portfolio solution is the white square. The value of λ corresponding to $\theta = 0.4$ is $\lambda = 0.34$ according to equation (11). The weighting vector for this classically optimised portfolio is $\omega_P = [0, 72, 0, 28]$; the optimum robust solution is $\omega_{PR} = [0, 67, 0, 33]$.

Since mean estimates are more uncertain for equities than for bonds, the proportion of equities in the robust portfolio is lower. The price of robustness is thus a lower expected return.

Summary

In this article, we have examined so-called robust optimisation methods. Their goal is to find optimum portfolio weightings under unfavourable parameter constellations.

Because the robust solutions are based on a worst-case approach, the precise definition of uncertainty sets is of key importance. Extreme scenario parameters or intervals can have a considerable effect on portfolio structure.

This is much less likely when using ellipsoidal specifications. In this case, the portfolio is dependent on the given level of confidence, where the relevant quantile value can be interpreted as a risk aversion parameter in relation to estimation accuracy.

In the coming articles of the series, we will look at market risk-based optimisation methods. We will examine how a portfolio must be constructed to satisfy certain requirements such as minimum expected shortfall.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Asset Allocation

Notes:

1 Robust optimisation can also be distinguished from stochastic optimisation. Although the latter also takes into account parameter uncertainties, it generally entails explicit distribution assumptions. This is not a prerequisite for robust optimisation. Stochastic optimisation methods will be the subject of a later article.

2 Detailed descriptions of robust optimisation methods are available in publications such as: Ben-Tal and Nemirovski (1998), Cornuejols and Tütüncü (2007), Fabozzi et al. (2007), Meucci (2005, Chapter 9), Scherer (2010, Chapter 5) and Tüntücü and König (2004).

3 The present description is derived from Schöttle (2007, Chapter 4).

4 All calculations were conducted using the free statistical program environment R 2.13.1 (R Development Core Team, 2011) and the packages Rsocp (Chalabi and Würtz, 2010) and Hmisc (Harrell, 2010).

Bibliography

- Ben-Tal, A. and A. Nemirovski (1998). Robust convex optimization. Mathematics of Operations Research 23(4), 769–805.
- Chalabi, Y. and D. Würtz (2010). Rsocp: An R extenstion library to use SOCP from R. R package version 271.1/r4910.
- Cornuejols, G. and R. Tütüncü (2007). Optimization Methods in Finance. Cambridge: Cambridge University Press.
- Fabozzi, F., S. Focardi, P. Kolm, and D. Pachamanova (Eds.) (2007). Robust Portfolio Optimization and Management. New Jersey: John Wiley & Sons.
- Harrell, F. (2010). Hmisc: Harrell Miscellaneous. R package version 3.8-3.
- Meucci, A. (2005). Risk and Asset Allocation. New York: Springer.
- R Development Core Team (2011). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Scherer, B. (2010). Portfolio Construction and Risk Budgeting (fourth ed.). London: Risk Books.
- Schöttle, K. (2007). Robust Optimization with Application in Asset Management. Dissertation, Technische Universität München, Munich.
- Tüntücü, R. and M. König (2004). Robust asset allocation. Annals of Operation Reserach 132, 132–157.

New approaches to portfolio optimisation: Part 4

In the first three parts of this series, we constructed portfolios with various optimisation methods, and subsequently calculated their VaR. In this and the next quarter's installment, we will take just the opposite approach: constructing portfolios that will not breach a pre-defined risk of loss limit (measured as VaR or CVaR) at a certain level of probability. To this end, we have replaced the accustomed mean-variance optimisations with a mean-VaR or mean-CVaR optimisation.

The high volatility in the wake of the most recent financial crisis has led to a renaissance in minimum-variance portfolio strategies. But for many investors, the risk of loss is more important than portfolio variance – due in no small part to regulatory guidelines defining capital adequacy in relation to VaR, such as Solvency II in Europe and Basel III.

In this article, we demonstrate how efficient mean-VaR and mean-CVaR portfolios can be constructed, and illustrate the point in a simulation study.

Efficient mean-VaR portfolios

What does it take to make a mean-VaR portfolio? And, what relationship does such a portfolio bear with the traditional Markowitz approach? The answer can be found based on Alexander and Baptista (1999, 2002, 2004, 2008) and De Giorgi (2002).

It is assumed that the returns on N investment instruments in a portfolio are jointly normally distributed: $\mathbf{r} \sim N(\mu, \Sigma)$. For a given confidence level $\alpha \in (1/2, 1)$, the VaR of the portfolio is thus:

$$(1) \quad VaR_\alpha = z_\alpha \sigma_W - \bar{r}$$

where z_α stands for the quantile value of the normal distribution, σ_W for the standard deviation of the portfolio and $\bar{r}$ for the expected portfolio return. The standard deviation of the portfolio is $\sigma_W = \sqrt{\omega' \Sigma \omega}$ – the portfolio return is $\bar{\mathbf{r}} = \omega' \mathbf{r}$. ω is the $(N \times 1)$ weighting vector.

As with the Markowitz approach, efficient mean-VaR portfolios can now be constructed with either a maximised portfolio return for a given VaR_α, or a minimised VaR for a given portfolio return.

In the latter case, all weighting vectors $\bar{\omega} \in \Omega$ result in efficient portfolio allocations for solutions of the following optimisation problem:

$$(2) \quad P_{\text{VaR}_\alpha} = \arg\min \text{VaR}_\alpha = z_\alpha \sigma_W - \bar{r}$$
$$\omega' \mu = \bar{r}$$
$$\omega' \mathbf{i} = 1$$

where $\mathbf{i}$ represents the $(N \times 1)$ unit vector.

As with Markowitz, this constitutes a quadratic target function with linear constraints. Alexander and Baptista (2002) showed that efficient mean-VaR portfolios are always simultaneously mean-variance efficient, and given the efficiency line must also satisfy the conditions defined by Merton (1972) for portfolio efficiency:

$$(3) \quad \frac{\sigma_W^2}{1/C} - \frac{(\bar{r} - A/C)^2}{D/C^2} = 1$$

Here, A, B, C and D represent the scalars:

$$(4) \quad A = \mathbf{i}' \Sigma^{-1} \mu$$

$$(5) \quad B = \mu' \Sigma^{-1} \mu$$

$$(6) \quad C = \mathbf{i}' \Sigma^{-1} \mathbf{i}$$

$$(7) \quad D = BC - A^2$$

Rearranging equation (1) to σ_W and inserting it into equation (3) results in the following conditions for efficient mean-VaR portfolios:

$$(8) \quad \frac{\left[(\text{VaR}_\alpha + \bar{r})/z_\alpha\right]^2}{1/C} - \frac{(\bar{r} - A/C)^2}{D/C^2} = 1$$

The efficiency line is dependent on the confidence level α. Below, we describe the shape of these lines for the lowest and highest confidence levels in the range, i.e. for $\alpha \to 1/2$ and $\alpha \to 1$. In the mean-VaR diagram, the efficient frontier for $\alpha \to 1/2$ becomes a straight line with a slope of -1, since z_α approaches zero, rendering VaR equal to the negative expected portfolio return. For the other extreme, the efficient frontier is equivalent to that of an efficient mean-variance portfolio, because $z_\alpha \to \infty$ and thus also $\text{VaR}_\alpha / z_\alpha = \sigma_W$, as the term $\bar{r}/z_\alpha$ approaches zero. The generalised formula for mean-VaR portfolios is arrived at through reordering the terms of (8) as follows:

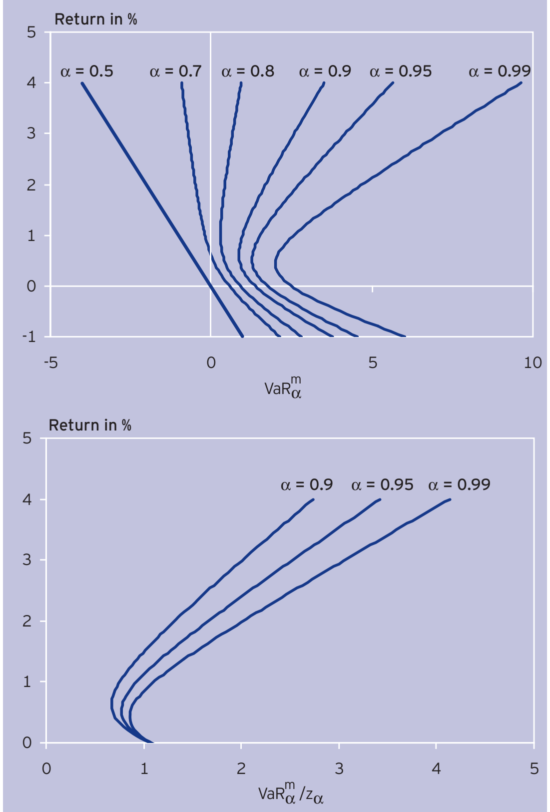
$$(9) \quad \text{VaR}_\alpha = -\bar{r} + z_\alpha \sqrt{\tfrac{1}{D}\left(C\bar{r}^2 - 2A\bar{r} + B\right)}$$

Equation (9) describes a band of efficiency lines dependent on confidence level α. The higher the confidence level, the further left the lines are bent backwards. Examples of these lines are illustrated in figure 1.

Figure 1 also shows that an efficiency line does not exist for all confidence levels α. This is due to the VaR definition in equation (1); thereafter, the VaR is a linear combination comprising two terms: (1) the product of a scalar and the standard deviation and (2) the portfolio return. Only if the first term is larger than the second, there is a solution for minimisation of VaR. This is the case when

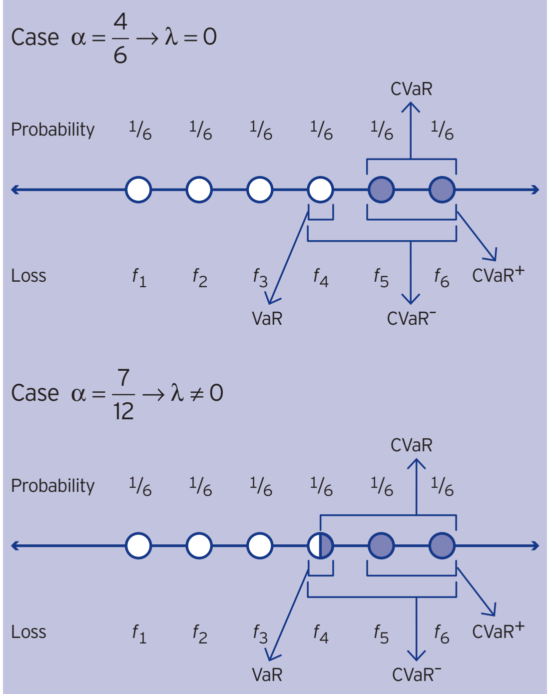
$$\alpha > \Phi\left(\sqrt{D/C}\right),$$

Figure 1: Efficiency lines of mean-VaR efficient portfolios



Source: Invesco. For illustrative purposes only.

Figure 2: Discrete loss distribution – VaR, CVaR, CVaR$^+$ and CVaR$^-$



Source: Invesco. For illustrative purposes only.

where Φ represents the distribution function of the normal distribution. A closed-form solution can then be directly found for the portfolio weightings at a given confidence level using:

$$(10)\quad \omega_{VaR_\alpha} = g + h\left[\frac{A}{C} + \sqrt{\frac{D}{C}\left(\frac{z_\alpha^2}{C(z_\alpha)^2 - D}\right)} - \frac{1}{C}\right]$$

where the vectors **g** and **h** are defined as:

$$(11)\quad \mathbf{g} = \frac{1}{D}\left[B\Sigma^{-1}\mathbf{i} - A\Sigma^{-1}\mu\right]$$

$$(12)\quad \mathbf{h} = \frac{1}{D}\left[C\Sigma^{-1}\mu - A\Sigma^{-1}\mathbf{i}\right]$$

A closed-form solution for VaR_α^m can also be found using:

$$(13)\quad VaR_\alpha^m = z_\alpha\sqrt{\frac{z_\alpha^2}{C(z_\alpha)^2 - D}} - \left[\frac{A}{C} + \sqrt{\frac{D}{C}\left(\frac{z_\alpha^2}{C(z_\alpha)^2 - D} - \frac{1}{C}\right)}\right]$$

Because efficient mean-VaR portfolios are always simultaneously efficient mean-variance portfolios, the solutions on the efficiency line (with the exception of the global minimum-variance portfolio) are also solutions for the mean-VaR optimisation at a given confidence level. This confidence level α_{MV} can be calculated with

$$(14)\quad \alpha_{MV} = \Phi\left(\sqrt{\frac{D}{C} + \frac{D^2/C^3}{(\bar{r} - A/C)^2}}\right)$$

Efficient mean-variance portfolios, however, are inefficient mean-VaR portfolios if the confidence level is lower than in equation (14). For its part, the global minimum-variance portfolio is always mean-VaR inefficient, always permitting portfolio allocations with higher portfolio returns and lower risk.

All of the above applies when a risk-free asset is permitted or expected shortfall (ES) is used in place of VaR. Ultimately, ES is also a quantile value. Unlike VaR, however, it represents a coherent risk measure and, as such, is preferable.

Efficient mean-CVaR portfolios

Thus far, we have worked on the assumption of normally distributed returns, which – as demonstrated in earlier articles – are not in line with the realities of the financial markets. Therefore, a method would be desirable that requires no specific distribution assumptions and delivers a coherent measure of risk similar to ES. In several articles, Uryasev and Rockafellar have proposed portfolio optimisations for Conditional Value-at-Risk (CVaR).[1] Like ES, the CVaR measure is defined as the average loss upon violation of VaR limits – with the difference that ES is usually defined for continuous random variables whereas the CVaR applies to discrete ones. Synonyms for this are mean excess loss, mean shortfall or tail VaR.

In addition to coherence, convexity is another advantage of CVaR, whereas VaR is not convex.[2] Using VaR optimisations, there is thus a danger of achieving a local optimum rather than a global one. However, a direct optimisation of CVaR is challenging, as it is dependent on VaR. The authors solve this problem by introducing additional risk measures and definition of CVaR as a weighted average. The resulting target function is linear, and can be specified as a linear programme or integrated into a mathematical programme as a constraint. Below, we illustrate the procedure and problem definition for discrete losses, based on the following risk measures:

- VaR^{α}: the α-quantile of the loss distribution
- $CVaR^{+}$: the average loss that is strictly greater than VaR
- $CVaR^{-}$: the expected loss that is greater than or equal to VaR

The CVaR measure is then defined as the weighted average of VaR and $CVaR^{+}$:

(15) $CVaR = \lambda VaR + (1 - \lambda) CVaR^{+}$

For the $0 \leq \lambda \leq 1$ weighting, $\lambda = (\Psi(VaR) - \alpha)/(1 - \alpha)$ applies. Here, $\Psi(VaR)$ is the probability that the loss will be lesser than or equal to VaR at a given confidence level. Figure 2 illustrates this for a discrete loss distribution.[3]

The horizontal axis in figure 2 display six equally probably losses $f_1, \ldots, f_6$ each.

In the first case, the confidence level is $\alpha = 4/6$, which equals the VaR-quantile. Consequently, the weighting factor λ equals zero. The risk measures CVaR and $CVaR^{+}$ are identical and equivalent to the average of the losses f_5 and f_6. The VaR equals f_4, and $CVaR^{-}$ is equal to the average of $f_4, \ldots, f_6$. This, the relationship between the risk measures can be described as $VaR < CVaR^{-} < CVaR = CVaR^{+}$.

This is not the case for $\alpha = 7/12$. Although the value of $CVaR^{+}$ is this same as in the first example, here it is $\lambda = 1/5$. CVaR is calculated as $CVaR = 1/5 VaR + 4/5 CVar^{+}$. For the losses: $CVaR = 1/5 f_4 + 2/5 f_5 + 2/5 f_6$. In this second example, the relationship between the risk measures is $VaR < CVaR^{-} < CVaR < CVaR^{+}$.

Now we will introduce the loss function $f(\omega, \mathbf{r})$, which is equivalent in the simplest case to the negative value of the scalar product of the weighting and return vectors. The CVaR measure can then be stated as:

$$(16)\quad F(\omega,\alpha) = (1-\alpha)\zeta + \frac{1}{J}\sum_{j=1}^{J}\left(f(\omega,\mathbf{r}_j)-\zeta\right)^{+}$$

Here, ζ represents VaR, and $y^{+} = \left(f(\omega,\mathbf{r}_j)-z\right)^{+}$ is defined as $y^{+} = \max\{0,y\}$.

Equation (16) can now be plugged into a linear programme, in which the portfolio solution has a minimal CVaR at a given confidence level:

$$(17)\quad P_{CVaR} = \underset{\omega\in\Omega,\zeta\in\mathbb{R}}{\arg\min}\ \zeta + \nu\sum_{j=1}^{J} y_j$$

$$y_j \geq f(\omega,\mathbf{r}_j) - \zeta$$

$$y_j \geq 0$$

$$\omega'\mu = \bar{r}$$

$$\omega' i = 1$$

In this case, $\nu = 1/(1 - \alpha)J$ is a constant factor. The statement of the problem in (17) can be expanded to include additional restrictions for the portfolio weighting range.

In practical implementation, it must be determined which data sets for $\mathbf{r}_j$ apply. To this end, two approaches are possible: either one simply applies historical returns or $\mathbf{r}_j$ is based on the results of a Monte Carlo simulation.

With CVaR as a constraint, a mathematical programme can be set up using

$$\zeta + \nu\sum_{j=1}^{J} y_j \leq C.$$

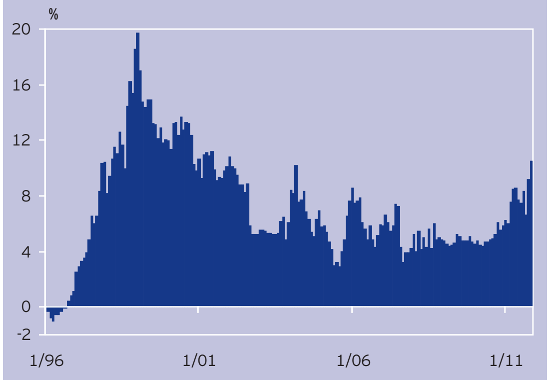
In this case, C designates the upper limit.

Empirical application

The following simulation will now compare solutions for a globally invested minimum-variance portfolio against those of a minimum-CVaR portfolio.[4]

Permissible portfolio investments are equities from the US, UK, eurozone and Japan, as well as Asian and Latin American emerging markets, which are tracked by the S&P 500, the FTSE 100, the EuroStoxx 50, the Topix, the MSCI Emerging Market Asia Performance Index and the MSCI Emerging Market Latin America Performance Index,

Figure 3: Relative performance minimum CVaR vs. minimum-variance



Source: Invesco. Based on data from March 1991 to November 2011. For illustrative purposes only.

respectively. The results do not take into account currency risks. A long-only restriction was applied for problem definition, as well as a confidence level of 95%. Simulated monthly returns encompass the period from March 1991 to November 2011. The comparative back-test was implemented on the basis of a 60-month moving window; portfolio returns were calculated using the current weighting vector and the prior-month performance of the individual equities.

This resulted in a total of 192 data points to be analysed. Figure 3 shows the outperformance of the CVaR approach as compared to the minimum-variance method, which amounted to 73 basis points per annum.

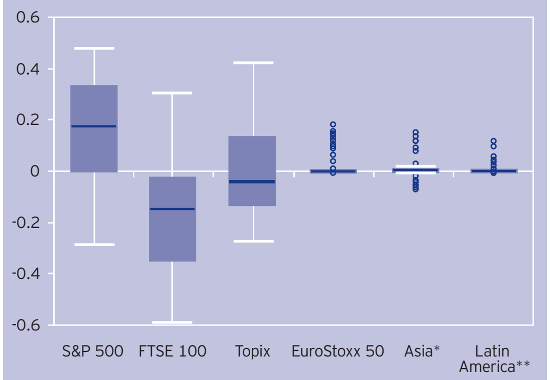
The boxplots in figure 4 illustrate the different portfolio allocations – the proportion of US equities under the CVaR approach was higher on average, at the expense of UK and Japanese holdings.

Average CVaR of the minimum-variance portfolio, at 7.86, is higher than the 7.47 of the minimum-CVaR portfolio. This is in line with expectations and any other result would have been very surprising.

Summary and outlook

A minimum-variance strategy does not offer investors optimum protection against loss risk. Results can be improved through approaches that optimise the portfolio on the basis of mean and value-at-risk rather than mean and variance. Such a method, however, can be further improved since financial market returns (contrary to the assumption of a VaR calculation) generally are not normally distributed. Accordingly, we have introduced the concept of optimisation on the basis of mean and conditional value-at-risk. A simulation study bears out the advantages of this approach.

Figure 4: Boxplots of weighting differences



* Asia: MSCI Emerging Market Asia Performance Index
** Latin America: MSCI Emerging Market Latin America Performance Index
Source: Invesco. Based on data from March 1991 to November 2011. For illustrative purposes only.
A boxplot provides a graphical overview for the empirical distribution of a time series on the basis of five indicators. The median is represented by a horizontal line, the first and third quartile constitute the lower and upper edges of the rectangle. The offset longitudinal line is the "whisker". This is defined as 1.5 times the interquartile distance between the values for the first and third quartiles. Observations that are greater than the whisker in absolute terms are classified as outliers, and indicated as such with circles in the graph.

In the next installment of this series, we will show how the method for determination of mean-CVaR portfolios can be adapted to further measures of risk, in particular portfolio drawdown.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Asset Allocation

Notes:

1 See Rockafellar and Uryasev (2000), Rockinger and Jondeau (2001), Rockafellar and Uryasev (2002), Uryasev and Rockafellar (1999), Rockafellar and Uryasev (2001), Uryasev (2005).
2 Rockafellar and Uryasev (2001).
3 Adapted from Uryasev (2005).
4 All calculations were conducted using the free statistical program environment R 2.14.0 (R Development Core Team, 2011) and the Rsocp package (Würtz et al., 2010).

Bibliography

- Alexander, G. and A. Baptista (1999, September): Value at risk and mean-variance analysis. Working Paper 9804, University of Minnesota, Minneapolis.

- Alexander, G. and A. Baptista (2002): Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis. Journal of Economic Dynamics & Control 26, 1159–1193.

- Alexander, G. and A. Baptista (2004, September): A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. Management Science 50(9), 1261–1273.
- Alexander, G. and A. Baptista (2008): Active portfolio management with benchmarking: Adding a value-at-risk constraint. Journal of Economic Dynamics & Control 32, 779–820.
- De Giorgi, E. (2002, August): A note on portfolio selection under various risk measures. Working Paper 9, National Centre of Competence in Research, Financial Valuation and Risk Management, Zurich, Switzerland.
- Merton, R. (1972, September): An analytical derivation of the efficient portfolio frontier. Journal of Financial and Quantitative Analysis 7, 1851–1872.
- R Development Core Team (2011). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Rockafellar, R. and S. Uryasev (2000): Optimization of conditional value-at-risk. The Journal of Risk 2(3), 21–41.
- Rockafellar, R. and S. Uryasev (2001, Juli): Conditional value-at-risk for general loss distributions. Research Report 2001-5, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.
- Rockafellar, R. and S. Uryasev (2002): Conditional value-at-risk for general loss distributions. Journal of Banking and Finance 26, 1443–1471.
- Uryasev, S. (2005): Conditional value-at-risk (CVaR): Algorithms and applications. Master in finance and risk management (finarm), Universita degli Studi di Milano, Milan, Italy.
- Uryasev, S. and T. Rockafellar (1999, June): Optimization of conditional value-at-risk. Research Report 99-4, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.
- Würtz, D., Y. Chalabi, W. Chen, and A. Ellis (2010, April): Portfolio Optimization with R/Rmetrics. Rmetrics Association & Finance Online, www.rmetrics.org. R package version 2110.79.

Invesco

In the first parts of this series we presented optimisation methods in which portfolio risk was measured using the volatility or the (conditional) Value-at-Risk. In the fifth part, we will measure risk using the drawdown.

The drawdown refers to the percentage loss of a portfolio between a local maximum and a subsequent local minimum. The duration of the drawdown is the period from the local maximum to the local minimum and then to the date on which the original local maximum is reached again. Illustratively speaking, the development runs from a peak and through a valley to the next peak, which is of at least equal height as the first one.

Understandably, investors and asset managers value less pronounced drawdowns.[1] The approach presented here thus raises the question of how the portfolio allocation should be ex ante so that a certain portfolio drawdown is not exceeded. Here, we turn to Chekhlov et al. (2000, 2005).[2]

Minimum drawdown portfolios

Formally, the drawdown of a portfolio consisting of N separate components at time t is the difference between the maximum previous portfolio value and the current portfolio value W:

$$(1) \quad \mathbb{D}(\omega,t) = \max_{0\le\tau\le t}\{W(\omega,\tau)\} - W(\omega,t)$$

Here, $y(t)$ stands for the $(N \times 1)$ price vector and ω for the $(N \times 1)$ weighting vector of the portfolio components, thus the portfolio value at time t is $W(\omega,t) = y(t)\omega$.

Based on this formal drawdown definition, Chekhlov et al. (2005) derive the following three functional measures of risk:

- maximum drawdown (MaxDD)
- average drawdown (AvDD)
- conditional drawdown (CDaR)

The CDaR is calculated similarly to the CVaR[3]. For a given data set for the time interval $[0, T]$; it is defined as:

$$(2) \quad \mathrm{CDaR}(\omega)_\alpha = \min_{\zeta}\left\{\zeta + \frac{1}{(1-\alpha)T}\int_0^T\left[D(\omega,t)-\zeta\right]^+\right\}dt$$

Thus, the portfolio drawdown should only exceed the threshold value ζ in $(1 - \alpha)\,T$ cases; α is the specified confidence level.[4]

The definition of CDaR in equation 2 includes the MaxDD and the AvDD as borderline cases. With $\alpha \to 1$, the CDAR approaches the maximum drawdown:

$$\mathrm{CDaR}(\omega)_{\alpha\to 1} = \mathrm{MaxDD}(\omega) = \max_{0\le t\le T}\{\mathbb{D}(\omega,t)\};$$

with $\alpha = 0$, we obtain the AvDD:

$$\mathrm{CDaR}(\omega)_{\alpha=0} = \mathrm{AvDD}(\omega) = {}^{1}\!/_{T}\int_0^T \mathbb{D}(\omega,t)\,dt$$

To optimise the portfolio drawdowns, conditions of inequality can be incorporated into a linear programme. It can be required, for example, that the maximum portfolio drawdown shall not exceed 10%: $\mathrm{MaxDD}(\omega) \le \nu_1 C$, where $0 \le \nu_1 \le 1$ and C represents the capital. Similarly, one could include $\mathrm{AvDD}(\omega) \le \nu_2 C$ as a constraint for the AvDD and $\mathrm{CDaR}(\omega) \le \nu_3 C$ as a constraint for the CDaR. A linear combination of all three requirements with $0 \le \nu_2, \nu_3 \le 1$ would also be conceivable.

The portfolio optimisation itself is then formulated in discrete steps; the target function to be maximised could be the average annualised portfolio return, e.g.

$$(3) \quad R(\omega) = \frac{1}{dC}\mathbf{y}'_T\,\omega$$

where d is the number of years in the time interval $[0,T]$.

Chekhlov et al. (2005) suggested the following three linear programmes, in which the MaxDD, the AvDD or the CDaR is included as a condition of inequality:

The maximisation of the average annualised portfolio return, with the constraint that the maximum drawdown (MaxDD) is no greater than the threshold value $\nu_1 C$, can be formulated as follows:

$$(4) \quad P_{\mathrm{MaxDD}} = \arg\max_{\omega\in\Omega,\,\mathbf{u}\in\mathbb{R}} R(\omega) = \frac{1}{dC}\mathbf{y}'_T\omega$$
$$u_k - \mathbf{y}'_k\omega \le \nu_1 C$$
$$u_k \ge \mathbf{y}'_k\omega$$
$$u_k \ge u_{k-1}$$
$$u_0 = 0$$

where with $\mathbf{u}$ a $(T + 1 \times 1)$ vector of slack variables was included in the LP. These slack variables represent the maximum portfolio value at point in time k with $1 \le k \le T$.

If an optimisation with a constraint for the average drawdown (AvDD) is to be made, only the first group of constraints needs to be changed. The following linear programme would therefore be used:

$$(5) \quad P_{\mathrm{AvDD}} = \arg\max_{\omega\in\Omega,\,\mathbf{u}\in\mathbb{R}} R(\omega) = \frac{1}{dC}\mathbf{y}'_T\omega$$
$$\frac{1}{T}\sum_{k=1}^{T}\left(u_k - \mathbf{y}'_k\omega\right) \le \nu_2 C$$
$$u_k \ge \mathbf{y}'_k\omega$$
$$u_k \ge u_{k-1}$$
$$u_0 = 0$$

The formulation of the linear programme for the conditional drawdown (CDaR) is, however, more complex. In addition to the slack variable **u**, which stands for the (local) maximum portfolio value, two further auxiliary variables must be used: ζ denotes the threshold value for the portfolio drawdown at given confidence level α, **z** describes the ($T \times 1$) vector of the values in excess of this endogenous parameter. The linear programme can then be written as:

$$(6) \quad P_{\text{CDaR}} = \underset{\omega \in \Omega,\, \mathbf{u} \in \mathbb{R},\, \mathbf{z} \in \mathbb{R},\, \zeta \in \mathbb{R}}{\arg\max} \; R(\omega) = \frac{1}{dC}\mathbf{y}_T'\omega$$

$$\zeta + \frac{1}{(1-\alpha)T}\sum_{k=1}^{T} z_k \leq \nu_3 C$$

$$z_k \geq u_k - \mathbf{y}_k'\omega - \zeta$$

$$z_k \geq 0$$

$$u_k \geq \mathbf{y}_k'\omega$$

$$u_k \geq u_{k-1}$$

$$u_0 = 0$$

Of course, further restrictions on the weighting vector ω can be added to the solution concepts in (4) to (6) (including budgetary restrictions, conditions of non-negativity, minimum and / or maximum constraints, group restrictions).

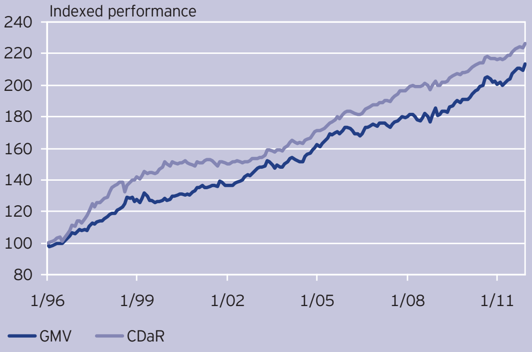
As with the CVaR solution, historical returns are normally used. Although, data which was simulated with multivariate models can also be used. However, regardless of the selected data design, we should highlight two drawbacks of MaxDD and AVDD risk functions: Firstly, the optimal portfolio allocation greatly depends on a singular drawdown (outlier return). Secondly, the concept of AvDD says nothing about the frequency of drawdowns, nor about the maximum loss. It only looks at the average, with all its advantages and disadvantages. The CDaR risk function does not have these weaknesses, cum grano salis, and is therefore the preferable drawdown concept.

Empirical application

Below we will compare the backtest results for the CDaR-optimum portfolio with the results for the global minimum-variance portfolio (GMV)[5]. The data base comprises the month-end values of selected stock performance indices and the monthly performance of ten-year government bonds from January 1991 to December 2011. We use performance indices for the USA (S&P 500), UK (FTSE100), Japan (TOPIX), Europe (EuroStoxx), Australia (ASX), Canada (TSX), Latin America (MSCI Latin America) and Asia (MSCI Asia), as well as government bonds from the USA, Germany, Japan and the United Kingdom.

The portfolio allocations were determined using a recursive data window, with an initial period of 60 months. The backtest thus consisted of a total of 192 periods. The drawdown was limited to 4% with a confidence level of 90%. Both optimisations were carried out as long-only portfolios with a 100 percent investment ratio. Currency risks were ignored in the performance measurement.

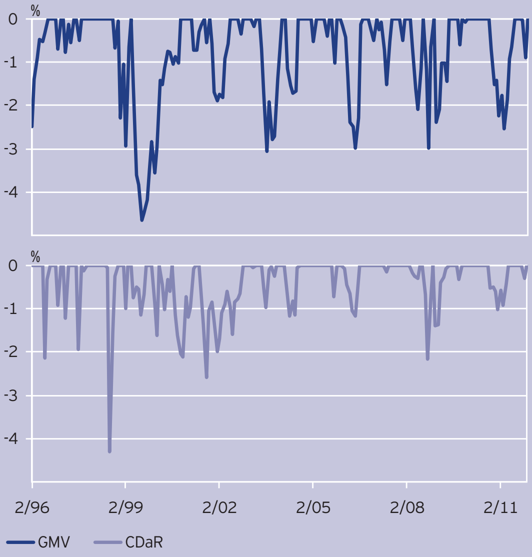
Figure 1: Backtested performance in comparison



Source: Invesco. Data as at December 2011.
For illustrative purposes only.

Figure 1 shows the performance of the CDaR-optimum portfolio and the global minimum-variance portfolio.[6] During the entire backtest period, the value of the CDaR-optimum portfolio is higher than that of the global minimum-variance portfolio. This is also reflected in the average annualised returns of both strategies. It is 5.27% for the CDaR approach and 4.88% for the GMV strategy.

Figure 2: Backtest portfolio drawdowns in comparison



Source: Invesco. Data as at December 2011.
For illustrative purposes only.

Table 1: The five largest drawdowns (backtest results)

	Portfolio	Begin	Trough	End	Drawdown	Overall duration (months)	Downward phase (months)	Recovery phase (months)
GMV	1	05/1999	08/1999	11/2000	4.65	19	4	15
	2	06/2003	08/2003	02/2004	3.06	9	3	6
	3	09/2008	10/2008	12/2008	3.00	4	2	2
	4	01/2006	06/2006	09/2006	2.98	9	6	3
	5	10/1998	02/1999	04/1999	2.93	7	5	2
CDaR	1	07/1998	08/1998	11/1998	4.31	5	2	3
	2	07/2001	09/2001	11/2002	2.59	17	3	14
	3	09/2008	10/2008	12/2008	2.16	4	2	2
	4	07/1996	07/1996	09/1996	2.14	3	1	2
	5	09/2000	12/2000	05/2001	2.12	9	4	5

Source: Invesco. Data as at December 2011. For illustrative purposes only.

Table 2: Portfolio parameters (backtest results)

Portfolio	GMV	CDaR
VaR (95%)	1.282	1.202
ES (95%)	1.710	1.618
Sharpe Ratio	0.158	0.198
Maximum Drawdown	4.654	4.308
Average Drawdown	1.464	1.243
Number of Drawdowns	27	24

Source: Invesco. Data as at December 2011.
For illustrative purposes only.

Figure 2 shows the portfolio drawdowns; table 1 also the five largest drawdowns. It transpires that the average losses of the GMV strategy are higher. The maximum loss also would have been incurred with the GMV optimisation. A striking point is that with the parameters selected for the CDaR portfolio, one would expect about 20 breaches of the drawdown limit ex ante, but there is only one ex post.

Furthermore, table 1 shows that four of the five largest portfolio drawdowns would have arisen in the GMV strategy and that the drawdown phases were longer on average.

As can be seen in table 2, at a confidence level of 95%, the risk measures VaR and ES (i.e those assuming normal distribution) are slightly higher with the GMV portfolio than with the CDaR approach (ex post). In contrast, the Sharpe ratio is about 25% lower. Furthermore, as already evident from figure 2, the maximum drawdown, the average drawdown and the number of drawdowns are higher with the GMV strategy.

These differences arise naturally from the different allocations of both strategies. Table 3 shows the minimum, average and maximum shares of the individual markets and asset classes.

On average, over 80% of the GMV portfolio consists of German and Japanese bonds, compared with

Table 3: Portfolio allocations (backtest results)

Market		GMV			CDaR		
		min.	mean	max.	min.	mean	max.
Equities	S&P 500	0.00	0.23	1.43	0.00	11.53	47.01
	FTSE 100	0.00	0.00	0.34	0.00	0.00	0.00
	TOPIX	4.49	5.90	6.71	0.00	4.51	12.65
	EuroStoxx	0.00	0.22	0.75	0.00	0.10	4.27
	ASX	0.00	1.39	2.93	0.00	0.04	1.34
	TSX	0.00	2.79	4.46	0.00	1.52	6.43
	MSCI EM Asia	0.00	0.00	0.00	0.00	0.53	2.29
	MSCI EM LA	0.00	0.00	0.13	0.00	0.35	2.15
Bonds	USA	0.00	0.54	4.24	0.00	0.63	6.91
	Germany	33.39	41.05	55.65	0.00	17.63	32.31
	Japan	32.33	40.60	47.33	8.20	23.53	44.13
	United Kingdom	0.00	7.28	11.22	0.00	0.44	7.90

Source: Invesco. Data as at December 2011. For illustrative purposes only.

around 40% in the CDaR portfolio. Furthermore, the equity component of the GMV strategy is significantly lower. At an average of 5.9% and maximum of 6.7%, only Japanese stocks are relatively well represented.

Summary and outlook

In this article, we have presented an optimisation method using the portfolio drawdown as a risk measure. In a comparative backtest, we have shown that this approach has advantages over a GMV strategy. The performance and other portfolio indicators are better, and the loss phases are both shorter and less pronounced than with the minimum-variance portfolio.

Having dealt with portfolio optimisation for one-sided risk measures in the last two parts, the next article looks at what portfolio diversification actually means. Here too, we will show alternatives to the Markowitz approach.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Asset Allocation

Notes:

1 See Chekhlov et al. (2005), which explains this using typical CTA mandates.

2 Chekhlov (2000) is reprinted in Chekhlov et al. (2004), a comparison of the optimisation for the drawdown and for other risk measures can be found in Krokhmaal et al. (2002).

3 Conditional Value-at-Risk, see "New approaches to portfolio optimisation: Part 4", Risk & Reward Q1/2012

4 If $(1 - \alpha)T$ is not an integer, the next lower integer is used.

5 See "New approaches to portfolio optimisation: Part 1", Risk & Reward Q2/2011, and "New approaches to portfolio optimisation: Part 2", Risk & Reward Q3/2011.

6 All calculations were performed with the free statistical programming environment R 2.14.0 (R Development Core Team, 2010) and fPortfolio (Würtz et al., 2010), Hmisc (Harrell et al., 2010), RODBC (Ripley and Lapsley, 2011), Rglpk (Theussl and Hornik, 2010), fPortfolio (Würtz et al., 2010) and Performance Analytics (Carl and Peterson, 2010).

Bibliography

- Carl, P. and B. Peterson (2010). Performance Analytics: Econometric tools for performance and risk analysis. R package version 1.0.3.2.
- Chekhlov, A., S. Uryasev, and M. Zabarankin (2000). Portfolio optimization with drawdown constraints. Research report 2000-5, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.
- Chekhlov, A., S. Uryasev, and M. Zabarankin (2004). Supply Chain and Finance, Volume 2 of Series on Computers and Operations Research, Chapter 13 Portfolio Optimization with Drawdown Constraints, pp. 209-228. Singapore: World Scientic Publishing Co. Pte. Ltd.
- Chekhlov, A., S. Uryasev, and M. Zabarankin (2005). Drawdown measure in portfolio optimization. International Journal of Theoretical and Applied Finance 8 (1), 13-58.
- Harrell, F. and et al. (2010). Hmisc: Harrell Miscellaneous. R package version 3.8-3.
- Krokhmal, P., S. Uryasev, and G. Zrazhevsky (2002, Summer). Risk management for hedge fund portfolios. Journal of Alternative Investments 5 (1), 10-29.
- R Development Core Team (2010). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Ripley, B. and M. Lapsley (2011). RODBC: ODBC Database Access. R package version 1.3-3.
- Theussl, S. and K. Hornik (2010). Rglpk: R/GNU Linear Programming Kit Interface. R package version 0.3-5.
- Würtz, D., Y. Chalabi, W. Chen, and A. Ellis (2010, April). Portfolio Optimization with R/Rmetrics. Rmetrics Association & Finance Online, www.rmetrics.org. R package version 2110.79.

Past as well as backtested performance is not indicative of future performance.

New approaches to portfolio optimisation: Part 6

In the sixth part of our series, we look at the concept of diversification. It is standard practice to measure portfolio diversification using the return covariances between the individual positions. However, as we demonstrate in this and the following parts of our series, this is only one possibility – and moreover, it is often misleading. We illustrate this by comparing classic allocation models with an alternative concept: the "most diversified portfolio" (MDP).

Diversification has at least two dimensions. "Diverse in terms of what?" is the first; "How should diversification be measured?" is the second. Those who use the variance-covariance or correlation matrix of the returns to measure diversification have already implicitly answered these questions. Firstly, they equate diversification with most diverging return distributions; secondly, they measure their dispersion using the paired squared deviations from the mean value.

However, dependencies between return processes are only correctly reproduced by the correlation matrix when the processes have a joint elliptical distribution. As demonstrated in previous issues of *Risk & Reward*, this should rarely be the case in practice.

Another factor to consider is that the variance-covariance matrix of the returns is a symmetrical dispersion measure and thus equally covers both positive and negative deviations from the position parameter.[1] For long-only investors, for whom only the left-hand side of a return distribution (i.e. the losses) counts from a risk perspective, a symmetrical dispersion measure is only an approximation of the dependencies for the left-hand side of the distribution.

This will, however, be the subject of the next part of this series. In the introductory article, we first present the concept of the "most diversified portfolio" (MDP) and, using an example, compare this with the global minimum variance portfolio (GMV) and the maximum Sharpe ratio portfolio (MSR).

Diversification ratio ...

In several articles, Choueifaty and Coignard (2008, 2011) analysed portfolio diversification, both theoretically and empirically. This requires a metric for the degree of diversification.

Below, we are solely concerned with long-only portfolios. Σ denotes the variance-covariance matrix of returns for N financial instruments, and σ denotes the $(N \times 1)$ vector of their volatility (as measured by the standard deviations). The authors have defined the diversification ratio (abbreviation: DR) for a given $(N \times 1)$ weight vector ω from the multitude of possible portfolio solutions Ω as:

$$(1)\quad DR_{\omega\in\Omega} = \frac{\omega'\sigma}{\sqrt{\omega'\Sigma\omega}}$$

The numerator is the weighted average volatility of the individual financial instruments, defined as a scalar product of the weight vector and the volatility vector. The denominator is the standard deviation of the portfolio. The higher the diversification ratio, the more diversified the portfolio.

The DR is at least one, which is the case with a portfolio consisting of just one single asset. Using equation (1), highly concentrated portfolios or portfolios with highly correlated return processes would be classified as not well-diversified. This can be illustrated in the following breakdown:[2]

$$(2)\quad DR_{\omega\in\Omega} = \frac{1}{\sqrt{\left(\rho + CR\right) - \rho CR}}$$

where ρ denotes the volatility-weighted average correlation, and CR denotes the volatility-weighted concentration ratio. ρ is given by:

$$(3)\quad \rho_{\omega\in\Omega} = \frac{\sum_{i\neq j}^{N}\left(\omega_i\,\sigma_i\,\omega_j\,\sigma_j\right)\rho_{ij}}{\sum_{i\neq j}^{N}\left(\omega_i\,\sigma_i\,\omega_j\,\sigma_j\right)}$$

CR is the normalised Herfindahl-Hirschmann Index (Hirschman, 1964), which lies within the interval $[1/N, 1]$:

$$(4)\quad CR_{\omega\in\Omega} = \frac{\sum_{i=1}^{N}\left(\omega_i\sigma_i\right)^2}{\left(\sum_{i=1}^{N}\omega_i\sigma_i\right)^2}$$

The partial derivatives of equation (2) are:

$$(5a)\quad \frac{\partial DR}{\partial \rho} = -\frac{1}{2}\frac{\frac{1-CR}{\sqrt{\rho+CR-\rho CR}}}{\rho + CR - \rho CR}$$

$$(5b)\quad \frac{\partial DR}{\partial CR} = -\frac{1}{2}\frac{\frac{1-\rho}{\sqrt{\rho+CR-\rho CR}}}{\rho + CR - \rho CR}$$

This results in highly concentrated portfolios, or portfolios with highly correlated return processes, being classified as not well-diversified.

Because of the relationship between the number of financial instruments and the concentration ratio, comparisons between portfolios are only permitted when the investment universes are equal in size. It should also be noted that in the case of a naïve – or equally weighted – portfolio allocation, the DR only depends on the volatility-weighted correlations.[3]

... and "most diversified portfolio" (MDP)
Based on the DR, Choueifaty et al. (2011) formulated the following condition, which the "most diversified portfolio" must satisfy:

$$(6) \quad P_{\mathrm{MDP}} = \arg\max_{\omega \in \Omega} \mathrm{DR}$$

In practice, the weights are determined in two steps. Conceptually, in step one synthetic investment instruments with the same volatility are defined. Then, the correlation matrix for these instruments, which is to be directly included in the portfolio optimisation, will be calculated,

Thus, the DR is maximised when $\omega' C \omega$ is minimised, whereby C is the correlation matrix of the original asset returns. Consequently, the target function is identical to that of a global variance minimal portfolio (GMV)[4] – but with the difference that the correlation matrix is used instead of the variance-covariance matrix, so that the portfolio solution with the lowest possible correlation between the assets is determined. Only in a second step, are the positions adjusted to take account of asset volatility.

The difference from the GMV optimisation is therefore the downstream consideration of the asset volatility. In this way, unlike with the GMV portfolio, the volatility of the financial instruments does not become a dominant driver of the allocation (minimisation of a quadratic form). This is only considered in the subsequent rescaling of the position variables. It is, therefore, by no means immaterial whether covariances or correlations are considered.

This can also be derived from the DR breakdown in equations (2) to (4). The risks of the respective financial instruments only occur as relative variables in the concentration ratio CR.

Finally, let us show two basic properties of an MDP (Choueifaty et al., 2011, page 7f.):

1. Each asset which is not part of the MDP is more strongly correlated with the MDP than the financial instruments included in the MDP. All financial instruments included in the MDP show the same correlation with the MDP.

2. The correlation of the long-only MDP with another long-only portfolio is always greater or equal to the ratio of both portfolio DRs.

The second property can also be formulated as follows: the more diversified a long-only portfolio, the more highly correlated is its return process with that of the MDP.

Empirical application
Below, the MDP concept is applied to a global equity portfolio. The analysis is based on the monthly returns of the MSCI performance indices, calculated in euros, for Australia, Brazil, Canada, Germany, Japan, Switzerland, the UK and the USA, from January 2003 to December 2009.

A naïve allocation, i.e. an equally weighted portfolio (EQW), is assumed as the benchmark. Such an allocation would be optimal if the individual asset returns would follow random walks, i.e. I(0) processes without drift. One would then expect no excess returns from individual financial instruments, but would assume all returns to have an expected value of zero. But the possibly different risks of the individual assets would remain completely disregarded.

We are, however, not comparing the MDP solution with the naïve EQW allocation only, but also with the global variance minimal portfolio (GMV), as well as the portfolio with the highest Sharpe ratio (MSR).

The method of analysis is a so-called fix-me strategy. The total period is divided into two sub-periods. The portfolio optimisations were first carried out with the data from the in-sample period (January 2003 to December 2007), a sample of 60 observations. A long-only restriction and a budget restriction were included as constraints. In a second step, the allocations obtained were used for an ex-ante simulation. The simulation period starts in January 2008 and covers 24 months. This was deliberately chosen to include the sub-prime crisis, a clearly unfavourable and highly volatile environment for equities.

Figure 1 shows the relative portfolio allocations (with the EQW portfolio as the benchmark) as lattice diagrams.[5] The large differences between the MDP, GMV and MSR portfolios can be clearly seen. In the MDP portfolio, Japan, Canada and Switzerland are overweighted compared to the naïve allocations and underweighted compared to all other markets. In the GMV approach, Switzerland and the UK are considerably overweighted, while Australia, Brazil and Germany have either no or only marginal presence in the portfolio. In contrast, the Australian and Brazilian equity markets are considerably overweighted in the MSR portfolio.

We check the plausibility of these allocations using the statistics for the individual equity markets in Table 1, which also indicate which countries have portfolio shares of at least 1%.

Figure 2 shows the marginal risk contributions of the individual countries. In the EQW portfolio, there are no distinct risk concentrations, whereas in both the

Figure 1: Simulated portfolio allocations compared

AU = Australia, BR = Brasil, CA = Canada, CH = Switzerland, DE = Germany, JP = Japan, UK = Great Britain, US = United States
Source: Invesco. Data as at January 2003 to December 2007. For illustrative purposes only.

GMV and MSR portfolio, the strongly overweighted markets have very high shares of total risk. The MDP solution lies in the middle.

Table 2 shows metrics for the four optimisation approaches. The standard deviation ("risk") is highest in the MSR portfolio and (by definition) lowest in the GMV portfolio. MDP and EQW are in between. This same ranking applies to the expected shortfall (ES), which was modified using the Cornish-Fisher expansion, at a confidence level of 95%. The degree of diversification is (also by definition) highest in the MDP portfolio, followed by the EQW solution. The GMV and MSR portfolios have similar values here. Since GMV and MSR are corner solutions, they have the highest degree of concentration.

For the four portfolio allocations, the performance was then determined in the second sub-period. Figure 3 shows the portfolio drawdowns.

During the period of declining equity markets until the spring of 2009, the drawdowns are lowest in the MSR and MDP portfolio, but only slightly higher in the GMV portfolio. During the subsequent recovery period, however, the MSR portfolio (relative to the

Table 1: Simulated portfolio allocations and metrics

	Portfolio allocation > 1%				Metrics		
	EQW	MDP	GMV	MSR	Standard deviation	Sharpe ratio	Average correlation
Australia	X			X	3.815	0.488	0.509
Brasil	X	X		X	7.836	0.505	0.453
Canada	X	X	X	X	4.096	0.431	0.471
Germany	X	X		X	4.638	0.400	0.505
Japan	X	X	X		4.647	0.156	0.324
Switzerland	X	X	X	X	2.921	0.346	0.458
UK	X		X		2.842	0.335	0.550
US	X		X		2.950	0.174	0.527

Source: Invesco. Data as at December 2007. For illustrative purposes only.

Figure 2: Marginal risk contributions of the countries

AU = Australia, BR = Brasil, CA = Canada, CH = Switzerland, DE = Germany, JP = Japan, UK = Great Britain, US = United States
Source: Invesco. Data as at January 2003 to December 2007. For illustrative purposes only.

other investment strategies) was able to more than offset the previous losses.

In end value, the MDP portfolio comes in second place, but the differences in performance are only marginal. But, as in earlier articles from this series, a minimum variance strategy was not found to be an optimal optimal allocation per se when the ex-ante portfolio risk or the portfolio drawdown is concerned. In contrast, the amplitude of the MDP drawdowns is lower, which could be due to the higher degree of diversification.

Summary and outlook

In this article, the concept of diversification was questioned. In general, portfolio diversification is reduced to a symmetrical dispersion measure, such as the variance-covariance matrix of returns. This is, however, just one possibility. Still, this metric was adhered to in this introductory article, and it was shown how a portfolio can be designed with maximum diversification on this basis.

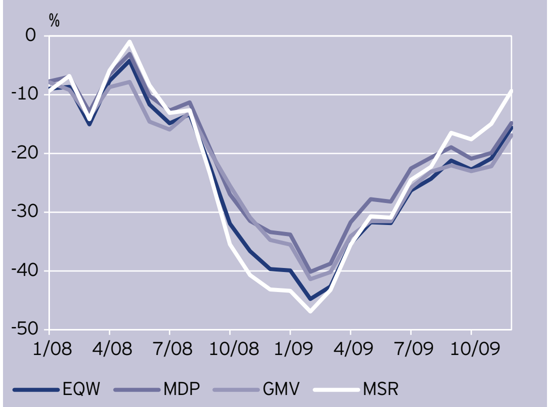
Our simulation study found that the most diversified portfolio is less concentrated than the minimum

Table 2: Simulated portfolio metrics at a glance

	EQW	MDP	GMV	MSR
Risk	3.244	3.176	2.568	3.897
ES (modified, 95%)	5.448	5.073	4.898	6.267
Diversification	1.300	1.358	1.222	1.207
Concentration	0.142	0.222	0.242	0.254

Source: Invesco. Data as at December 2007.
For illustrative purposes only.

Figure 3: Simulated portfolio drawdowns compared



Source: Invesco. Data as at December 2009. For illustrative purposes only.

variance portfolio and the maximum Sharpe ratio portfolio. During the sub-prime crisis, this meant lower losses.

An alternative concept of diversification will be presented in the next article. In particular, we will show how the dependencies between the assets on the left-hand side of the joint distribution can be used, which is of particular interest to long-only investors.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Asset Allocation

Notes:

1 This point was dealt with in Markowitz (1959, chapter 9). Markowitz suggested semi-variance as an alternative.

2 A proof can be found in Choueifaty et al. (2011), Appendix A.

3 If the excess returns of the financial instruments are proportional to the volatility, the DR is also proportional to the Sharpe ratio. Portfolio allocations which maximise the DR are then also Maximum Sharpe ratio portfolios or tangent portfolios (MSR). It should be noted that the excess returns are not directly inserted in equation (1). Nevertheless, the financial instruments with a higher or lower implied expected excess return can be identified in the comparison of the maximum DR portfolio with the maximum Sharpe ratio portfolio.

4 See New approaches to portfolio optimisation: Part 5, Risk & Reward Q1/2012.

5 All calculations were made with the free statistical programming environment R 2.15.0 (see R Development Core Team, 2010) and the package portfolio (see Würtz et al., 2010), FRAPO (see Pfaff, 2012), Hmisc (see Harrell et al., 2010), lattice (see Sarkar, 2008) and Performance Analytics (see Carl and Peterson, 2010).

Bibliography

- Carl, P. and B. Peterson (2010): Performance Analytics: Econometric tools for performance and risk analysis. R package version 1.0.3.2
- Choueifaty, Y. and Y. Coignard (2008): Towards maximum diversification. Journal of Portfolio Management 34(4), 40–51
- Choueifaty, Y., T. Froidure, and J. Reynier (2011): Properties of the most diversified portfolio. Working paper, TOBAM
- Harrell, F. et al. (2010): Hmisc: Harrell Miscellaneous. R package version 3.8-3
- Hirschman, A. (1964): The paternity of an index. The American Economic Review 54(5), 761
- Markowitz, H. (1959): Portfolio selection: Efficient diversification of investments. Monograph 16, Cowles Foundation for Research in Economics at Yale University, New York, London
- Pfaff, B. (2012): Financial Risk Modelling and Portfolio Optimisation with R. London: John Wiley & Sons, Ltd. (forthcoming)
- R Development Core Team (2010): R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0
- Sarkar, D. (2008): Lattice: Multivariate Data Visualization with R. New York: Springer
- Würtz, D., Y. Chalabi, W. Chen, and A. Ellis (2010, April): Portfolio Optimization with R/Rmetrics. Rmetrics Association & Finance Online, www.rmetrics.org. R package version 2110.79

New approaches to portfolio optimisation: Part 7

Investors diversify to prevent their portfolio positions from losing value at the same time. However, the classic diversification measures not only capture the simultaneous losses but also the simultaneous gains. In this article, we present a diversification measure which only records the high losses that occur simultaneously.

A portfolio is considered diversified and less risky when the return covariances of the individual positions remain within limits and the standard deviation of the portfolio returns is correspondingly low. From the perspective of an investor who may only enter into long positions, this definition falls too short – because the standard deviation captures price fluctuations independently of their direction, so it is high not only with considerable losses, but also with strong gains. To add to that, the return co-variances only correctly reflect the dependencies between the portfolio positions if their returns are jointly normally distributed or follow another joint elliptical distribution. This is rarely the case in practice.

Since only negative returns are a risk for a long-only investor, a risk measure which only covers the dependencies at the lower end of the joint distribution would be more appropriate. Diversification would then mean that only a few portfolio positions would lose value at the same time.

The closely linked concepts of the copula and the tail dependence were already introduced in previous methodology articles. In this article, we show how the tail dependence coefficient can be considered in portfolio design. First, we will present parametric estimators for the tail dependence between the random variables, and then the non-parametric estimators. Using an empirical example, we will then show how optimum portfolios can be designed when only the lower tail dependencies are considered.

Tail dependencies

A measure of dependence between two random variables at the ends of their joint distribution function is the tail dependence coefficient (TDC).[1] Since, in this article, we assume that an investor may only enter into long positions, the following observations are limited to the left tail of the joint distribution function of the returns, i.e. to the lower tail dependence coefficient.

For this, we consider the distribution function of two random variables X and Y:

$$(1) \quad F_{X,Y}(x,y) = \mathrm{P}(X \le x, Y \le y) \text{ for } (x,y) \in \mathbb{R}^2$$

This bivariate distribution function can also be written in relation to a copula C:

$$(2) \quad F_{X,Y}(x,y) = C(F_X(x), F_Y(y))$$

where $F_X(x)$ and $F_Y(y)$ stand for the marginal distributions of X and Y. In the selected bivariate example, a copula is a joint distribution function of the marginal distributions $C = \mathrm{P}(U \le u, V \le v)$ with $U = F_X(x)$ and $V = F_Y(y)$, which relates the two-dimensional space $[0,1]^2$ to the one-dimensional space $[0,1]$. If the limit exists, the lower tail dependence coefficient is defined as:

$$(3) \quad \lambda_L = \lim_{u \to 0} \frac{C(u,u)}{u}$$

Because it can be interpreted as a conditional probability, it takes a value between 0 and 1, where 0 is achieved in the independence copula and 1 in a co-monotone copula. The lower tail dependence coefficient is also invariant to strictly monotonic transformations of the random variables (X,Y) and only dependent on the used copula, but not on the marginal distributions of the random variables.

The tail dependence coefficient between the two random variables can be determined empirically with a parametric and non-parametric estimation procedure. Thus, when using the Clayton copula, the lower tail dependence coefficient is a function of its parameter δ:

$$(4) \quad \lambda_L^{\text{Clayton}} = 2^{-1/\delta\pi}$$

and can be determined with a parametric estimation procedure.[2]

But the tail dependencies can also be directly estimated from the data pairs of two random variables, hence from a non-parametric procedure.[3] Again, we limit ourselves to the lower tail dependence coefficient. The non-parametric estimators for λ_L are derived from the empirical copula. In a sample with the size N and the data pairs $(X_1, Y_1), \dots, (X_N, Y_N)$ with the ranking $X_{(1)} \le X_{(2)} \le \dots \le X_{(N)}$ and $Y_{(1)} \le Y_{(2)} \le \dots \le Y_{(N)}$, it is defined as:

$$(5) \quad C_N\left(\frac{i}{N}, \frac{j}{N}\right) = \frac{1}{N} \sum_{l=1}^{N} I\left(X_l \le X_{(i)} \wedge Y_l \le Y_{(j)}\right)$$

with $i, j = 1, \dots, N$. The indicator function I takes the value one if $X_l \le X_{(i)} \wedge Y_l \le Y_{(j)}$ is fulfilled; otherwise it is zero. According to the definition, the empirical copula for $i = j = 0$ also takes the value zero.

Three non-parametric estimators

In the literature, three estimators are presented for λ_L.

The first estimator, λ_L, is an approximation for the derivative of $\lambda_L^{(1)}(N,k)$ with respect to u using the slope of a secant in the vicinity of u. In this case, the probability u is equated with k/N:

$$(6) \quad \lambda_L^{(1)}(N,k) = \left[\frac{k}{N}\right]^{-1} \cdot C_N\left(\frac{k}{N}, \frac{k}{N}\right)$$

With the second estimator, $\lambda_L^{(2)}(N,k)$, the partial derivative is approximated using the slope of a simple regression of the marginal probabilities i/N for $i = 1, \dots, k$ on the copula values. This estimator is defined as:

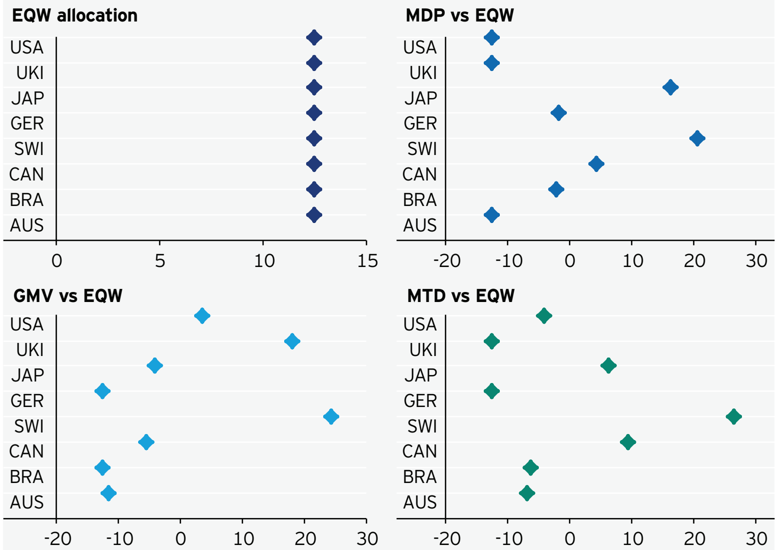
$$(7)\quad \lambda_L^{(2)}(N,k)=\left[\sum_{i=1}^{k}\left(\frac{i}{N}\right)^2\right]^{-1}\cdot\sum_{i=1}^{k}\left[\frac{i}{N}\cdot C_N\left(\frac{i}{N},\frac{i}{N}\right)\right]$$

The third option is to use a mixed copula, namely a weighted combination of an independence copula and co-monotone copula. The lower tail dependence coefficient is then the weight parameter. This estimator is defined as:

$$(8)\quad \lambda_L^{(3)}(N,k)=\frac{\sum_{i=1}^{k}\left(C_N\left(\frac{i}{N},\frac{i}{N}\right)-\left(\frac{i}{N}\right)^2\right)\left(\left(\frac{i}{N}\right)-\left(\frac{i}{N}\right)^2\right)}{\sum_{i=1}^{k}\left(\left(\frac{i}{N}\right)-\left(\frac{i}{N}\right)^2\right)^2}$$

The threshold parameters k for all three estimators are specified. Similarly to the peaks-over-threshold method in the extreme value theory, there must also be a balance here between bias and variance of the estimator. If too low a k were chosen, the estimation of the lower tail dependence would be inaccurate. For larger values of k, although the estimation error is smaller, this is at the price of greater distortion of the measured lower dependence. J. Dobrić and F. Schmid (2005) have shown that with $k \sim \sqrt{N}$, all three estimators are consistent and asymptotically undistorted.

Figure 1: Portfolio allocations compared



EQW: Equal-weighted portfolio; MTD: Minimum Tail-Dependence portfolio;
GMV: Global Minimum Variance portfolio; MDP: Most Diversified Portfolio.
The charts show the absolute country positions of the EQW and the active country positions of the other portfolios relative to the EQW.
Source: Invesco. For illustrative purposes only.

Table 1: Portfolio allocations and metrics

Asset	Statistics			Allocation > 1%			
	Standard deviation	Average TDC	Average correlation	EQW	MDP	GMV	MTD
USA	2.950	0.490	0.527	X		X	X
UKI	2.842	0.571	0.550	X		X	
JAP	4.647	0.327	0.324	X	X	X	X
GER	4.638	0.469	0.505	X	X		
SWI	2.921	0.408	0.458	X	X	X	X
CAN	4.096	0.429	0.471	X	X	X	X
BRA	7.836	0.429	0.453	X	X		X
AUS	3.815	0.469	0.509	X			X

Source: Invesco. For illustrative purposes only.

Tail dependence and portfolio design

Next, we will examine how the estimators for the lower tail dependence can be considered in portfolio design. We show two possibilities.

MTD instead of MDP

In the last article of this series, we presented the concept of the "Most Diversified Portfolio" (MDP).[4] In the first step of this two-step approach, a portfolio allocation is determined based on the correlations of the individual assets; in the second step, the weights are rescaled using the individual assets' volatilities.

The same two-step method can also be used here. In the first step, the pairwise tail dependencies between the individual assets are determined with one of the three estimators shown above, and compiled in a tail dependence matrix. It is symmetric, similar to a correlation matrix; on the main diagonal, there are only values of 1, as the correlation of an individual asset with itself is 1. In the second step, the weight vector which has thus been determined is rescaled using the individual assets' risks. The result is a minimum tail-dependent portfolio (MTD), which is characterized by a minimal probability of simultaneous high losses of the individual assets.

Low β only in loss phases

The tail dependencies can also be used for security-selection strategies with a benchmark index.[5] The procedure is similar to a low β strategy, i.e. the selection of securities, which only track the movements of the benchmark index at a lower rate and therefore not only participate in index losses, but also in index gains, at a lower rate. The β coefficient of an individual asset is the ratio of its covariance with the index and the index variance. The numerator thus contains a symmetric dispersion measure, and the covariance only properly reproduces the dependencies between jointly elliptically distributed random variables.

Tail dependencies instead of the β coefficient have two advantages: firstly, no problematic distribution assumption is required; and secondly, the focus is clearly on the loss side. One would therefore give greater weight to individual assets which do not generally have a low correlation with the benchmark index, but especially during loss phases.[6]

Empirical application

Here, the concept of a MTD is applied to a global equity portfolio. As in the last article of this series, the analysis is based on the monthly returns of the MSCI performance indices, calculated in euros, for Australia, Brazil, Canada, Germany, Japan, Switzerland, the UK and the US, from January 2003 and December 2009. Again, we have divided the overall time period into two sub-periods as part of a fix-me strategy. The portfolios were optimized on the basis of the in-sample period from January 2003 to December 2007 (60 months). For the allocation defined in this way, we determine the performance in the second sub-period (24 months). This design was chosen in order to include the sub-prime crisis, which was unfavourable for equities.

Figure 1 shows, also similarly to the last article, the relative portfolio allocations, with the equally weighted portfolio (EQW) as the benchmark, as lattice diagrams.[7] Table 1 indicates which countries have portfolio shares of more than 1% and also shows the standard deviations of the returns for each country, as well as the average tail dependencies and correlations with the other countries. One can see that with the MTD approach, the portfolio mainly comprises six markets (Germany and the UK are missing), and five markets with the global minimum variance portfolio (GMV) and MDP approaches.

The lack of German stocks in the MTD portfolio can be explained by the fact that the tail dependence of Germany with the equally weighted portfolio is hardly smaller than the tail dependence of the US, but the German market has a much larger standard deviation. The largest overweight is Switzerland (by about 27 percentage points, figure 1), since both the average tail dependence with other countries, as well as the standard deviation, is lowest here.

The low standard deviation is also the reason for the overweight of Switzerland (by almost 25 percentage points) in the GMV, while Australia, Brazil and Germany remain virtually unconsidered here. Although the German market is hardly more volatile than the Japanese market (as with Australia and Brazil too), it is more strongly correlated with the equally weighted portfolio.

Similarly, in the MDP, only five markets are represented with more than 1%. Australia, the UK and the US, the markets with the highest average correlation with the other countries, remain unconsidered.

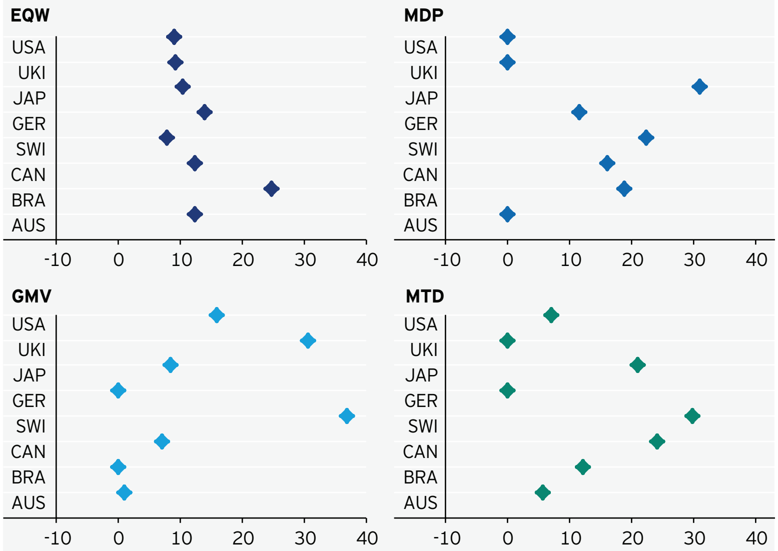
Figure 2 shows the marginal risk contributions of the individual countries. In the EQW, there is no clear concentration of risk; the highest risk contribution comes from Brazil with just under 25%. More pronounced risk bundles, however, can be seen in the other three portfolios. With the MDP approach, Japan and Switzerland have the highest portion of the overall risk, at 31% and 22.5% respectively. The concentration is even greater with the GMV strategy with risk contributions of almost 37% for Switzerland and almost 31% for the UK, so these two countries account for more than two thirds of the overall risk. With the MTD approach, the concentration of risk is lower; about 54% of the standard deviation of returns is attributable to Switzerland and Canada.

Finally, table 2 shows indicators for the different optimisation approaches. By definition the standard deviation risk of the GMV approach is the lowest. The MTD portfolio is only slightly more volatile. The riskiest according to these criteria is the equal weighting strategy, at least ex post.

However, measured on the modified expected shortfall, with a confidence level of 95%, the MTD approach would have the lowest risk. But the MDP portfolio is not far behind, while the GMV strategy provides only little diversification. A comparison of the degree of concentration is only partly possible, as the individual approaches contain different numbers of markets.

Based on the portfolio weights thus determined, we calculated the portfolio development in the second sub-period (figure 3). During the bear market until the spring of 2009, the equally weighted portfolio lost the most and the MTD portfolio the least, closely followed by MDP and GMV. As a whole, an investor with the GMV approach would have lost the most, and one with the MTD strategy would have lost the

Figure 2: Marginal risk contributions of the countries



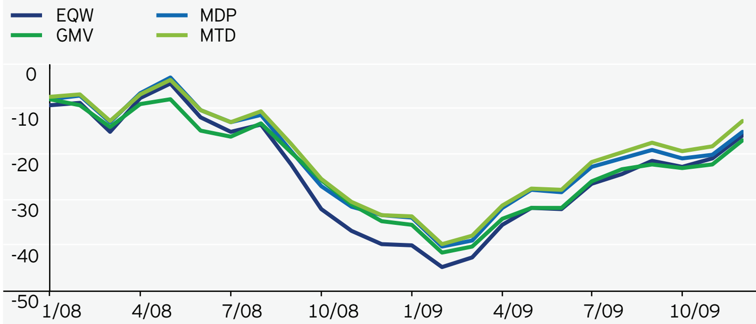
Source: Invesco. For illustrative purposes only.

Table 2: Portfolio metrics at a glance

Statistic	EQW	MDP	GMV	MTD
Risk	3.244	3.176	2.568	2.877
ES (modified, 95%)	5.448	5.073	4.898	4.793
Diversification	1.300	1.358	1.222	1.342
Concentration	0.142	0.222	0.242	0.216

Source: Invesco. For illustrative purposes only.

Figure 3: Portfolio drawdowns compared



Source: Invesco, as at 31 December 2009. For illustrative purposes only.

least (almost 17% and 12.5%). With EQW and MDP, the loss would have been around 15%.

Summary and outlook

For investors who may only enter into long positions, the lower tail dependencies are a more suitable measure of diversification than covariances and correlations – because they only consider the simultaneous occurrence of losses (but not of gains) and, furthermore, do not require problematic distribution assumptions. The advantages of this approach were also confirmed in a simulation study.

In the next methodology article, we will look at a diversification concept which is based on the contributions of individual assets to overall risk. For example, a portfolio could be considered diversified if no individual asset contributes more than a certain percentage to portfolio risk. The portfolio risk could, for example, be measured with the standard deviation, but also with a one-sided risk measure such as the CVaR.

Dr. Bernhard Pfaff, Portfolio Manager,
Global Asset Allocation, Invesco Global Strategies

Notes

1 Overviews can be found in, for example, Coles et al. (1999) and Heffernan (2000).

2 For the sake of completeness, it should be mentioned that there are no tail dependencies with a Gaussian copula, and that although the Student t-copula implies tail dependencies because of its symmetry, the coefficient measures the degree of dependence on both the left and right end of the distribution.

3 An overview is given by, for example, J. Dobrić and F. Schmid (2005); Frahm et al. (2005); Schmidt and Stadtmüller (2006).

4 See B. Pfaff, New approaches of portfolio optimisation: Part 6, Risk & Reward Q3/2012, pp. 20-24.

5 See Malevergne and Sornette (2008).

6 The tail dependence coefficient should always be supplemented by a (one-sided or two-sided) security-specific measure of risk, since it does not indicate the level of risk of an individual asset.

7 All calculations were made with the free statistical programming environment R 2.15.0 (R Development Core Team, 2010) and the package portfolio (Würtz et al., 2010), FRAPO (Pfaff, 2012), Hmisc (Harrell and et al., 2010), lattice (Sarkar, 2008) and Performance Analytics (Carl and Peterson, 2010).

Bibliography

Carl, P. and B. Peterson (2010): PerformanceAnalytics: Econometric tools for performance and risk analysis. R package version 1.0.3.2.

Coles, S., J. Heffernan, and J. Tawn (1999): Dependence measures for extreme value analysis. Extremes 2(4), 339–365.

Frahm, G., M. Junker, and R. Schmidt (2005): Estimating the tail dependence coefficient: Properties and pitfalls. Insurance: Mathematics and Economics 37(1), 80–100.

Harrell, F. and et al. (2010): Hmisc: Harrell Miscellaneous. R package version 3.8-3.

Heffernan, J. (2000): A directory of coefficients of tail dependence. Extremes 3(3), 279–290.

J. Dobrić and F. Schmid (2005): Nonparametric estimation of the lower tail dependence λL in bivariate copulas. Journal of Applied Statistics 32(4), 387–407.

Malevergne, Y. and D. Sornette (2008): Extreme Financial Risks – From Dependence to Risk Management. Berlin, Heidelberg: Springer-Verlag.

Pfaff, B. (2012): Financial Risk Modelling and Portfolio Optimisation with R. London: John Wiley & Sons, Ltd. (forthcoming).

R Development Core Team (2010): R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Sarkar, D. (2008): Lattice: Multivariate Data Visualization with R. New York: Springer. ISBN 978-0-387-75968-5.

Schmidt, R. and U. Stadtmüller (2006): Nonparametric estimation of tail dependence. The Scandinavian Journal of Statistics 33, 307–335.

Würtz, D., Y. Chalabi, W. Chen, and A. Ellis (2010, April): Portfolio Optimization with R/Rmetrics. Rmetrics Association & Finance Online, www.rmetrics.org. R package version 2110.79.

All analysis shown is based on simulated portfolios.

New approaches to portfolio optimisation: Part 8

The more evenly risk is distributed across individual portfolio components, the better diversified the portfolio. This is the main idea behind the diversification concepts introduced in this article.

In the two previous parts of this series, we dealt with the topic portfolio diversification and introduced concepts like the *most diversified portfolio* (MDP) and the *minimum tail-dependent portfolio* (MTD). Here, we will cover another diversification concept, the *equal risk contribution portfolio* (ERC) – also known as a risk parity strategy – as formalized by Qian.[1] His starting point are the contributions of individual financial instruments to the standard deviation risk of the portfolio. According to Qian, a portfolio's diversification increases as the differences between these individual risk contributions diminish – to put it differently: a portfolio is more diversified the more evenly the load (risk) is distributed across all risk-bearing components (the individual securities).

A similar concept is put forward by Boudt, who in place of the contributions to the standard deviation measures the contributions to a one-sided (i.e. focussed only on expected loss) risk measure like CVaR. Below, we present two variations of this: the *minimum CVaR concentration portfolio* (MCC) and the *budgeted CVaR contribution portfolio* (BCC).[2]

Portfolios with restrictions on risk contributions
Our examination starts with the risk contribution of the *i*-th financial instrument to the total risk of a portfolio, where $i = 1, \dots, N$:

$$(1) \quad C_i M_{\omega\in\Omega} = \omega_i \frac{\partial M_{\omega\in\Omega}}{\partial \omega_i}$$

In this formula, $M_{\omega\in\Omega}$ represents a linearly homogeneous risk measure and ω_i stands for the portfolio weighting of the i-th financial instrument. $M_{\omega\in\Omega}$ can be used for either the standard deviation of the portfolio or as a one-sided risk measure such as VaR or CVaR, since all three measures constitute linearly homogeneous functions of the portfolio weighting.

According to Euler's homogeneous function theorem, total risk is equal to the sum of the risk contributions from equation (1). The risk contributions can also be expressed as proportions of overall risk, by dividing the total by $M_{\omega\in\Omega}$:

$$(2) \quad \%C_i M_{\omega\in\Omega} = \frac{C_i M_{\omega\in\Omega}}{M_{\omega\in\Omega}} \times 100$$

We employ this relative expression in our empirical analysis.

Equal contributions to standard deviation risk...
The formula for the ERC portfolio is arrived at through the application of portfolio standard deviation

$$\sigma(\omega) = \sqrt{\omega'\Sigma\omega} \quad \text{for } M_{\omega\in\Omega},$$

where ω represents the ($N \times 1$) weighting factor and Σ the variance-covariance matrix of returns. The minor diagonal elements of this matrix are σ_{ij} (covariance of returns from the i-th and j-th financial instruments), the main diagonal elements are σ_i^2 (variance of returns from the i-th financial instruments). This results in the following partial derivatives:

$$(3) \quad \frac{\partial\sigma(\omega)}{\partial\omega_i} = \frac{\omega_i\sigma_i^2 + \sum_{i\neq j}^{N}\omega_j\sigma_{ij}}{\sigma(\omega)}$$

These N partial derivatives are proportional to the i-th line of $(\Sigma\omega)_i$. Accordingly, the optimisation problem in an ERC portfolio with additional long-only and budget restrictions can be formulated as follows:

$$(4) \quad P_{\text{ERC}}: \omega_i(\Sigma\omega)_i = \omega_j(\Sigma\omega)_j, \ \forall i, j$$
$$0 \le \omega_i \le 1, \ for\ i = 1, \dots, N$$
$$\omega'\mathbf{i} = 1$$

where i is the ($N \times 1$) unit vector. The solution is arrived at using numerical optimisation (sequential quadratic programming).

One closed-form solution presupposes identical correlation of all conceivable pairs of portfolio values. Then, the weighting of the i-th financial instrument is determined as the inverse of relative volatility. More volatile financial instruments receive a lower portfolio weighting, more stable ones are weighted higher.[3] In this case, the ERC solution is identical with the Maximum Sharpe Ratio Portfolio (MSR).

By definition, the standard deviation risk of the ERC portfolio is greater than that of a minimum-variance solution like the Global Minimum Variance Portfolio (GMV). But, it is lower than that of a naïve solution like the Equal Weighted Portfolio (EQW).

...and CVaR risk
Below, we determine the risk contributions of individual financial instruments using CVaR as risk measure in place of standard deviation. By applying the CVaR function for $M_{\omega\in\Omega}$ in equation (1), the CVaR contribution of the i-th financial instrument becomes:

$$(5) \quad C_i\,\text{CVar}_{\omega\in\Omega,\alpha} = \omega_i \frac{\partial\,\text{CVar}_{\omega\in\Omega,\alpha}}{\partial\omega_i}$$

Now, confidence level α is added as a further parameter to the formula. It must be either specified a priori for all quantile-based risk measures or determined in the context of the optimisation.

Calculation of partial derivatives for these risk contributions is dependent on the assumed joint distribution function of returns. Under the assumption of normal distribution (or generally under the assumption of elliptical joint distribution), the partial derivatives can be computed quite easily. Consequently, either normal distribution or the one-sided risk measure derived from the Cornish-Fisher expansion are used for practical application.

Under the normal distribution assumption, portfolio CVaR is:

$$(6) \quad \mathrm{CVar}_{\omega\in\Omega,\alpha} = -\omega'\mu + \sqrt{\omega'\Sigma\omega'}\,\frac{\phi(z_\alpha)}{\alpha}$$

where the ($N \times 1$) vector μ represents expected returns, ϕ the density function of the standard normal distribution and z_α the α-quantile of the standard normal distribution. The marginal contribution of the i-th financial instrument to portfolio CVaR is thus:

$$(7) \quad C_i\,\mathrm{CVar}_{\omega\in\Omega,\alpha} = \omega_i\left[\mu_i + \frac{(\Sigma\omega)_i}{\sqrt{\omega'\Sigma\omega'}}\,\frac{\phi(z_\alpha)}{\alpha}\right]$$

Scaillet (2002) demonstrated that the contribution to CVaR is equivalent to the contribution to portfolio loss if the loss contributions are higher than the VaR at a given confidence level α.

Note also that the portfolio allocation – under the assumption $\mu = 0$ with equal CVaR contributions – is always equivalent to that of an ERC portfolio. The assumption of zero expected returns is founded in the random walk hypothesis. Under this premise, the first term in equation (7) is zero and the target function deviates from that of an ERC optimisation only through inclusion of the constant term $\phi(z_\alpha)/\alpha$, one which has no effect on the solution.

Compared to the ERC optimisation, this approach offers two advantages: for one, a coherent risk measure is applied, and secondly, restrictions can be set with respect to portfolio CVaR contributions. The optimisation can even be centered around the restrictions, for instance, by using a MinMax formula as found in Boudt et al. (2011a):

$$(8) \quad C_{\omega\in\Omega,\alpha} = \min\{\max C_i\,\mathrm{CVaR}_{\omega\in\Omega,\alpha}\}$$

The authors refer to this approach – which minimises the maximum CVaR contribution – as a Minimum CVaR Concentration (MCC) optimisation. It results in uniform CVaR contributions as a solution with interval budgets (Budgeted CVaR Contribution or BCC optimisation).

Empirical application*

We will now apply the three optimisation approaches (ERC, MCC and BCC) to a global mixed portfolio.

The financial instruments will comprise equities and fixed income paper from the US, UK, Japan and the eurozone, as well as a commodities index. The equity markets will be covered via the S&P 500, FTSE 100, TOPIX and EuroStoxx[4], the bond markets via US, UK, Japanese and German 10yr government bonds and the commodity markets via the GSCI performance index.

The optimisations are based on monthly returns from February 1991 to October 2012. No short positions are included and currency risks have been abstracted out. For the BCC optimisation, the maximum CVaR contribution was set to 20% at a confidence level of 95%, which was also applied for the MCC optimisation.

As in the previous parts of this series, we have separated the overall sample into two phases. Using the returns from February 1991 to October 2007, we optimized the portfolios; on this basis, we then determined their performance in the period November 2007 to October 2012. By way of comparison, we will also look at the performance of the evenly weighted EQW portfolio (which starts with all financial instruments constituting 1/9 or roughly 11%) as well as that of the MDP portfolio.[5]

Table 1 shows the portfolio shares of the nine financial instruments; figure 1 shows the relative weightings compared to an EQW portfolio.

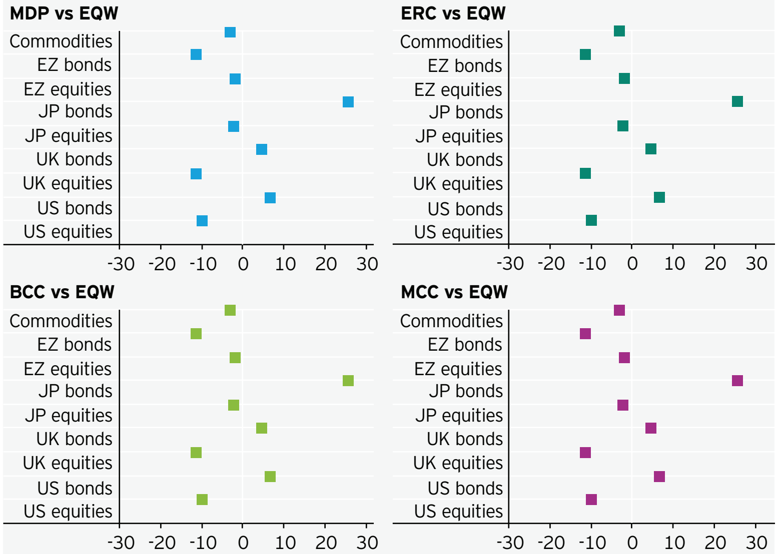
As can be seen, the MDP portfolio does not contain UK equities or German bonds, and the share of US equities is small, at about 1.55%. Instead, there is a markedly larger weighting of US, UK and Japanese bonds, eurozone equities and commodities than in

Table 1: Portfolio allocations

Markets	MDP	ERC	BCC	MCC
US equities	1.55	6.13	6.91	6.22
US bonds	18.23	14.25	13.87	4.27
UK equities	0.00	5.83	5.65	4.87
UK bonds	15.93	13.70	11.54	13.81
JP equities	9.29	6.48	0.77	5.96
JP bonds	37.17	24.12	27.84	35.29
EZ equities	9.41	5.33	7.18	2.70
EZ bonds	0.00	17.31	22.49	22.94
Commodities	8.41	6.85	3.76	3.92

Source: Invesco. For illustrative purposes only.

Figure 1: Portfolio allocations compared



EQW: Equal-weighted portfolio; MDP: Most diversified portfolio; ERC: Equal risk contribution portfolio; BCC: Budgeted CVaR contribution portfolio; MCC: Minimum CVaR concentration portfolio.
The charts show the active positions of the other portfolios relative to the EQW.
Source: Invesco. For illustrative purposes only.

the ERC, BCC and MCC portfolios. Except for the weighting of Japanese equities, the ERC and BCC portfolios are similar. There are virtually no Japanese equities in the BCC portfolio, but in the ERC approach they have a meaningful weighting.

We also calculated risk – measured as standard deviation risk and modified Expected shortfall (ES) at 95% confidence – as well as the diversification and concentration ratios (table 2).[6]

Table 2: Portfolio metrics

Statistic	EQW	MDP	ERC	BCC	MCC
Standard deviation	1.802	1.227	1.228	1.199	1.118
ES (modified, 95%)	3.415	1.941	1.913	1.759	1.580
Diversification	1.849	2.180	2.094	1.926	2.040
Concentration	0.136	0.167	0.115	0.129	0.138

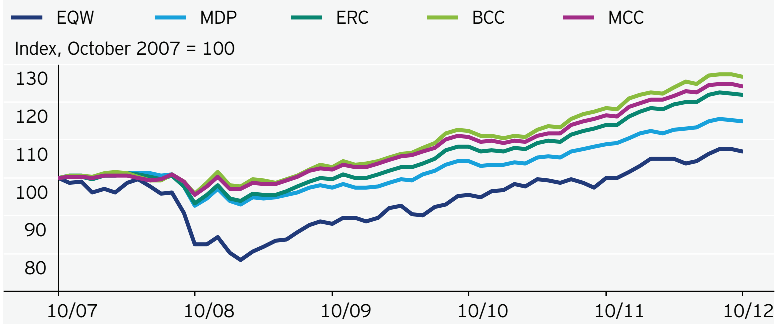
Source: Invesco. For illustrative purposes only.

Table 3: Contributions to the ES (at 95% confidence)

Markets	EQW	MDP	ERC	BCC	MCC
US equities	21.653	2.617	13.284	14.349	12.339
US bonds	-2.947	11.163	10.771	14.230	4.055
UK equities	20.503	0.000	12.969	12.616	10.965
UK bonds	-1.555	6.493	9.924	11.952	13.455
JP equities	23.524	26.330	16.028	0.972	13.999
JP bonds	-1.640	11.475	1.006	5.646	15.644
EZ equities	29.002	21.749	14.886	19.424	7.241
EZ bonds	-1.967	0.000	8.009	16.298	16.150
Commodities	13.426	20.174	13.122	4.512	6.152

Source: Invesco. For illustrative purposes only.

Figure 2: Absolute performance



Source: Invesco, as at 31 October 2012. For illustrative purposes only.

Standard deviation risk and ES are highest in the case of the EQW approach; MDP and ERC are in the middle; the BCC and MCC portfolios carry the least risk ex post.

By definition, the MDP portfolio has the highest diversification ratio, followed closely by the ERC solution, while the BCC (and EQW) exhibits the lowest degree of diversification.

When it comes to the concentration ratio, on the other hand, that is lowest in the ERC portfolio and highest for the MDP approach – presumably because it leaves out a number of financial instruments.

Table 3 shows the percentage contributions to modified ES at a confidence level of 95%.

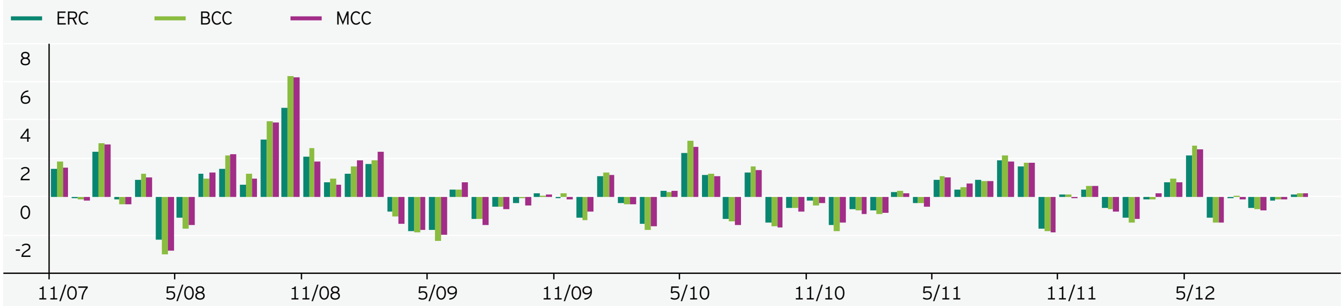
Note the negative ES contributions from bonds of all countries in the evenly weighted EQW portfolio, by which they serve a hedging purpose. It can also be seen that more than 90% of portfolio ES comes from the equity portion. There is also risk bundling evident under MDP: as much as two-thirds of portfolio ES derives from Japanese and eurozone equities and commodities. For the ERC, BCC and MCC solutions, the ES contributions are more evenly distributed; the maximum ES contribution was lowest under the ERC allocation, at just over 16% (Japanese equities), followed a close second by the MCC solution (slightly above 16% for eurozone bonds).

Figures 2, 3 and 4 show absolute and relative performance second phase based on the original allocation in table 1.

The terminal values of the three new approaches are higher than those of the EQW and MDP portfolios (figure 2). At the outset of the subprime crisis, the EQW portfolio sees a dramatic decline – like all other allocations, which also saw (albeit smaller) declines. At the start of 2009, all five portfolios then begin a major comeback.

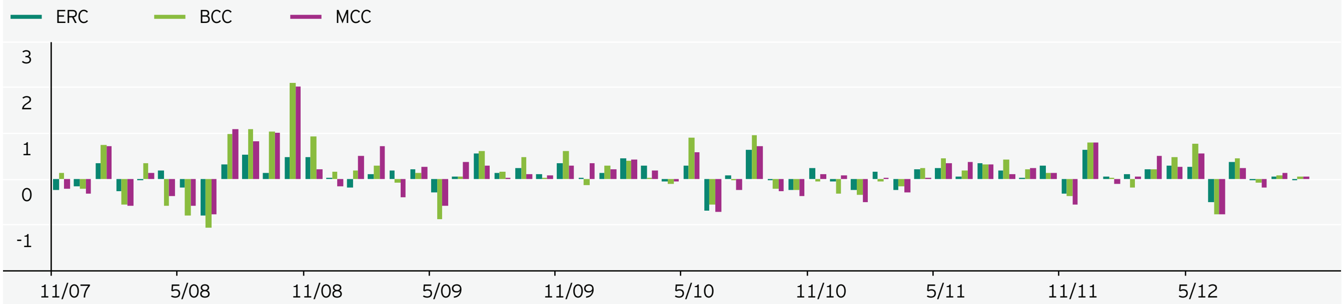
Looking at the differences in returns of the ERC, BCC and MCC vs. EQW and MDP (figures 3 and 4), the following is evident: First, the ERC, BCC and MCC returns are higher in roughly half of the months and lower in the other half than under the EQW allocation. Second, ERC, BCC and MCC have a markedly larger advantage over EQW than over

Figure 3: Performance relative to EQW



Source: Invesco, as at 31 October 2012. For illustrative purposes only.

Figure 4: Performance relative to MDP



Source: Invesco, as at 31 October 2012. For illustrative purposes only.

MDP. Third, returns under the three new concepts are higher in around two-thirds of all months than under the MDP portfolio.

Summary and outlook

In this article, we interpreted diversification as the most uniform possible distribution of portfolio risks across individual financial instruments. We defined portfolio risk as both standard deviation of portfolio returns (ERC) and CVaR (BCC and MCC). As a rule, none of the three approaches permits analytical solutions, so that numerical optimisation methods are required. The empirical simulation, however, supports the hypothesis that approaches aiming at uniform risk contributions can have advantages.

Now that we have completed an in-depth look at the risk side of the equation over the past several articles, the next parts of the series will cover returns. We will introduce optimisation approaches that take into account differing levels of expected return.

Dr. Bernhard Pfaff, Portfolio Manager,
Global Asset Allocation, Invesco Global Strategies

Notes:

1 Qian (2005, 2006, 2011). The concept, however, has been in existence since 1996 when it was coined by the investment firm Bridgewater. Zhu et al. (2010) describe solutions with restrictions or budgets for the individual risk components.
2 Boudt et al. (2007, 2008), Peterson and Boudt (2008), Boudt et al. (2010, 2011a), Ardia et al. (2010)
3 For more information here and below, see Maillard et al. (2010).
4 All performance indices with the exception of the EuroStoxx.
5 All calculations were conducted using the free statistical program environment R 2.15.2 (see R Development Core Team, 2010) and the packages FRAPO (see Pfaff, 2013), Hmisc (see Harrell and et al., 2010), lattice (see Sarkar, 2008), PortfolioAnalytics (see Boudt et al., 2011b), PerformanceAnalytics (see Carl and Peterson, 2010) and timeSeries (see Würtz and Chalabi, 2012).
6 For the definition of diversification ratio and concentration ratio, see part 6 of this series, Risk & Reward, Q3/2012.

Bibliography

Ardia, D., K. Boudt, P. Carl, K. Mullen, and B. Peterson (2010). Differential Evolution (DEoptim) for non-convex portfolio optimization.

Boudt, K., P. Carl, and B. Peterson (2010, April). Portfolio optimization with cvar budgets. Presentation at r/finance conference, Katholieke Universteit Leuven and Lessius, Chicago, IL.

Boudt, K., P. Carl, and B. Peterson (2011a, September). Asset allocation with conditional value-at-risk budgets. Technical report, http://ssrn.com/abstract=1885293.

Boudt, K., P. Carl, and B. Peterson (2011b). PortfolioAnalytics: Portfolio Analysis, including Numeric Methods for Optimization of Portfolios. R package version 0.6.1/r1849.

Boudt, K., B. Peterson, and C. Croux (2007, September). Estimation and decomposition of downside risk for portfolios with non-normal returns. Working Paper KBI 0730, Katholieke Universteit Leuven, Faculty of Economics and Applied Economics, Department of Decision Sciences and Information Management (KBI), Leuven.

Boudt, K., B. Peterson, and C. Croux (2008). Estimation and decomposition of downside risk for portfolios with non-normal returns. The Journal of Risk 11(2), 79–103.

Carl, P. and B. Peterson (2010). Performance-Analytics: Econometric tools for performance and risk analysis. R package version 1.0.3.2.

Harrell, F. and et al. (2010). Hmisc: Harrell Miscellaneous. R package version 3.8-3.

Maillard, S., T. Roncalli, and J. Teiletche (2010). The properties of equally weighted risk contribution portfolios. The Journal of Portfolio Management 36(4), 60–70.

Peterson, B. and K. Boudt (2008, November). Component var for a non-normal world. Risk. Reprint in Asia Risk.

* All analysis shown is based on simulated portfolios.

Pfaff, B. (2013). Financial Risk Modelling and Portfolio Optimisation with R. London: John Wiley & Sons, Ltd.

Qian, E. (2005). Risk parity portfolios: Efficient portfolios through true diversification. White paper, PanAgora, Bostan, MA.

Qian, E. (2006). On the financial interpretation of risk contribution: Risk budgets do add up. Journal of Investment Management 4(4), 1–11.

Qian, E. (2011, Spring). Risk parity and diversification. The Journal of Investing 20(1), 119–127.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Sarkar, D. (2008). Lattice: Multivariate Data Visualization with R. New York: Springer. ISBN 978-0-387-75968-5.

Scaillet, O. (2002). Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14(1), 74–86.

Würtz, D. and Y. Chalabi (2012). timeSeries: Rmetrics - Financial Time Series Objects. R package version 2160.94.

Zhu, S., D. Li, and X. Sun (2010, Fall). Portfolio selection with marginal risk control. Journal of Computational Finance, 1–26.

New approaches to portfolio optimization: Part 9

In previous parts of this series, we concentrated on alternative approaches to optimizing portfolio risk. Here we examine methods which deal explicitly with expected levels of return. Specifically we introduce the Black-Litterman (BL)[1] model, an approach which is able to take into account individual return expectations.

The Black-Litterman model consists of five components. At the heart of the concept is a Bayesian estimation of the equilibrium returns, under the multivariate normal distribution assumption, and the expected levels of return. The explicit account taken of individual return expectations most likely contributed in particular towards the initially high acceptance and spread of this concept among portfolio managers.

The five components of the model are:

1. the Capital Asset Pricing Model (CAPM; Sharpe, 1964)
2. reverse optimization (Sharpe, 1974)
3. the Bayesian estimation method
4. the concept of a universal hedge ratio (Black, 1989)
5. mean-variance optimization (Markowitz, 1952)

1. The Capital Asset Pricing Model
The BL approach is based on the CAPM. In this equilibrium model for the capital market, supply and demand are equal for all assets and the market is permanently cleared. In the μ - σ space, the equilibrium corresponds to the tangent point between the capital market line and the upper branch of the efficient frontier.

The equilibrium solution, the Maximum Sharpe Ratio Portfolio, is efficient if account is taken of the return of the risk-free asset.[2] Knowledge of the prices and, by extension, the market capitalizations ω_{MKT} then makes it possible to determine the implied returns of every asset.

2. Reverse optimization
The second component is reverse optimization – "reverse" because, unlike elsewhere, no optimal allocation is determined for pre-defined return expectations; instead, a calculation is made of the implied return expectations on which the actual allocation is based.

We begin by making a straight forward calculation, free of restrictions, to establish which weighting vector ω maximizes a quadratic utility function:

(1) $$\arg\max_{\omega \in \Omega} \omega'\mu - \frac{\lambda}{2}\omega'\Sigma\omega$$

μ denotes the expected excess returns, $\omega'\Sigma\omega$ the standard deviation risk and $\lambda > 0$ the risk aversion parameter. The utility maximizing allocation, ω_{OPT}, is then:

(2) $$\omega_{OPT} = \left[\lambda\Sigma\right]^{-1}\mu$$

Usually it does not concur with the allocation of the market portfolio from the CAPM, ω_{MKT}. As a rule, the two allocations are only identical if the expected excess returns are the same.

The reverse optimization comes into play if we equate ω_{OPT} with ω_{MKT} and use the formula to determine the implied excess returns of the market portfolio. Left multiplication of equation (2) by $\lambda\Sigma$ results in:

(3) $$\pi = \lambda\Sigma\omega_{MKT}$$

with π denoting the implied equilibrium returns of the market portfolio.

3. The Bayesian estimation method
The implied equilibrium returns flow into the third component of the BL model, the Bayesian estimation method. The returns show an a-priori distribution with expected value π. An investor with no individual return expectations would therefore do best to replicate the market portfolio ω_{MKT}.

The question now is how to merge an investor's individual return expectations with the implied market expectations. In the BL approach, the Bayesian a-posteriori distribution is used to derive this. *A* denotes the investor's (multivariate) return expectations, and *B* the implied excess returns in equilibrium. According to the Bayes' theorem, the joint distribution, P(*A*, *B*), is then:

(4) $$\begin{aligned} P(A,B) &= P(A|B)P(B) \\ &= P(B|A)P(A) \end{aligned}$$

As we are interested in the conditional return distribution in relation to the equilibrium returns,

(5) $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

is derived by equating and solving the two equations in (4).

If we use E(**r**) to denote the expected excess returns, equation (5) can also be written as:

(6) $$P(E(\mathbf{r})|\pi) = \frac{P(\pi|E(\mathbf{r}))P(E(\mathbf{r}))}{P(\pi)}$$

With the BL approach, it is not necessary to formulate return expectations for all assets. A further point is that the return expectations can be shown both in absolute terms as well as in relation to other financial instruments. Thus, using the normal distribution the individual expectations are modelled as:

(7) $$P\,E(\mathbf{r}) \sim \mathcal{N}(\mathbf{q},\Omega)$$

with P a ($K \times N$) pick matrix, **q** the ($K \times 1$) vector for the absolute/relative return expectations (also known as "views") and Ω a ($K \times K$) diagonal matrix denoting the uncertainty of the individual expectations. As all minor diagonal elements of the matrix Ω equal zero, under the normal distribution assumption this implies that the return expectations are unrelated to each other.[3]

The distribution of the conditional equilibrium returns for given return expectations is then:

$$(8) \quad \pi \big| \mathrm{E}(\mathbf{r}) \sim \mathcal{N}\left(\mathrm{E}(\mathbf{r}), \tau\Sigma\right)$$

We used $\mathrm{E}(\pi) = \mathrm{E}(\mathbf{r})$ here as, in equilibrium, the expected value of the individual return estimates is commensurate with the market returns. The scalar τ indicates the market's uncertainty over the implied excess returns.

Hence, the a-posteriori distribution of the conditional random variables $\mathbf{r} = \mathrm{E}(\mathbf{r})|\pi$ is multivariate normally distributed with distribution parameters:

$$(9) \quad \mathrm{E}(\mathbf{r}) = \left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}\mathbf{q}\right]$$

$$(10) \quad \mathrm{Var}(\mathbf{r}) = \left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}$$

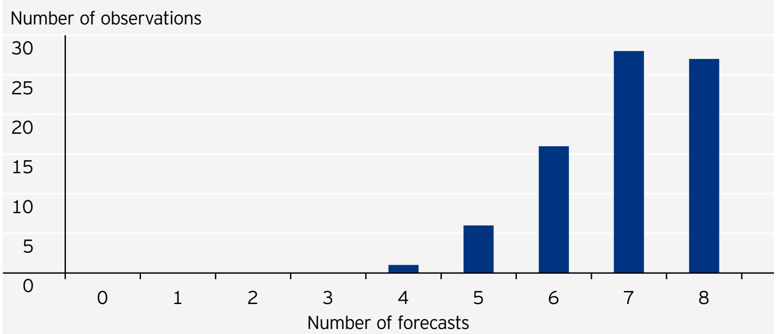
4. The concept of a universal hedge coefficient

The fourth element of the BL model, the concept of a universal hedge ratio, is already contained in equation (9). The position parameter of the normal distribution is a weighted mean from the market equilibrium returns and the individual expectations. Each of these return factors is scaled with the inverse variance-covariance matrices, the returns expected by the market $(\tau\Sigma)^{-1}$ and the individual return expectations $P'\Omega^{-1}$.

The ratio $\Omega_{kk}/\tau = \mathbf{p}'_k\Sigma\mathbf{p}$ indicates the relative significance of the market expectations and the individual expectations for the *k*th financial instrument. The right side of the equation shows the effect of the return expectations for the *k*th asset on the portfolio variance. This equation can also be used to implicitly determinate the confidence level Ω_{kk} at predefined τ. The advantage is that the point estimation for $\mathrm{E}(\mathbf{r})$ is unrelated to the assumed τ.[4]

If we consider the marginal case of no individual return expectations of any kind ($P = 0$), the market expectations are the point estimations for $\mathrm{E}(\mathbf{r})$. The other marginal case is deemed to exist if the investor has no doubts over his individual expectations

Figure 1: Frequency distribution of the return expectations



Source: Invesco. Based on 78 in-example periods of 96 observations each, the first one ending in September 2006, the last one ending in February 2013.

(technically formulated: if they are given without any uncertainty). In this case, the point estimations for $\mathrm{E}(\mathbf{r})$ correspond to the individual expectations.

5. Mean-variance optimization

The distribution parameters determined by equations (9) and (10) can be used, for example, in a mean-variance portfolio optimization. An interesting point is that only the portfolio weights for which there are individual return expectations deviate from their market capitalization weights. An investor who is more upbeat about an asset than the market will hold more of this asset without reducing his exposure to other assets – assuming, of course, there are no budget restrictions.

Application

In a comparative simulation study, we now apply the BL approach to a mixed portfolio of stocks and fixed income. The portfolio comprises generic futures on the S&P 500, the DAX, the FTSE 100 and the Nikkei 225 for stocks. The bond market is modelled by futures on 10-year sovereign debt issues from the US, Germany, the UK and Japan.[5] We use the month-end values from October 1998 to March 2013, i.e. a total of 174 observation vectors. For the portfolio optimization, a Maximum Sharpe Ratio model was chosen with a non-negativity restriction for the weights and a budget restriction.

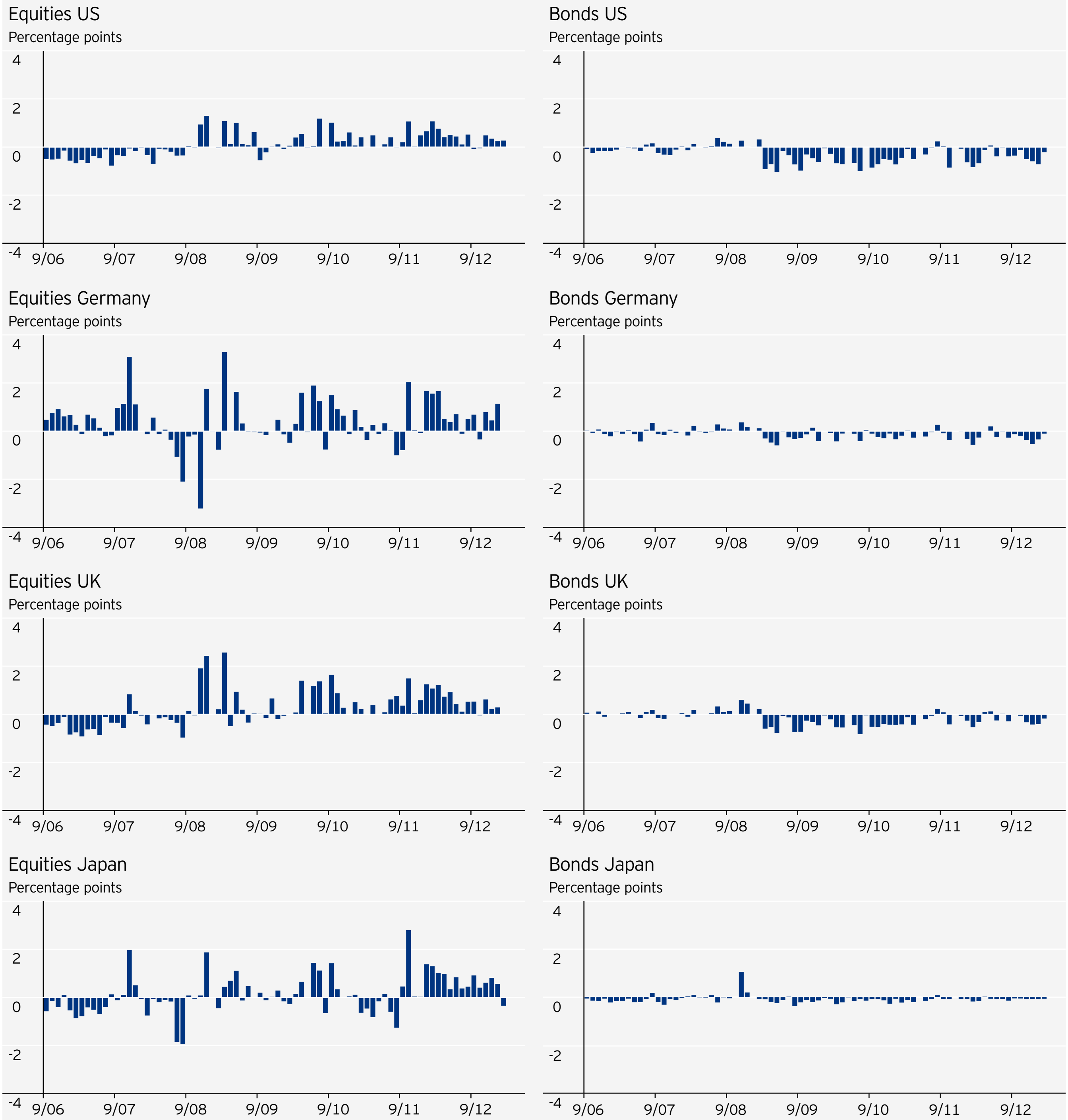
The simulation is based on a recursive estimate of the position and dispersion parameters. The first in-sample period was from October 1998 to September 2006, i.e. 8 years or 96 observation sets long. Based on this and the subsequent data sets, the a-priori and a-posteriori estimates for the return vectors and variance-covariance matrices were determined. The individual return expectations for the eight financial instruments were calculated on the basis of the (ex ante) one-step forecasts of a vector error correction model (VECM) at a cointegration rank of 2.[6]

The VECM for the logarithms of the prices was specified with one lag in the endogenous variables and a deterministic constant in the co-integration relations. The recursive estimates were also based on this model structure. Account was only taken of an individual return expectation for a future if its last month-end value stood outside the forecast interval at a confidence level of 5%. The idea behind this is first, to only take account of forecasts with a certain weight while second, generating a sufficient number of individual return expectations. Using this rule, a total of 542 return expectations were generated, based on 78 in-sample periods; the first one ending in September 2006 and the last one ending in February 2013. In monthly average terms, this corresponds to around six out of eight maximum possible forecasts. Figure 1 shows the frequency distribution of the forecasts.

The first element for determining the a-posteriori return vectors in accordance with equation (9) thus exists with the "view" vectors **q**.

A further component of this equation is the weighting scalar $\tau \in [0, 1]$. For $\tau \to 0$, the portfolio allocation increasingly depends on the a-prori returns, for $\tau \to 1$ increasingly on the individual

Figure 2: Differences between a-priori and a-posteriori returns



Source: Invesco. Based on 78 in-sample periods of 96 observations each, the first one ending in September 2006, the last one ending in February 2013.

return expectations. In our simulation study we model the sensitivity of this parameter reading on the basis of portfolio indicators and the portfolio progression. To do this we set the values for τ during the entire pseudo-ex ante period at 0.1, 0.5 and 1.0.[7]

Furthermore, we were the first to determine τ on the basis of an adaptive rule. A crucial factor in this concept is the latest portfolio return. But since the portfolio return $r_p \in \Re$ is $\tau \in [0, 1]$, the real numerical sequence has to be projected into the interval [0, 1]. One way to do this is by a logistic function (e.g. the distribution function of the normal distribution), but the sequence can also be asymmetrically modelled to allow implicit mapping of various degrees of risk aversion. For the normal distribution function, τ will be smaller than 0.5 for negative portfolio returns so that the individual return expectations would then only have a slight

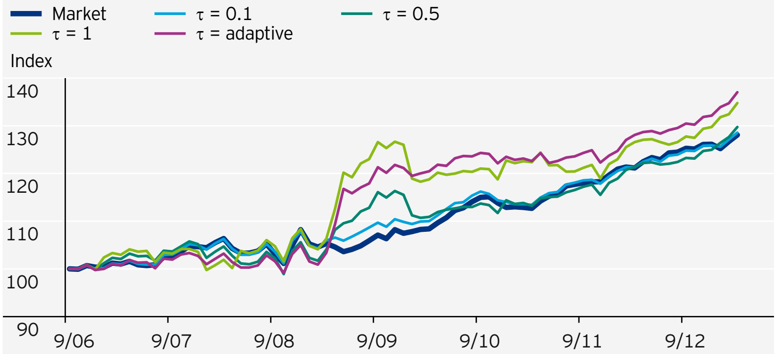
Table 1: Comparison of portfolio allocations

In %		Equities				Bonds			
		USA	Germany	UK	Japan	USA	Germany	UK	Japan
Market	Minimum	0.00	3.38	0.00	0.00	23.85	0.00	0.00	39.86
	Mean	0.21	6.66	0.00	1.89	33.83	10.99	0.00	46.42
	Maximum	1.90	8.03	0.00	6.18	47.98	27.39	0.00	52.87
τ = adaptive	Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	5.30	9.01	4.93	5.15	20.80	20.17	1.34	33.30
	Maximum	98.96	72.44	34.53	97.58	93.82	83.27	35.07	89.73

Source: Invesco. Based on 78 in-sample periods of 96 observations each, the first one ending in September 2006, the last one ending in February 2013.

influence on the portfolio allocation. By the same token, greater weight is attached to the individual expectations in the case of positive portfolio returns. Within the framework of an adaptive learning process, this offers a degree of protection against forecasting errors.

Figure 3: Portfolio development



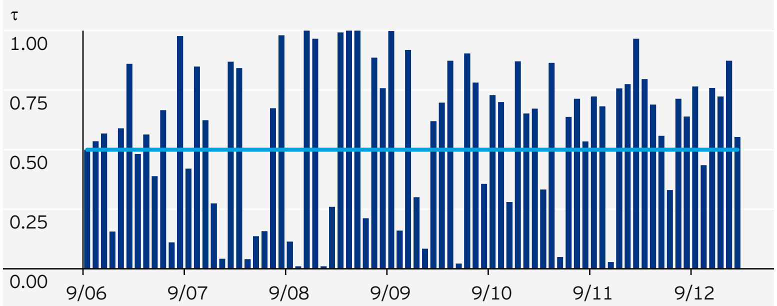
Source: Invesco. Backtest period September 2006 to March 2013.

Table 2: Portfolio ratios

Statistics	Market	τ = 0.1	τ = 0.5	τ = 1.0	τ = adaptive
Max. drawdown	4.92	4.87	6.47	6.63	4.04
ES (modified, 99%)	3.46	4.35	4.76	7.38	2.40
Sharpe ratio	0.09	0.08	0.07	0.05	0.17

Source: Invesco. Backtest period September 2006 to March 2013.

Figure 4: Development of τ



Source: Invesco. Backtest period September 2006 to March 2013.

Figure 2 highlights the differences in the return levels between the BL portfolio (with adaptively determined values for τ and filtered point forecasts for the VEC models) and the a-priori distributions (with random estimators for the a-priori returns).

The return differences lie at between -4 and +4 percentage points which suggests large allocation differences between the market portfolio and the BL approach (table 1). Differences can also be seen in the portfolio developments of the two concepts (figure 3, table 2).

Figure 3 clearly shows that, when the subprime crisis was at its peak in the years 2007 to 2009, the portfolio value would have moved more or less sideways in the case of all five optimization parameterizations. At the beginning, the performances of the BL models of τ = 0.1 and adaptive adjustment and the market allocation were virtually identical. From late autumn 2007 onwards, the BL optimizations in which the individual expectations were given a higher weight dropped back slightly only to reverse again as the equity market started to rebound. This was when it paid off to attach a higher weight to the individual return expectations (τ = 0.5 and τ = 1.0) as they had been more optimistic for stocks.

For τ = 1.0 and, with reservations, for τ = 0.5, the performance would have been extremely volatile. The advantages of an adaptive adjustment become apparent from spring 2009. While the extreme gain for τ = 1.0 is not fully tracked, the loss at year-end 2010 turned out significantly lower than with τ = 0.5 and τ = 1.0. The situation at year-end 2012 was no different. Hence adaptively setting τ can indeed offer a degree of protection against forecasting errors. Figure 4 shows the development of the adaptively adjusted value for τ; in approximately two thirds of the cases, a positive portfolio return was achieved (τ ≥ 0.5).

All versions of the BL approach culminate in higher end values than the market portfolio. This can probably be attributed to a more active portfolio allocation in connection with the individual return expectations which, however, also go hand in hand with higher portfolio reallocations.

Added to this, the BL portfolio with τ = 1 shows the highest drawdown (table 2). Holding all other factors constant, forecasting errors have a greater impact here than with other optimization models. With τ = 0.5, the drawdown is similarly high. On the other hand, the ratios for the market portfolio and the BL approach are similar at $\tau = 0.1$; the difference is only five basis points. The strategy with adaptive setting of τ scores the best.

For the one-sided risk measure, we also computed the modified expected shortfall for the back-test period at a confidence level of 99%. The smaller the influence of the individual return expectations, the lower the ES in most cases. However, an exception is the adaptive-τ strategy which shows the lowest risk. In conjunction with the highest portfolio end value, this also culminates in a far higher Sharpe ratio.

Summary and outlook

In this article, we have introduced the BL model which takes into account individual return expectations. However, the weightings of the market expectations and individual expectations are not defined a priori. We have presented a concept for determining the weights on the basis of an adaptive learning rule and tested it in a comparative simulation study. The new concept achieved the highest Sharpe ratio and the lowest drawdown.

In the next part of this series, we will present the concept of copula-opinion pooling which can be viewed as a continuation of the BL approach.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Strategies

Notes:

1 For more information here and below, see Black and Litterman (1990, 1991, 1992) and He and Litterman (2002).
2 See Risk & Reward, Q2/2011 and the literature quoted there.
3 However, this implication is questionnable. It is illogical to simultaneously assume a joint return distribution of market returns and unrelated individual return expectations. In a later contribution to this series, we will introduce an extension to the BL approach which avoids this inconsistency.
4 He and Litterman (2002)
5 Thomson Reuters DataStream was used as the data source. The mnemonic sequence of the time series is as follows: ISPCS04, GDXCS04, LSXCS04, ONACS04, CTYCS04, GGECS04, LIGCS04 and JGBCS04.
6 All the calculations were carried out using a free statistical programming environment R 3.0.1 (R Core Team, 2013) as well as the packages BLCOP (see Gochez, 2011), fPortfolio (Würtz et al., 2010), Hmisc (see Harrell and et al., 2010), PerformanceAnalytics (Carl and Peterson, 2013) and urca (Pfaff, 2008a) and vars (Pfaff, 2008b).
7 Differing views on the value to be used for τ can be found in publications on this subject. See He and Litterman et al (2002); Meucci (2010); Satchell and Scowcroft (2000); Walters (2009, 2010). The suggestion was also made (Walters, 2010) to take the inverse of the sample size for τ – since τ expresses the uncertainty over the a-priori returns which, however, can be estimated with ever increasing accuracy the greater the sample size. This rule more or less corresponds to the portfolio solutions based on a-priori distribution since with $\tau \in [0.006; 0.01]$ (as assumed here) the individual expectations play only a subordinate role.

Bibliography

- Black, F. (1989). Universal hedging: Optimizing currency risk and reward in international equity portfolios. Financial Analysts Journal, 16–22.
- Black, F. and R. Litterman (1990). Asset allocation: combining investor views with market equilibrium. Technical report, Goldman Sachs Fixed Income Research.
- Black, F. and R. Litterman (1991). Global asset allocation with equities, bonds, and currencies. Technical report, Goldman Sachs Fixed Income Research.
- Black, F. and R. Litterman (1992). Global portfolio optimization. Financial Analysts Journal 48(5), 28–43.
- Carl, P. and B. Peterson (2013). PerformanceAnalytics: Econometric tools for performance and risk analysis. R package version 1.1.0.
- Gochez, F. (2011). BLCOP: Black-Litterman and copula-opinion pooling frameworks. R package version 0.2.6.
- Harrell, F. (2010). Hmisc: Harrell Miscellaneous. R package version 3.8-3.
- He, G. and R. Litterman (2002). The intuition behind black-litterman model portfolios. Working paper, Goldman Sachs Group, Inc. Quantitative Strategy Group, http://ssrn.com/abstract=334304.
- Markowitz, H. (1952). Portfolio selection. The Journal of Finance 7(1), 77–91.
- Meucci, A. (2010). The Encyclopedia of Quantitative Finance, Chapter The Black-Litterman Approach: Original Model and Extensions. John Wiley & Sons.
- Pfaff, B. (2008a). Analysis of Integrated and Cointegrated Time Series with R (Second ed.). New York: Springer. ISBN 0-387-27960-1.
- Pfaff, B. (2008b). Var, svar and svec models: Implementation within R package vars. Journal of Statistical Software 27(4).
- R Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Satchell, S. and A. Scowcroft (2000). A demystification of the black-litterman model: Managing quantitative and traditional portfolio construction. Journal of Asset Management 1(2), 138–150.
- Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium. Journal of Finance, 425–442.
- Sharpe, W. (1974). Imputing expected security returns from portfolio composition. Journal of Financial and Quantitative Analysis, 463–472.
- Walters, J. (2009, February). The black-litterman model in detail. http://www.blacklitterman.org.
- Walters, J. (2010). The factor tau in the black-litterman model. http://www.blacklitterman.org.
- Würtz, D., Y. Chalabi, W. Chen and A. Ellis (2010). Portfolio Optimization with R/Rmetrics. Rmetrics Association & Finance Online, www.rmetrics.org. R package version 2130.80.

New approaches to portfolio optimization: Part 10

In the last contribution to this series we introduced the Black-Litterman model. Unlike older concepts it is able to take account of individual return expectations but also has its disadvantages due to its restrictive model assumptions. The *Copula Opinion Pooling* introduced by Meucci offers a solution.

There can be little doubt that the possibility of taking relative and absolute return expectations into account when optimizing a portfolio was one decisive reason for the huge popularity of the Black-Litterman model (BL model) among portfolio managers. However, the concept also has its disadvantages which should not be overlooked: for example, the expected returns in the BL model also depend on the definition of the market portfolio, and in this series we have made repeated reference to the forever problematic assumption of joint normally distributed returns in capital market models. Not only does this contradict reality, but it also means that consideration can only be given to mean and standard deviation expectations.

In this article we introduce the *Copula Opinion Pooling* (COP) model propagated by Meucci (2005, 2006b, 2006a, 2010c, 2010a, 2010b). The COP model avoids the weaknesses of the BL model by being more flexible and offering further means of specification.

Copula Opinion Pooling

Both models share the combination of the *a-priori* distribution of the market and the investor's individual expectations into an *a-posteriori* distribution which forms the basis of the allocation derived by portfolio optimization. The main difference between the two models is the manner in which the *a-priori* distribution of the market and the subjective expectations are to be specified. Nor is it possible to derive the *a-posteriori* distribution with the COP model in closed form, making Monte Carlo simulations necessary.

In the BL model, the market returns are determined using the CAPM equilibrium model; the return expectations for the assets can be specified either as absolute or relative return forecasts, based on the assumption of normally distributed market returns and return expectations. In practice, however, these core assumptions are too restrictive. This is the point at which the COP model takes over. It allows market returns to be more flexibly modeled and also creates possibilities of specifying expectations.

With the COP model, an *a-priori* distribution of the returns on the financial assets need not necessarily be defined. A factor-based risk model, with an accompanying statistical distribution, can also be used as the starting point; the *a-priori* distribution can be stated for any ($N \times 1$) random vector. The assumption of joint normally distributed random variables can also be dispensed with, and the copula concept used instead. The copula is a function which allows a joint probability distribution to be derived for random variables with any marginal distributions using Monte Carlo simulation.

For the statistical modeling of financial market returns, GARCH models could be used or absolute distributions such as the Generalized Hyperbolic Distribution, if appropriate. The dependencies between the financial assets can then be mapped using a Student's t-Copula which takes specific account of the marginal dependencies between the assets.

The second difference between the COP model and the BL model is the manner in which the return expectations ("views") are expressed. With the BL model, only the expected values of the returns are entered into the optimization – or, using the technical terms, the position parameters μ of the normal distribution. With the COP approach, however, the full distribution function is factored in. This enables the user to express interval forecasts in addition to point forecasts. Similar to the BL model, the user must determine a confidence level c_k at $k = 1,\ldots,K$ as well as the so-called *Pick* matrix for each of his K expectations.

All in all, however, the second key component of the COP model also offers the user greater flexibility than the BL model. But this increased flexibility comes at a price. Unlike with the BL model, the *a-posteriori* distribution cannot normally be derived in closed form. In most cases, a Monte Carlo simulation is needed.

The COP model consists of the following five steps (Meucci, 2006a) which we explain in more detail below:

1. The rotation of the *a-priori* distribution into the *view* space;
2. the calculation of the *view* distribution function and the *a-priori* copula;
3. the calculation of the *a-posteriori* marginal distributions of the *views;*
4. the determination of the joint *a-posteriori* distribution function of the *views*; and
5. the determination of the joint *a-posteriori* distribution of the market returns (market distribution).

The starting point of the COP model is a Monte Carlo simulation $\mathcal{M}$ for N random variables of the size J (for example, the returns of the financial instruments in a portfolio). This ($J \times N$) matrix can be obtained by, for example, generating random variables using the underlying copula model. The j^{th} line of this matrix is a potential tuple of the joint market distribution. Its characteristics are derived from the marginal distribution model selected and the dependencies between the random variables resulting from the chosen copula.

In a **first step** the ($J \times N$) simulated values $\mathcal{M}$ are transferred to the coordinates system of the *views*, $\mathcal{V}$. Formally, this corresponds to a matrix multiplication:

$\mathcal{V} = \mathcal{M}\bar{P}'$, with P representing an invertible ($N \times N$) *Pick* matrix as in the BL model. The qualitative difference between the *Pick* matrix in the BL model and the *Pick* matrix $\bar{P}$ in the COP model are the differing dimensions. In the BL model, the dimension of the *Pick* matrix was ($K \times N$), with each line corresponding to an absolute and a relative return expectation. In the COP model, by contrast, the dimension is ($N \times N$). Here, too, the first K lines of P contain the financial assets, the remaining ($N - K$) lines are filled out with the (orthogonal) complement $P^{\perp}$ of the dimension (($N - K$) $\times$ N) so that $\bar{P} = (P|P^{\perp})'$.

In a **second step**, the first K columns of $\mathcal{V}$ are arranged according to size in order to generate the ($J \times K$) matrix $\mathcal{W}$. The column elements of $\mathcal{W}$ satisfy the relations $\mathcal{W}_{1,k} \le \mathcal{W}_{2,k} \le \ldots \le \mathcal{W}_{J,k}$ for $k = 1,\ldots,K$.

The *a-priori* distribution function of the k^th *view* can also be generated using the arrangement statistics: $C_{j,k} = F_k(\mathcal{W}_{J,k}) = j/(J+1)$.

In a **third step**, the *a-posteriori* marginal distributions $\tilde{F}_{\cdot,k}$ of the K *views* are determined. In the COP model, this takes place by calculating weighted means. The marginal distributions are as follows:

$$(1) \quad \tilde{F}_{j,k} = c_k \hat{F}_k\left(\mathcal{W}_{j,k}\right) + \left(1 - c_k\right)\frac{j}{J+1}$$

with $c_k \in [0, 1]$ denoting the confidence level of the k^th *view* and $\hat{F}_k$ the corresponding distribution function. As already mentioned, the user must express his *views* as distribution functions. As with the BL model, this can be the normal distribution function. For equally distributed expectations within a predefined interval [a, b], the distribution function for the k^th *view* can also be specified as

$$(2) \quad \hat{F}_k(\upsilon) = \begin{cases} 0 & \upsilon \le a_k \\ \dfrac{\upsilon - a_k}{b_k - a_k} & \upsilon \in \left[a_k, b_k\right] \\ 1 & \upsilon \ge b_k \end{cases}$$

Equally feasible are multi-peak density functions or truncated distribution functions. In broad principle, the user can state any distribution function for a view.

In a **fourth step** the *a-posteriori* marginal distributions thus obtained must be summarized into the joint *a-posteriori* distribution of the *views*, $\tilde{\mathcal{V}}$, i.e. a quantile function. The takes place by means of a linear interpolation of the empirical copula elements $C_{j,k}$ using the value pairs $(\mathcal{W}_{\cdot,k}, \tilde{F}_{\cdot,k})$.

In the **fifth and final step** the *views'* quantile function needs to be transformed back into the coordinates system of the market distribution. A quick reminder: The dimension of matrix $\mathcal{V}$ is ($J \times K$) and that of the panel $\mathcal{M}$ of the market distribution ($J \times N$). To achieve equal rankings, the final ($N - K$) columns of the matrix $\mathcal{V}$ are joined to the matrix $\tilde{\mathcal{V}}$ from the right ($\bar{\mathcal{V}} = (\tilde{\mathcal{V}}|\mathcal{V}_{\cdot, N-K:N})$) and multiplied from the right with the inverted *Pick* matrix $\bar{P}$: $\bar{\mathcal{M}} = \bar{\mathcal{V}}\,(\bar{P}')^{-1}$. This leads to J realisations, i.e. Monte Carlo-based simulation values, of the N random variables of the market distribution. These form the basis for further portfolio optimization.

Application

In a comparative backtest, the Copula Opinion Pooling method is used for optimizing a mixed portfolio of stocks and fixed income. In the equities area, the portfolio consists of generic futures on the S&P 500, DAX, FTSE 100 and Nikkei 225 and in the fixed income area of ten year sovereign bonds from the US, Germany, the UK and Japan.[1] The backtest uses the monthly readings from October 1998 to June 2013, i.e. a total of 177 observation vectors of the eight synthetic future prices. The entire period was divided up into an in-sample period and a simulation period. The in-sample period encompasses the first 120 discrete monthly returns – i.e. it begins in November 1998 and ends in October 2008. The 56 month-long simulation period extends from November 2008 to June 2013.

In a first step the descriptive statistics were determined. An Anderson-Darling normality test was also carried out (Stephens, 1986). Table 1 shows the results.[2]

The statistics are typical for financial market data: the equity market returns are characterized by a negative skew, with a larger kurtosis than is usual for a normal distribution. The only equity market where the null hypothesis of normally distributed returns can be said to apply is the Japanese equity market. The bond market returns also exhibit a negative skew but, with the exception of Japan, the urtosis is lower in relation to the normal distribution. The null hypothesis of normally distributed returns can be said to apply to the US and the UK bond markets.

Table 1: Statistics for the monthly returns in the in-sample period

		Mean	Standard deviation	Minimum	Maximum	Skewness	Kurtosis	AD test (p-value)
Equities	US	-0.30	4.53	-19.54	8.77	-0.95	4.85	0.00
	Germany	-0.24	6.90	-29.41	16.67	-1.12	5.90	0.00
	UK	-0.32	4.45	-15.21	7.66	-1.15	4.26	0.00
	Japan	-0.29	5.90	-22.59	11.15	-0.61	3.63	0.18
Bonds	US	0.24	1.68	-4.80	4.24	-0.23	2.95	0.76
	Germany	0.15	1.35	-3.09	2.80	-0.31	2.30	0.00
	UK	0.02	1.41	-3.63	3.10	-0.30	2.56	0.23
	Japan	0.17	1.15	-4.91	2.49	-1.31	6.43	0.00

Source: Invesco. Based on monthly data for November 1998 to October 2008.

Table 2: Marginal distribution parameters of the Student's t-Copula (estimates)

		Mean	Standard deviation	Skewness	Degrees of freedom
Equities	US	4.12	5.68	-2.12	10.82
	Germany	5.13	6.60	-1.54	4.40
	UK	4.45	5.47	-3.65	6.45
	Japan	-0.29	5.90		
Bonds	US	0.24	1.68		
	Germany	1.72	2.07	-2.81	2604.56
	UK	0.02	1.41		
	Japan	0.94	0.97	-1.22	3.70

Source: Invesco. Based on monthly data for November 1998 to October 2008.

Table 3: Dependencies between the returns

		Equities				Bonds			
		US	Germany	UK	Japan	US	Germany	UK	Japan
Equities	US	1.000	0.807	0.809	0.591	-0.142	-0.244	-0.172	-0.222
	Germany	0.361	1.000	0.786	0.534	-0.208	-0.355	-0.237	-0.221
	UK	0.363	0.335	1.000	0.571	-0.141	-0.224	-0.169	-0.199
	Japan	0.170	0.140	0.159	1.000	-0.118	-0.169	-0.171	-0.285
Bonds	US	0.008	0.006	0.008	0.009	1.000	0.364	0.305	0.152
	Germany	0.005	0.002	0.005	0.007	0.076	1.000	0.838	0.270
	UK	0.007	0.005	0.007	0.007	0.061	0.406	1.000	0.143
	Japan	0.005	0.005	0.006	0.004	0.033	0.053	0.032	1.000

Correlations light green, marginal dependence coefficients highlighted in dark green.
Source: Invesco. Based on monthly data for November 1998 to October 2008.

Due to these results, a skew Student's t-distribution was used for the marginal distributions of the US, German and UK equity market as well as for the bond markets of Germany and Japan. The returns of the remaining markets were assumed to be normally distributed. As joint density function, we selected the Student's t-Copula on account of possible marginal dependencies between the individual markets and an excess kurtosis.

The estimators of the unknown parameters of this distribution model were calculated using the *Inference from Margins* method. In a first step, the parameters of the marginal distributions were determined using the maximum likelihood method. Table 2 shows the results.

The parameters of the Student's t-Copula were estimated by numerically maximizing the pseudo likelihood function based on the percentiles of the marginal distributions. Kendall rank correlation coefficients were used as the starting values and the degrees of freedom were set at 5. The estimator for the degrees of freedom was determined at 7.69. Table 3 shows the dependencies between the returns of the eight financial instruments in the upper triangle matrix and its marginal dependence coefficients in the lower triangle matrix.

The equity market returns are strongly correlated with each other, with the Nikkei 225 returns still being the most independent. Marginal dependencies can also clearly be seen between the equity returns but here, too, Japan is an exception. All equity market returns are negatively correlated with all bond market returns. However, the dependencies between the individual bond markets are lower than between the individual equity markets, and the same also applies to the marginal dependencies. The only bond markets to exhibit a close correlation are the German and the UK bond markets, and the marginal dependence coefficient is also the highest for this pair.

A goodness of fit test for the Copula model revealed a t-value of 0.01386 at a p-value of 0.716, so that the zero hypothesis of the specified Copula model still applies. The model for the *a-priori* market distribution has thus been defined. Based on the estimates for the joint distribution, 10,000 random return vectors were generated for the eight financial instruments. These data were used for the entire simulation period.

Using a recursively estimated vector error correction model of the logs of the future prices, the *views* for the simulation period were specified by a lag in the endogenous variables. Use was made of the one-step forecasts for a confidence level of 5% at an assumed cointegration rank of 2 if they differed significantly from zero.

We weighted the *views* in two ways: (1) constantly at a confidence level of 5%, based on the assumption of equally distributed expectations between the confidence limits, and (2) adaptively at the confidence levels of the individual forecasts, i.e. the most recent forecast.[3]

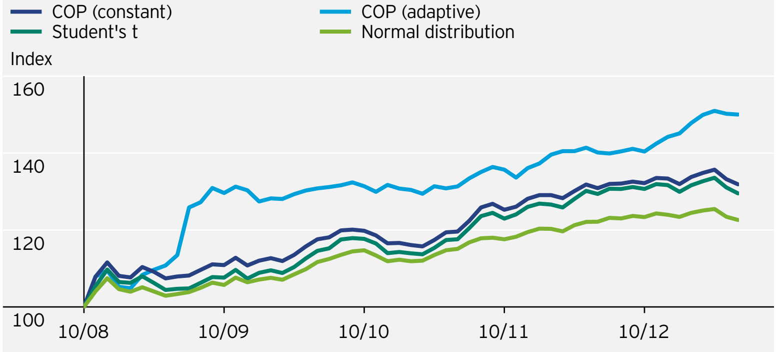
Table 4: Portfolio weights in %

		COP (constant)	COP (adaptive)	Buy-and-hold (Student's t)	Buy-and-hold (normal distribution)
Equities	US	0.0	6.5	0.0	0.0
	Germany	2.4	5.3	1.8	4.0
	UK	0.0	5.0	0.0	0.0
	Japan	0.8	8.2	0.0	0.0
Bonds	US	57.6	33.1	57.8	47.4
	Germany	23.1	23.4	21.4	0.0
	UK	0.0	2.1	0.0	0.0
	Japan	16.1	16.4	19.0	48.6
Total		**100**	**100**	**100**	**100**

Average weights in the simulation period (November 2008 to June 2013) for the COP methods; weights in November 2008 for the buy and hold strategies.
Source: Invesco.

The *a-posteriori* distributions generated from the *a-priori* market distribution and the *views* form the basis for the portfolio optimizations. The optimization problem was formulated as an optimal CVaR allocation at the confidence level of 1%, with non-negativity conditions for the weights and a budget restriction. To estimate the extent to which this performance depends on the market distribution model as well as the *views* and the weightings of the *views* in the *a-posteriori* distribution, the optimal allocations for a buy and hold strategy were also determined using the selected *a-priori* market distribution (Student's t-Copula) and assuming joint normally distributed returns. Table 4 shows the weights of the eight markets.

Figure 1: Portfolio development



Source: Invesco. Simulation results for November 2008 to June 2013.

Table 5: Performance statistics

	COP (constant)	COP (adaptive)	Buy-and-hold (Student's t)	Buy-and-hold (normal distribution)
Return (annualized, %)	6.11	9.08	5.71	4.47
Risk (annualized, standard deviation)	5.66	7.26	5.25	3.83
Sharpe ratio	1.08	1.25	1.09	1.16
CVaR (modified, 99%)	3.31	1.86	3.41	3.43
Maximum drawdown	3.75	4.24	4.80	4.24

Source: Invesco. Simulation results for November 2008 to June 2013.

The optimal CVaR solution assuming joint normally distributed returns would exhibit a high concentration in only three assets: in German equities, US bonds and Japanese bonds. The two bond markets would have accounted for 96% of the investment volume. All the other asset concepts would have resulted in greater diversification, with the adaptive COP method at the forefront. This is also the only method that would have factored in the simulation period for all eight markets. Most importantly, the average equity allocation of 25% would have been the highest of all the methods, and the average bond allocation of 75% the lowest.

Figure 1 depicts the portfolio development of the four investment concepts, table 5 the portfolio statistics.

A buy and hold strategy based on the normal distribution assumption would have generated the lowest end value. The end value of the buy and hold strategy based on the *a-priori* market distribution would have been higher. This relative outperformance is thanks largely to the better performance in 2010 and from the second half of 2011. The COP model with constantly weighted *views* would have led to a virtually identical performance of 5%. The reason is that for this assumption, the influence of the *views* on the *a-posteriori* distribution is quite low. Nonetheless the portfolio drawdown would have turned out lower at the beginning of the simulation period.

However, the adaptive weighting of the *a-priori* distribution and the *view* distributions would have resulted in the highest end value – due above all to the performance contributions in 2009 but also to the lower losses at year-end 2011 and at the end of the simulation period.

The annualized returns in table 5 correspond to the portfolio developments in figure 1. However, the risks of the various strategies are more interesting. The standard deviation risk for the buy-and-hold strategy based on the normal distribution assumption is the lowest and for the adaptive COP strategy the highest. However, the reverse is the case for the modified CVaR at a confidence level of 99% as one-sided risk measure.

The maximum portfolio drawdown also offers some interesting insights. For the adaptive COP strategy it is almost as high as for the buy and hold strategy based on the normal distribution assumption. For the COP strategy with constantly weighted *views* it is the lowest and for the Student's t-Copula model the highest. The *views*, albeit of marginal influence, most likely avoided higher portfolio losses. On the other hand, however, the risk with the adaptive COP strategy is that individual *views* might be weighted too heavily and then not be confirmed – especially at turning points.

Summary and outlook

In this article we introduced the Copula Opinion Pooling method propagated by Meucci. As an extension to the Black Litterman model, it offers alternatives for the normal distribution and allows individual expectations to be modelled far more generally. We applied the model in a simulation study to achieve CVaR-optimal mixed portfolios.

Even if the COP model is more flexible than the BL model, it still cannot take account of all forms of return expectations. Most importantly, it is unable to model non-linear return expectations or rankings.

In our next article we therefore introduce the *Entropy Pooling* method also propagated by Meucci.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Asset Allocation

Notes:

1 Data source was Thomson Reuters Datastream. The mnemonics of the time series in the order mentioned above are: ISPCS04, GDXCS04, LSXCS04, ONACS04, CTYCS04, GGECS04, LIGCS04 and JGBCS04.

2 All the calculations were carried out using the free statistical programming environment R 3.0.1 (R Core Team, 2013) as well as the packages copula (Hofert et al., 2013; Yan, J., 2007; Kojadinovic, I. and Yan, J., 2010; Hofert, M. and Mächler, M., 2011), Hmisc (Harrell, 2013), MASS (Venables and Ripley, 2002), nortest (Gross and Ligges, 2012), parma (see Ghalanos, 2013), PerformanceAnalytics (Carl and Peterson, 2013), signal (Ligges, 2013), sn (Azzalini, 2013), timeSeries (Würtz and Chalabi, 2013), urca (Pfaff, 2008a) and vars (Pfaff, 2008b).

3 See Part 9 of this series, Risk & Reward 3rd quarter 2013.

Literature

- Azzalini, A. (2013). R package sn: The skew-normal and skew-t distributions. Università di Padova, Italia. R package version 0.4-18.
- Carl, P. and B. Peterson (2013). PerformanceAnalytics: Econometric tools for performance and risk analysis. R package version 1.1.0.
- Ghalanos, A. (2013). parma: portfolio allocation and risk management applications. R package version 1.03.
- Gross, J. and U. Ligges (2012). nortest: Tests for Normality. R package version 1.0-2.
- Harrell, F. and Dupont, C. (2013). Hmisc: Harrell Miscellaneous. R package version 3.10-1.1.
- Hofert, M., I. Kojadinovic, M. Maechler, and J. Yan (2013). copula: Multivariate Dependence with Copulas. R package version 0.999-6.
- Hofert, M. and Mächler, M. (2011). Nested archimedean copulas meet R: The nacopula package. Journal of Statistical Software 39(9), 1–20.
- Kojadinovic, I. and Yan, J. (2010). Modeling multivariate distributions with continuous margins using the copula R package. Journal of Statistical Software 34(9), 1–20.
- Ligges, U. (2013). signal: Signal processing. R package version 0.7-3.
- Meucci, A. (2005). Risk and Asset Allocation. New York: Springer.
- Meucci, A. (2006a). Beyond black-litterman in practice: A five-step recipe to input views on non-normal markets. Risk 19(9), 114–119.
- Meucci, A. (2006b). Beyond black-litterman: Views on non-normal markets. Risk 19(2), 96–102.
- Meucci, A. (2010a, October). The black-litterman approach: Original model and extensions. Working paper, Symmys, http://ssrn.com/abstract=1117574.
- Meucci, A. (2010b). The Encyclopedia of Quantitative Finance, Chapter The Black-Litterman Approach: Original Model and Extensions. John Wiley & Sons.
- Meucci, A. (2010c, December). Fully flexible views: Theory and practice. Working paper, Symmys, http://ssrn.com/abstract=1213325.
- Pfaff, B. (2008a). Analysis of Integrated and Cointegrated Time Series with R (Second ed.). New York: Springer. ISBN 0-387-27960-1.
- Pfaff, B. (2008b). Var, svar and svec models: Implementation within R package vars. Journal of Statistical Software 27(4).
- R Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Stephens, M. (1986). Goodness-of-Fit Techniques, Chapter Tests based on EDF statistics. New York: Marcel Dekker.
- Venables, W. N. and B. D. Ripley (2002). Modern Applied Statistics with S (Fourth ed.). New York: Springer. ISBN 0-387-95457-0.
- Würtz, D. and Y. Chalabi (2013). timeSeries: Rmetrics - Financial Time Series Objects. R package version 3010.97.
- Yan, J. (2007). Enjoy the joy of copulas: With a package copula. Journal of Statistical Software 21(4), 1–21.

New approaches to portfolio optimization: Part 11

In the last two parts of the series, we showed how individual return expectations can be modelled using the Black-Litterman (BL) method and its extension known as Copula Opinion Pooling (COP)[1]. Despite offering an alternative to the strict normal distribution assumption of the BL model, COP is still unable to model non-linear return expectations. The Entropy Pooling (EP)[2] method offers a solution.

Like COP, the EP method is based on a combination of an *a-priori* market distribution and individual return expectations, known as *views*. From these two elements, an *a-posteriori* distribution is derived.

The EP model, propagated by Meucci, measures the difference between two distributions (in our case the *a-priori* and the *views* distribution) using the Kullback-Leibler divergence.[3] This measure is also known as relative entropy, hence the name of the model.

The EP method can also be broken down into three stages:

1. The determination of the *a-priori* market distribution
2. The specification of the individual expectations
3. The minimization of the relative entropy and determination of the *a-posteriori* distribution

The determination of the *a-priori* market distribution
The starting point is an N-dimensional random vector of risk factors, $\mathbf{x}$. In the simplest case, it corresponds to the return vector of the portfolio assets. The prices $P_{t+\tau}$ of the portfolio assets can be determined from the realizations of $\mathbf{x}$ in the investment period τ using the deterministic function $P_{t+\tau} = P(\mathbf{x}, \mathrm{I}_t)$. I_t denotes all the information available at time t. The joint density function f^M of the random vector $\mathbf{x}$ is the *a-priori* market distribution.

In principle, this distribution can be determined non-parametrically as well as parametrically. In a non-parametric estimate, the observations (realizations) of the risk factors are used but the small sample size makes this impracticable.

With a parametric estimate this problem does not arise. Using a distribution assumption or a Copula model as a basis, a sample of the size J can be generated for the N risk factors using a Monte Carlo simulation, similar to the COP method. A ($J \times N$) matrix $\mathcal{M}$ is thus derived, its lines denoting possible realizations of the joint market distribution and its columns possible realizations of the n[th] risk factor.

What's also needed is an assumption about the probability of the individual realizations of the market distribution. The general assumption is that they are all equally probable; i.e. their probabilities are all $1/J$.

The specification of the individual expectations
In a second step, the individual expectations are modelled. For each of the $k = 1, \ldots, K$ *views* a function $g_k(\mathbf{x})$ is specified, depending on the risk factors $\mathbf{x}$. The *views* $V = \mathbf{g}(\mathbf{x})$ thus generate a K-dimensional random vector, with the individual functions g_k grouped together in the vector function $\mathbf{g}(.)$. We term the joint distribution of the *views* f^V.

Unlike in the classic BL model, the individual expectations $g_k(\mathbf{x})$ can be expressed as non-linear functions. Thus, alongside classic estimates of the expected returns, nominal or ordinal scaled expectations on location parameters can also be modelled, as well as rankings of the expected returns, volatilities and dependencies between the risk factors.

The *views* are then merged with the market distribution $\mathcal{M} = (\mathbf{x}_1', \ldots, \mathbf{x}_J') = X$; the j-k[th] element of $\mathcal{V}$ is $\mathcal{V}_{j,k} = g_k(X_{j,1}, \ldots, X_{j,N})$. The probabilities $\mathbf{p}^V$ of the market distribution are therefore modified by the individual expectations.

But how must the new ($J \times 1$) probability vector be formulated so that the *views* are fulfilled? The *views* can be formulated as a system of K inequations: $\underline{a} \leq A\mathbf{p}^V \leq \bar{a}$, with $\underline{a}$ and $\bar{a}$ denoting ($K \times 1$)-dimensional vectors of the lower and upper limits. As (if $J > K$)[4] we are dealing with an underdetermined equation system, several possibilities exist to modify the probability vector in such a way that it does justice to individual expectations. With the EP approach, the objective is to find the one which intervenes the least in the *a-posteriori* distribution.

The minimization of the relative entropy and determination of the *a-posteriori* distribution
To achieve this, the relative entropy is minimized. As already mentioned, the distance between two probability functions is measured using the Kullback-Leibler divergence. For density functions it is defined as

$$(1) \quad \mathbb{D}\left(f^V, f^M\right)_8 = \int_{-\infty}^{\infty} f^V \left(x\left[\ln f^V(\mathrm{x}) - \ln f^M(\mathrm{x})\right] dx\right),$$

for discrete distributions as

$$(2) \quad \mathbb{D}\left(\mathbf{p}^V, \mathbf{p}^M\right)_d = \sum_{j=1}^{J} p_j^V \left[\ln p_j^V - \ln p_j^M\right].$$

Equations (1) and (2) reveal two things: if the density and probability functions are equal, the distance is zero. The relative entropy is an asymmetric variable, i.e. $\mathbb{D}\left(f^V, f^M\right)_S \neq \mathbb{D}\left(f^M, f^V\right)_S$.

With equation (2) the *a-posteriori* distribution is derived by solving the following optimization problem:

$$(3) \quad \tilde{\mathbf{p}} \equiv \underset{\underline{a} \leq A\mathbf{p}^V \leq \bar{a}}{\arg\min} \left\{\mathbb{D}\left(\mathbf{p}^V, \mathbf{p}^M\right)_d\right\}.$$

Table 1: Statistics for returns in the first sub-period

		Mean	Standard deviation	Minimum	Maximum	Skewness	Kurtosis	AD test (p-value)
Equities	US	-0.08	2.44	-17.48	10.31	-0.70	8.89	0.00
	Germany	-0.08	3.51	-15.59	17.41	-0.54	6.45	0.00
	UK	-0.09	2.60	-12.80	14.52	-0.16	7.52	0.00
	Japan	-0.02	3.05	-22.98	10.66	-0.97	9.26	0.00
Bonds	US	0.06	0.84	-3.02	2.89	-0.16	3.76	0.15
	Germany	0.03	0.71	-2.96	2.30	-0.37	3.94	0.01
	UK	0.00	0.73	-3.23	2.52	-0.36	4.52	0.00
	Japan	0.04	0.58	-2.73	2.37	-0.86	6.71	0.00

Source: Invesco. Based on weekly data from 11 November 1998 to 22 October 2008.

The elements of the ($J \times 1$) solution vector $\mathbf{p}$ are the *a-posteriori* probabilities being sought, based on the assumption of complete certainty about the *views;* i.e. that the confidence is 1. However, this is generally not the case. Similar to the COP method, the *a-posteriori* probabilities (as determined in accordance with equation (3)) and the probabilities of the market model are weighted with the confidence of the *views* (or pooled, which explains the second half of the term "entropy pooling"):

(4) $\bar{\mathbf{p}} = (1-c)\mathbf{p}^M + c\tilde{\mathbf{p}}$,

with $c \in [0.1]$ denoting the confidence of the *views*.[5]

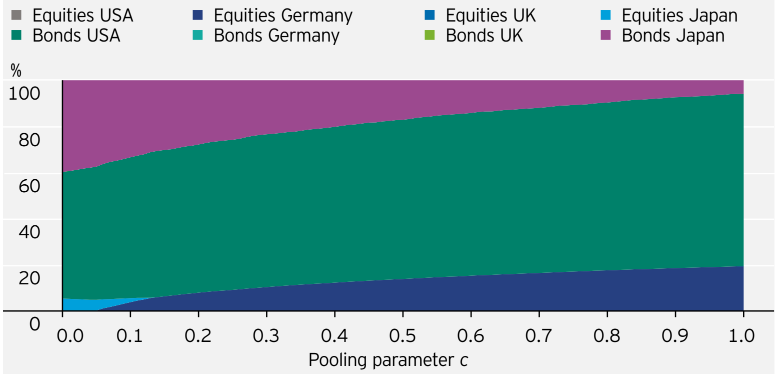
Comparative simulation

We now use the entropy pooling method to optimize a mixed portfolio of stocks and bonds. On the equities side, the portfolio is made up of generic futures on the S&P 500, DAX, FTSE 100 and Topix. The bond market is modelled through futures on ten-year sovereign bonds from the US, Germany, the UK and Japan.[6]

Unlike the previous article, here we base our calculations on weekly instead of monthly data, using Wednesday's closing prices. This higher data frequency is more suitable for formulating individual expectations in the new model.

For generating the individual expectations, an ARMA(1.0)-GARCH(1.1) model is now used instead of a vector error correction model. Thus, both the point estimates of the return forecasts as well as their conditional volatilities determine the *a-posteriori* distribution.

Figure 1: CVaR-optimal allocation depending on the pooling parameter



Source: Invesco. Based on weekly data from 11 November 1998 to 22 October 2008.

The analysis period begins on 4 November 1998 and ends on 25 September 2013 so that there are 778 price and 777 return vectors. Similar to our previous article in this series, we began by forming an in-sample period (until 22 October 2008; 520 return vectors) immediately followed by a simulation period.

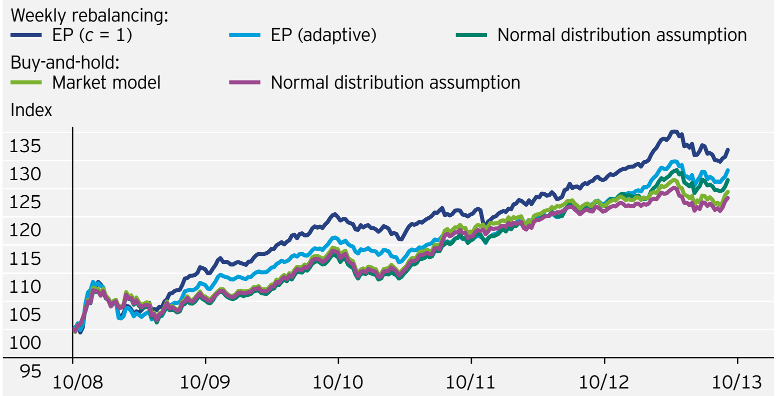
To start with we calculated descriptive statistics for the first sub-period and carried out the Anderson-Darling Normality Test.[7] The results can be seen in table 1.[8]

Table 2: Average portfolio allocation in %

		Weekly rebalancing			Buy-and-hold	
		EP (c = 1)	EP (adaptive)	Normal distribution assumption	Market model	Normal distribution assumption
Equities	US	3.01	1.66	0.28	0.00	0.00
	Germany	4.23	2.54	3.34	0.00	0.16
	UK	0.47	0.09	0.07	0.00	0.00
	Japan	6.06	6.78	3.42	5.41	4.65
Bonds	US	16.31	22.44	32.92	54.85	49.37
	Germany	22.09	15.46	19.12	0.00	0.00
	UK	2.33	0.08	0.00	0.00	0.00
	Japan	45.51	50.96	40.85	39.74	45.82

Source: Invesco. Calculations are based on simulation results from 22 October 2008 to 25 September 2013.

Figure 2: Simulated performance



Source: Invesco. Simulation results for 22 October 2008 to 25 September 2013.

Table 3: Performance statistics

	Weekly rebalancing			Buy-and-hold	
	EP (c = 1)	EP (adaptive)	Normal distribution assumption	Market model	Normal distribution assumption
Return (annualized, %)	5.77	5.17	4.88	4.53	4.34
Risk (annualized, standard deviation)	4.16	3.87	3.70	3.99	3.74
Sharpe ratio	1.39	1.33	1.32	1.14	1.16
CVaR (modified, 95%)	1.35	1.25	1.06	1.11	1.04
Maximum drawdown	5.61	6.15	5.20	5.25	4.97
Concentration	0.29	0.34	0.31	0.46	0.46

Source: Invesco. Calculations are based on simulation results from 22 October 2008 to 25 September 2013.
The performance results shown are hypothetical (not real) and were derived by back-testing using a simulated portfolio which does not factor in all the economic and market conditions that can impact results. There can be no assurance that the simulated results can be achieved in the future.

The statistics reflect the typical features of financial market returns: the equity returns display a negative skew and the kurtosis is larger than with a normal distribution. However, the null hypothesis of normally distributed returns cannot be rejected for the US bond market. For this reason, a negatively skewed Student's t-distribution was assumed for the marginal distributions of German, UK and Japanese ten year yields and a normal distribution for US yields. As in our previous article, we define the joint density function using a Student's t Copula. For the market model, the results do not differ qualitatively from those of the previous article.

Using the t-Copula thus specified, we generated 100,000 possible realizations of the market distribution M and determined the *views* distribution at maximum confidence in one's expectations (c=1). The sample moments were entered into an optimum CVaR allocation at a confidence of 95% (subject to a budget and non-negativity restriction).

Since the pooling parameter *c* has to be set, we first examined the extent to which the allocation depends on its value. The averaged *a-posteriori* probabilities were varied for *c* = 0.01,...,1.00 in hundredth steps and used for the portfolio optimization. Figure 1 depicts the allocations of the CVaR-optimal portfolios.

All in all the weights can be seen permanently changing. As the certainty of the *views* increases, German stocks and US bonds are given greater weightings at the expense of Japanese stocks and bonds.

In a next step we carried out simulations for the second sub-period. Based on the individual ten-year expectations, the *a-posteriori* probabilities were computed – assuming complete confidence (c=1) as well as factoring in adaptive expectations for c.[9] In addition, a simulation was carried out using the normal distribution assumption and buy-and-hold strategies based on the data available on 22 October 2008 – on the one hand for the market model M and on the other hand for the normal distribution assumption.

Table 2 depicts the average allocations in the simulation period. The entropy concept results in a higher equity weighting. A striking point with all five approaches is the high weighting of Japanese bonds.

Figure 2 depicts the performance, table 3 the statistics of the five investment concepts. The EP strategies lead to the highest end value due to a marked outperformance towards the middle of 2009 and only a small loss at year-end 2011. The two buy-and-hold strategies do not differ substantially. The reason for the slightly diverging performance is the different weights given to US and Japanese bonds.

The performance statistics in table 3 show the return and risk-adjusted yield of the EP strategy (at full confidence) to be the highest. However, for both EP concepts, the CVaRs at 95% confidence and the drawdowns are also the highest.

Summary and outlook

This article on the EP model concludes our exposition of Bayesian optimization. Our next article will deal with the theme of expected utility maximization as a supplement to modern portfolio theory.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Asset Allocation

Notes:
1 See Meucci (2006a), Meucci (2006b)
2 See Meucci (2010)
3 See Kullback and Leibler (1951), Kullback (1959)
4 For the parametric version of the EP method, this condition can be readily fulfilled since the user can determine the size of the simulation.
5 A point to note is that solving the optimization problem using equation (3) is impractical, so the dual form of the problem is maximized instead. In this case, the number of the variables to be optimized reduces from J to K i.e. to the number of *views*. The dual formulation of the problem allows the optimization task to be solved simply and quickly.
6 Data source is Thomson Reuters DataStream. The mnemonics of the time series in the order mentioned above are: ISPCS04, GDXCS04, LSXCS04, JSXCS04, CTYCS04, GGECS04, LIGCS04 and JGBCS04.
7 See Stephens (1986)

8 All the calculations were carried out using the free statistical programming environment R 3.0.1 (R Core Team, 2013) as well as the packages copula (Hofert et al., 2013; Yan, J., 2007; Kojadinovic, I. and Yan, J., 2010; Hofert, M. and Mächler, M., 2011), Hmisc (Harrell, 2013), MASS (Venables and Ripley, 2002), nortest (Gross and Ligges, 2012), parma (see Ghalanos, 2013), PerformanceAnalytics (Carl and Peterson, 2013), sn (Azzalini, 2013) and timeSeries (Würtz and Chalabi, 2013).

9 As in our previous article in this series, we transformed the portfolio returns into the interval (0,1) using the standard normal distribution function.

Literature

- Azzalini, A. (2013). R package sn: The skew-normal and skew-t distributions. Università di Padova, Italia. R package version 0.4-18.
- Carl, P. and B. Peterson (2013). PerformanceAnalytics: Econometric tools for performance and risk analysis. R package version 1.1.0.
- Ghalanos, A. (2013). parma: portfolio allocation and risk management applications. R package version 1.03.
- Gross, J. and U. Ligges (2012). nortest: Tests for Normality. R package version 1.0-2.
- Harrell, F.and Dupont, C. (2013). Hmisc: Harrell Miscellaneous. R package version 3.10-1.1.
- Hofert, M., I. Kojadinovic, M. Maechler, and J. Yan (2013). copula: Multivariate Dependence with Copulas. R package version 0.999-6.
- Hofert, M. and Mächler, M. (2011). Nested archimedean copulas meet R: The nacopula package. Journal of Statistical Software 39(9), 1–20.
- Kojadinovic, I. and Yan, J. (2010). Modeling multivariate distributions with continuous margins using the copula R package. Journal of Statistical Software 34(9), 1–20.
- Kullback, S. (1959). Information Theory and Statistics. New York: Jon Wiley & Sons.
- Kullback, S. and R. Leibler (1951). On information and sufficiency. Annals of Mathematical Statistics 22(1), 79–86.
- Meucci, A. (2006a). Beyond black-litterman in practice: A five-step recipe to input views on non-normal markets. Risk 19(9), 114–119.
- Meucci, A. (2006b). Beyond black-litterman: Views on non-normal markets. Risk 19(2), 96–102.
- Meucci, A. (2010, December). Fully flexible views: Theory and practice. Working paper, Symmys, http://ssrn.com/abstract=1213325.
- R Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Stephens, M. (1986). Goodness-of-Fit Techniques, Chapter Tests based on EDF statistics. New York: Marcel Dekker.
- Venables, W. N. and B. D. Ripley (2002). Modern Applied Statistics with S (Fourth ed.). New York: Springer. ISBN 0-387-95457-0.
- Würtz, D. and Y. Chalabi (2013). timeSeries: Rmetrics - Financial Time Series Objects. R package version 3010.97.
- Würtz, D., Y. Chalabi, M. Miklovic, C. Boudt, and P. Chausse (2013). fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling. R package version 3010.82.
- Yan, J. (2007). Enjoy the joy of copulas: With a package copula. Journal of Statistical Software 21(4), 1–21.

New approaches to portfolio optimization: Part 12

In this final part of the series, we show how optimum portfolio allocation can be achieved by means of expected utility maximization. In this case the portfolio weights are treated as random variables. On this basis, we determine the allocation leading to the investor's expected utility.

The starting point of portfolio optimization is the utility function of the investor:

$$(1) \quad U = \lambda\omega\mu - (1-\lambda)\omega'\Sigma\omega$$

with (for N portfolio assets) ω denoting the $(N \times 1)$ vector of the portfolio weights, μ the $(N \times 1)$ vector of the return expectations, Σ the $(N \times N)$ variance-covariance matrix and $\lambda \in (0, 1)$ the risk aversion parameter. The utility maximizing portfolio thus depends on μ, Σ and λ.

In an earlier article in this series we showed that even smaller changes in these parameters can have major consequences for optimum allocation. At that time we proposed resorting to robust estimators or robust optimization methods.[1] In this article we present an alternative.

Unlike before, we now consider the portfolio weights as random variables with a probability distribution. As optimum allocation depends on the returns and risks of the portfolio assets and both are considered as random variables, the weight vector is also a random variable.

Rossi et al. (2002) and Marschinski et al. (2007) take up at this point. They interpret the utility function as the logarithm of a density function whose parameters include the weight vector ω. The optimum allocation is the expected value of ω. That explains why reference is also made to a "probabilistic" interpretation of the utility function.

In the following we make use of the general function $u = u(\omega, U, \theta)$. U denotes the utility function and θ all its arguments (e.g. expected returns, risk factors and/or dispersion measures). The expected utility is proportional to the logarithm of the probability expression:

$$(2) \quad \omega \sim P(\omega|U,\theta) = Z^{-1}(\nu,U,\theta)\exp(\nu u(\omega,U,\theta))$$

In equation (2) Z is a constant so that the area below the density function is normalized to one. This constant is defined as:

$$(3) \quad Z(\nu,U,\theta) = \int_{\mathfrak{D}(\omega)} [d\omega] \exp(\nu u(\omega,U,\theta))$$

Equation (2) also contains the convergence constant ν which is defined as $\nu = pT^{\gamma}$, with T representing the sample range. With $p = 1$ and $\gamma = ½$, an asymptotic convergence to the utility maximizing allocation is derived with the rate , $\sqrt{T}$. For $\nu \to \infty$ the distribution converges to the utility maximizing allocation, for $\nu \to 0$ to a portfolio with equal-weighted assets.

With

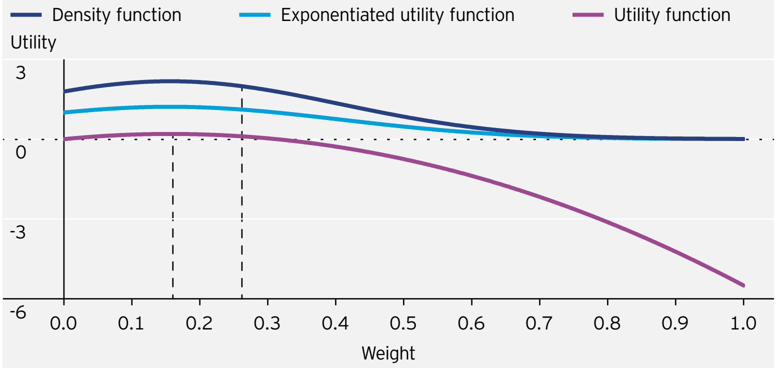
$$(4) \quad \bar{\omega}(U,\theta) = Z^{-1}(\nu,U,\theta) \int_{\mathfrak{D}(\omega)} [d\omega]\, \omega \exp(\nu u(\omega,U,\theta))$$

the portfolio solution ω can then be determined as an expected value.

The precise extent to which the probabilistic interpretation of the utility function differs from the classic utility maximizing allocation is illustrated in the following example. We assume a portfolio with two assets, a risky asset with excess return of 5 percentage points and a standard deviation risk of 4% and a risk-free investment. The risk aversion parameter λ is assumed to be 0.5, the constant ν is assumed to be 1.

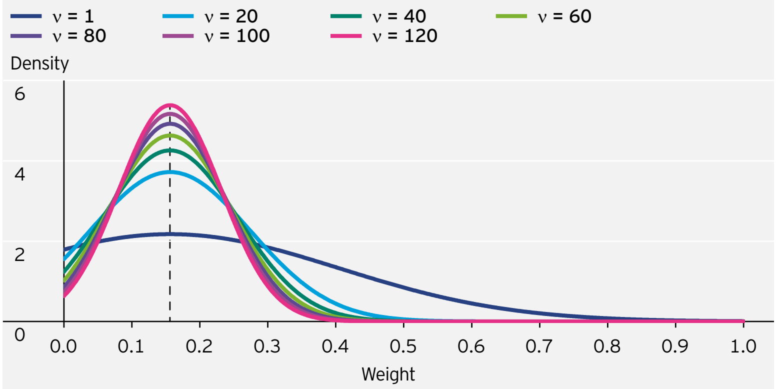
Figure 1 depicts the utility function as a purple line; the utility maximizing share of the risky asset amounts to around 15.6%. The exponentiated utility function is shown as a light blue line, the density function as a dark blue line. The area beneath the exponentiated utility function is unequal to one; it only becomes a valid density function through the normalization constant. Its expected value amounts

Figure 1: Utility maximizing and probabilistic optimum allocation



Source: Invesco, own calculations. For illustrative purposes only.

Figure 2: Asymptotic convergence of the probabilistic utility function



Source: Invesco, own calculations. For illustrative purposes only.

to around 26.7% and is thus greater than that of the utility maximizing allocation. The values of the utility function are 0.2 and 0.1.

In this example, figure 2 illustrates the convergence of the probabilistic utility function to the utility maximizing allocation. The values assumed by constant ν are 1, 20, 40, 60, 80, 100 and 120 ($p = \gamma = 1$). With increasing sample size, the distribution collapses over the utility maximizing allocation.

Determining the distribution function
According to equation (4), the optimum portfolio is an N-dimensional integral multiplied by a constant. Due to the general impracticability of an analytical solution, the distribution function is determined according to the Markov-Chain Monte Carlo model, known as the MCMC model. This involves searching the state space of the distribution parameter and evaluating it along a Markow chain.[2] The only input size required by the MCMC model is a density function as in equation (2). Detailed explanations can be found in Gilks et al. (1995) and Brooks et al. (2011).

The drawback with this standard form of the MCMC model is the generally high auto correlation which greatly slows down the process of searching the entire state space of the distribution parameter due to the many steps required. For this reason, Duane et al. (1987) introduced a version of the MCMC model which not only avoids this weakness but also leads in most cases to a higher acceptance rate for the individual steps. This method is also known as the Hybrid-Monte-Carlo algorithm, or quite simply the HMC algorithm. Compared with the classic MCMC model, (1) the search of the state space is made in larger steps, (2) the autocorrelation is lower and (3) the acceptance rate is higher. The distribution function can therefore be determined more quickly. But the advantages of the HMC algorithm also come at a price: Besides the density function itself, the gradient of its distribution parameter is now also required. Neal (2011) offers a detailed description of the HMC algorithm which is used in the following simulation study.

Simulation
With the concept of the probabilistic utility function we now optimize a mixed portfolio of equities and bonds similar to previous articles in this series. The portfolio comprises generic futures on S&P 500, DAX, FTSE 100 and Nikkei and futures on ten-year sovereign bonds from the US, Germany, the UK and Japan.[3] We used month-end values from October 1998 to September 2013.

Similar to Marschinski et al. (2007), the structure of the analysis is geared to Michaud (1989, 1998):

1. On the basis of the discrete percentage returns, the expected returns and the variance-covariance matrix are calculated (μ, Σ). These estimates are considered as true values of the data pool, assuming that the yields are jointly normally distributed.

2. The utility maximizing allocation is determined on the basis of these true parameters.

3. With the distribution parameters, K random data sets are produced with sample range L.

4. For each of the K data sets (1) the utility maximizing allocation and (2) the expected value of the probabilistic utility function are determined.

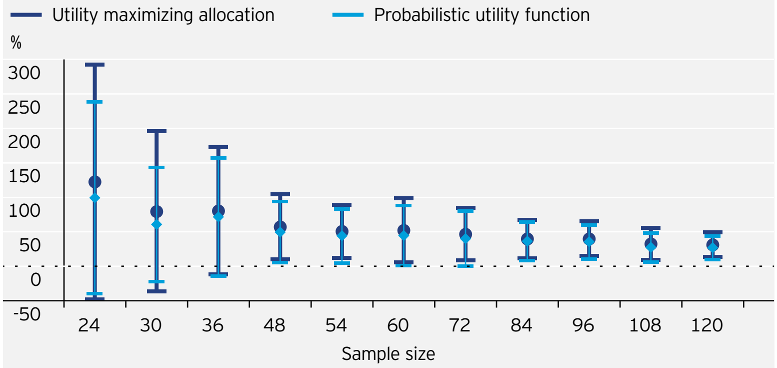
5. Finally, for each data set the deviations ("distances") of the two solutions from the true allocation (in accordance with point 2) are calculated.

For samples with 24, 30, 36, 48, 54, 60, 72, 84, 96, 108 and 120 observations, 100 data sets were generated ($K = 100$). The Markow chain was 250 in length, with the first 150 data sets for the distribution parameters not included in the determination of the expected values ("burn-in periods"). The constant ν was in each case equated with the sample size. The percentage deviations of the utility from the true utility were used as distances.[4]

Figure 3 shows the empirical distribution of these distances, described by mean and standard deviation.

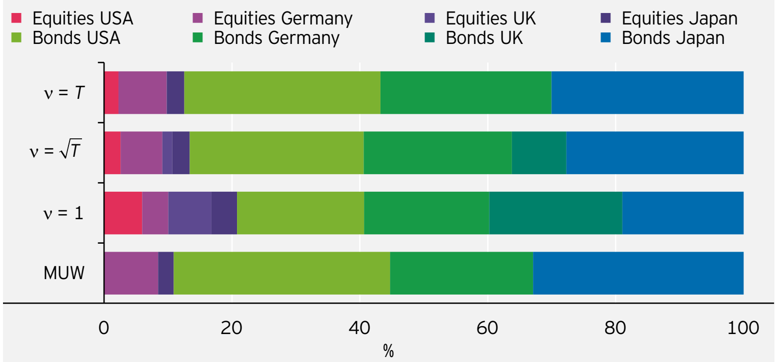
In the case of the utility maximizing allocation, smaller samples show strikingly large deviations from the true utility. This is attributable to the sensitivity of this concept to outliers. With the probabilistic utility function, the diversification is lower.

Figure 3: Deviation from true utility



Source: Invesco. Based on month-end data from October 1998 to September 2013.

Figure 4: Portfolio weights



Source: Invesco. Based on month-end data from October 1998 to September 2013.

The greater the sample size, the more closely the moments of the two solutions diverge even if the deviations from the true utility are somewhat smaller in the case of the probabilistic utility optimization.

Finally, based on the returns for the entire observation period we compare the allocations for different values of ν with the utility maximizing allocation. The parameter ν assumed the values $\nu_1 = 1$, $\nu_2 = \sqrt{T}$ and $\nu_3 = T$. The risk aversion parameter λ was set at 0.9. Figure 4 illustrates the allocations.

The utility maximizing allocation (shown in figure 4 as "MUW") turns out to be highly-concentrated; approximately two thirds of it comprise Japanese and US government bonds. Apart from the DAX, there were evidently no significant exposures to equities. The probabilistic optimization reveals a lower concentration for all values of ν. Even with $\nu = T$, i.e. a sample range of 179 observations, the allocation would have been more balanced.

Summary

For the probabilistic interpretation of utility functions, the portfolio weights themselves are considered as random variables; in this case, the optimum allocation is their expected value. In broad principle, this approach is suitable for every utility function, in other words also for functions with one-sided risk measures or risk neutrality. However, for simplification purposes or better comparison we have used the function familiar from the Markowitz approach.

The simulation study has shown that the probabilistic interpretation of utility functions can lead to far lower portfolio concentrations than the traditional utility maximizing allocation. It therefore serves as an alternative to robust estimators or optimization processes.

Dr. Bernhard Pfaff, Portfolio Manager,
Invesco Global Asset Allocation

Notes:

1 Risk & Reward, 3rd and 4th Quarter 2011
2 The validity of each step can be determined, for example, using the Metropolis-Hastings algorithm.
3 Thomson Reuters DataStream was used as the data source. The mnemonics of the time series in the order mentioned above are: ISPCS04, GDXCS04, LSXCS04, ONACS04, CTYCS04, GGECS04, LIGCS04 and JGBCS04.
4 All calculations were carried out using the free statistical programming environment R 3.0.2 (see R Core Team, 2013) and the GRIMS package (see Neal, 2011).

Literature

- Brooks, S., A. Gelman, G. Jones and X.-L. Meng (ed.) (2011). Handbook of Markow Chain Monte Carlo. Boca Raton, FL: Chapman & Hall / CRC.
- Duane, S., A. Kennedy, B. Pendleton, and D. Roweth (1987). Hybrid monte carlo. Physical Letters B195, 216-222.
- Gilks, W., S. Richardson and D. Spiegelhalter (1995). Markov Chain Monte Carlo in Practice. Interdisciplinary Statistics. Boca Raton, FL.: Chapman & Hall / CRC.
- Marschinski, R., P. Rossi, M. Tavoni, and F. Cocco (2007). Portfolio selection with probabilistic utility. Annals of Operations Research 151, 223-239.
- Michaud, R. (1989). The markowitz optimization enigma: Is optimized optimal. Financial Analyst Journal 45, 31-42.
- Michaud, R. (1998). Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation. New York: Oxford University Press.
- Neal, R. (2011). Handbook of Markov Chain Monte Carlo, Chapter MCMC using Hamiltonian dynamics, pp. 113-162. Handbooks of Modern Statistical Methods. Boca Raton, FL: Chapman & Hall / CRC.
- R Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rossi, P., M. Tavoni, F. Cocco and R. Marschinski (2002, November). Portfolio selection with probabilistic utility, bayesian statistics and markov chain monte carlo. eprint arXiv arXiv:cond-mat/0211480, 1-27. http://arxiv.org.

New approaches to portfolio optimization: A summary

This is the final article in the series which we started in the 2nd quarter 2011 on the 60th anniversary of Markowitz's groundbreaking work "Portfolio Selection". In a series of twelve articles we presented the classic Markowitz optimization approach, its developments and extensions and the resultant portfolio concepts:

Part	Issue	Theme
1	Q2 2011	Mean-variance optimization
2	Q3 2011	Robust estimators
3	Q4 2011	Robust optimization methods
4	Q1 2012	VaR optimum and CVaR optimum portfolios
5	Q2 2012	Drawdown optimum portfolios
6	Q3 2012	"Most diversified portfolios"
7	Q4 2012	Tail-dependent portfolio optimization
8	Q1 2013	Equal risk contribution portfolios
9	Q3 2013	The Black-Litterman approach
10	Q4 2013	Copula opinion pooling
11	Q1 2013	Entropy pooling
12	Q2 2013	Probabilistic utility functions

The classic Markowitz optimization approach, when put into practice, frequently results in a strong concentration on a few financial instruments. Added to this, even the slightest parameter change has major consequences for optimum portfolio allocation.

We therefore presented methods which counter these weaknesses: robust estimators mitigate the influence of outliers and, by extension, lower portfolio concentration while robust optimization methods ensure greater stability in the event of parameter changes.

Consideration was then given to various risk indicators. Markowitz deems a portfolio to be optimum if the relation between return expectations and return variance ("volatility") is appropriate. But not all investors share this view. Other alternatives are portfolio-VaR (or CVaR) and portfolio drawdown. We have shown the consequences that these have for portfolio optimization.

Views also diverge when it comes to the meaning of portfolio diversification. Traditionally it is measured on the basis of the return covariances between the individual positions. But here, too, there are alternatives – the "Most Diversified Portfolio", diversification measures which only capture simultaneous losses and a portfolio concept in which all positions make equal contributions to aggregate risk.

Finally, we dealt with the question of modelling individual return expectations, the classic Black-Litterman approach along with its extensions Copula Opinion Pooling and Entropy Pooling. The final article dealt with expected utility maximization as a basis for portfolio optimization.

The risk-return paradigm is still a set component of quantitative optimization methods today. But risks are modelled and defined differently due not least to the financial market crisis of the last ten years. Optimization methods were introduced which only consider losses; market risks were modelled more precisely and interdependencies captured more efficiently.

Depending on the investor's preference, one or other of these approaches can be optimum. But this is where the difficulty lies: in the final analysis, whichever approach is chosen remains a subjective decision – but one of great importance.