The Core Factor –
a fast and accurate factor reduction technique

Christoff Gössl,
Unicredit Markets and Investment Banking, London
December 2007

Summary

In credit portfolio modelling the normal copula approach and multi-factor models have become an industry-wide standard to describe the asset correlation structure. To calculate the portfolio loss distribution for these models, due to its analytical intractability, in general Monte Carlo methods, non-trivial numerical integrations, or structural simplifications have to be applied. Still, for more than 3-4 factors especially determination of risk contributions and optimisation routines become a very time-consuming endeavour. Aim of this paper is to present a new factor reduction approach using a single underlying factor to calculate the portfolio loss distribution. Focusing on risk contributions rather than asset correlations it is shown that the portfolio correlation structure could be captured in just one factor to an astonishing degree of accuracy. The paper outlines the basic idea and calibration of this factor reduction approach, and for illustrative as well as realistic, heterogeneous portfolios the new model is applied in the calculation of loss distributions and risk contributions. The capabilities of this approach are further demonstrated when used in an optimisation setting.

Keywords: Factor models; factor reduction; correlation; optimisation.

Remark: The opinions expressed here are those of the author, and do not reflect the views of Unicredit MIB or its staff. Written correspondence can be send to christoff.goessl@HVBEurope.com, or to 120 London Wall, London EC2Y 5ET, UK.

1. Introduction

In the last decade in credit portfolio modelling the normal copula approach and multi-factor models have become an industry-wide standard to describe the asset correlation structure. Allowing to construct mutual asset correlations between members of the portfolio as the sum of dependencies on common external factors, the multi-factor approach represents a very flexible and powerful modelling tool. If any drawbacks exist, one would argue, then it is the fact that for every minute source of correlation an independent factor is to be added to the model. To calculate the loss distribution for these multi-factor portfolio models, Monte Carlo methods, non-trivial numerical integrations, or structural simplifications are applied. Most practitioners, however, resort to Monte Carlo simulations, which work very well even for the sometimes very large number of factors. This should be attributable to the fact that for every instrument usually only a handful of factors have non-zero loadings, virtually reducing the integration dimensions in the Monte Carlo algorithm. However, difficulties arise when statistics have to be calculated where in a reasonable time frame Monte Carlo simulations are not accurate enough and are therefore no longer an option. For example for risk contributions far in the tail, Monte Carlo results can be fairly volatile. This is especially unwanted in optimisation algorithms, where such random fluctuations in the parameters are frequently punished with an immediate termination of the algorithm. To avoid these drawbacks, a number of workarounds have been found. In some cases loss amounts are 'blurred' by some mixing distributions to ensure more continuous risk contributions of instruments and thus reduce volatilities. Other approaches, based on a direct calculation of the loss distribution, focus on a reduction in the number of underlying portfolio factors in the numerical integration. This can be achieved either by just simplifying the factor model per se, or by using a full factor model and then condensing the information contained in it by some algorithm. In this article we want to present a new approach for the last type of solution.

One of the most popular and widely applied factor reduction techniques should be the Principal Components Analysis (PCA), as e.g. presented in a credit context in Andersen et al. (2003). Having its origins in multivariate statistics, the underlying idea of the PCA is to reduce the number K of variables observed for a set of individuals by using a small number K^* of linear recombinations. Simultaneously, one wants to transfer as much discriminating power as possible, i.e. variation, from the old to the fewer new variables. Mathematically it turns out that the linear transformations are simply the set of the K^* Eigenvectors with the largest Eigenvalues for the covariance matrix. This fact makes the PCA also easily applicable in a credit framework. In credit portfolio modelling the PCA factor reduction works very well, as long as the number K^* is chosen large enough to really cover most of the portfolio correlation characteristics. If K^* is not large enough, this approach has two important shortcomings. The first is, because the PCA extracts 'the principal components', the part of the correlation matrix discarded usually still describes some correlation 'structure'. In practice this means often, that systematically the features of small subsets are cut off, and not just some random characteristics will get lost. You could compare this to the case where in a linear regression an important variable is missing and therefore the errors are no longer *iid* normal. We will give an example for this fact later on. The second problem is based in the construction of the new factors. The factors are calculated using information solely from the correlation matrix. Important figures like exposure sizes, default probabilities, or loss given defaults do not enter the calculations. In the PCA all assets are treated equally. However, as in the course of a factor reduction correlation information has to be discarded, it would appear natural to weight large contributors to the loss distribution higher than, e.g., small highly rated assets. In conclusion, though PCA is an extremely powerful tool in the context of multivariate normal distributions, as a factor reduction technique with respect to the calculation of loss distributions it is not optimal.

In this paper a new factor reduction approach is presented particularly designed to preserve as much information as possible for the loss distribution, while reducing the number of factors needed to just one core factor. In section 2, the general idea is outlined and how a different target function based on risk contributions could be defined to calibrate the new core factor model. Section 3 demonstrates the performance of the approach in synthetic and realistic data sets. The illustrative examples thereby include the calculation of loss distributions and risk contributions, as well as the application in an optimisation framework. The paper closes with a short discussions of open questions and future perspectives for this approach.

2. The Core Factor

This section shall outline the underlying idea of how to extract a single core factor from a multi-factor model to approximate the loss distribution. In the following let us consider a portfolio of N credit risky instruments or assets $a_i,\ i = 1, \ldots, N$, each with exposure V_i. Let the probabilities of default (PD) of these assets be denoted by $p_i,\ i = 1, \ldots, N$. For simplicity, let us assume that all assets are non-amortising with the same fixed maturity, and all to have a loss given default of 100%. Further, suppose the respective instruments follow a K-factor asset value model, as generally introduced by Merton (1974) or Vasicek (1991). Let $\beta_{ik},\ k = 1, \ldots, K$ denote the factor loadings for asset i. In this framework it is assumed that an instrument or firm defaults when its asset value process X_i falls below the firm's liabilities or a certain default frontier. The corresponding default frontier or threshold is determined by its individual PD via $DF_i = \Phi^{-1}(p_i)$. In its discrete version on a fixed horizon, the underlying asset return process and mutual covariances are defined as follows:

$$\begin{aligned} X_i &= \sum_{k=1}^{K} \beta_{ik} Z_k + \sqrt{1 - \sum_{k=1}^{K} \beta_{ik}^2}\; Z_i^* \qquad (1) \\ \text{with} \qquad & \mathrm{Covar}(X_i, X_j) = \sum_{k=1}^{K} \beta_{ik}\beta_{jk} \\ \text{and} \qquad & Z_k, Z_i^* \sim \mathbf{N}(0,1) \;\; i.i.d., \;\; i = 1, \ldots, N, \; k = 1, \ldots, K \end{aligned}$$

If we define the default variable $D_i = \mathbf{1}_{(X_i < DF_i)}$ as 'asset i defaults', then we have the following relations:

$$\begin{aligned} P(D_i) &= P(X_i < DF_i) \; = \; p_i \qquad (2) \\ \mathrm{Var}(D_i) &= p_i \times (1 - p_i) \qquad (3) \\ \mathrm{Covar}(D_i, D_j) &= \mathbf{E}(D_i D_j) - \mathbf{E}(D_i) \times \mathbf{E}(D_j) \\ &= \mathbf{E}\left(X_i < \Phi^{-1}(p_i) \cap X_j < \Phi^{-1}(p_j)\right) \; - \; p_i p_j \\ &= \mathbf{\Phi_2}\left(\Phi^{-1}(p_i), \Phi^{-1}(p_j), \rho_{ij}\right) \; - \; p_i p_j \qquad (4) \end{aligned}$$

For these formulas the asset correlation has to be specified. But as the asset value process' variance at the fixed horizon is defined to be 1, the asset correlation equals the asset covariance $\rho_{ij} = \sum_{k=1}^{K} \beta_{ik} \times \beta_{jk}$. Aggregation over the portfolio via $D_{PF} = \sum_{i=1}^{N} D_i$ yields

$$\begin{aligned} \mathrm{Var}(D_{PF}) &= \sum_{i=1}^{N}\sum_{j=1}^{N} V_i V_j \; \mathrm{Covar}(D_i, D_j) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} V_i V_j \; \left(\mathbf{\Phi_2}(\Phi^{-1}(p_i), \Phi^{-1}(p_j), \rho_{ij}) \; - \; p_i p_j \right) \end{aligned}$$

and as risk contributions to the portfolio variance we have for asset i

$$\frac{\partial \mathrm{Var}(D_{PF})}{\partial V_i} = 2 \sum_{j=1}^{N} V_j \; \left(\mathbf{\Phi_2}(\Phi^{-1}(p_i), \Phi^{-1}(p_j), \rho_{ij}) \; - \; p_i p_j \right) \qquad (5)$$

However, at this point we want to introduce a new separate model with only a single core factor β^{CF}, with $\beta_i^{CF}, \; i = 1, \ldots, N$, the respective factor loadings for the individual assets. For the new model, everything stays the same as for the old one except for the covariances, i.e. the correlations:

$$\begin{aligned} \text{Multi Factor Correlation for } i \neq j\text{:} \qquad \rho_{ij} &= \sum_{k=1}^{K} \beta_{ik} \times \beta_{jk} \qquad (6) \\ \text{Core Factor Correlation for } i \neq j\text{:} \qquad \rho_{ij}^{CF} &= \beta_i^{CF} \times \beta_j^{CF} \qquad (7) \\ \text{and in both cases for } i = j\text{:} \qquad \rho_{ij} &= \rho_{ij}^{CF} \; = \; 1 \end{aligned}$$

Therefore we would also only have to change the loss distribution's variance

$$\mathrm{Var}_{CF}(PF) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} V_i V_j \;\left(\mathbf{\Phi_2}(\Phi^{-1}(p_i), \Phi^{-1}(p_j), \rho_{ij}^{CF}) \;-\; p_i p_j\;\right) \tag{8}$$

and the portfolio risk contributions, which should be straightforward.

In a multi-factor normal copula world, the factor model defines the covariance/correlation matrix, which in turn defines together with PDs, etc., the risk contributions. And these again define the loss distribution's variance and other statistics. However, this almost follows some Markov-*esque* principle. If we know the correlation matrix, we no longer need the factor model, and if we have, e.g., the risk contributions for the variance, we no longer need the pairwise correlations to calculate the variance. This means on the other hand, that we condense a sometimes over-specified factor model into pairwise correlations and further into a set of just N risk contributions. Each set of parameters on its own is sufficient to calculate the variance. Therefore, if we would have a new model where we have to calibrate a set of N free parameters to approximate a loss distribution or its variance, we probably would focus on the level with the fewest parameters to match, but which is still sufficient for our purposes. The underlying idea of this core factor approach is to fit the new one-factor model to replicate a set of given variance contributions. If we can replicate a given portfolio loss variance, we would argue, the same model should also allow us to approximate the whole loss distribution reasonably well. To put this in a more mathematical wording, let us define a distance function for the calibration of the new model according to

$$\mathbf{D}_{CF} \;=\; \sum_{i=1}^{N} V_i \left(\frac{\partial \mathrm{Var}(PF)}{\partial V_i} \;-\; \frac{\partial \mathrm{Var}_{CF}(PF)}{\partial V_i}\right)^2 \tag{9}$$

$$\frac{\mathbf{D}_{CF}}{2} \;=\; \sum_{i=1}^{N} V_i \left(\sum_{j=1}^{N} V_j \;\left(\mathbf{\Phi_2}(\Phi^{-1}(p_i), \Phi^{-1}(p_j), \rho_{ij}) \;-\; \mathbf{\Phi_2}(\Phi^{-1}(p_i), \Phi^{-1}(p_j), \rho_{ij}^{CF})\right)\right)^2 \;, \tag{10}$$

i.e. penalising differences in the variance risk contributions in the two models. The vector $\widehat{\beta^{CF}}$ minimising the distance function

$$\widehat{\beta^{CF}} \;=\; \min_{\beta^{CF}} \; \mathbf{D}_{CF} \tag{11}$$

is the vector of the core factor loadings for the new model.

So, what have we achieved by this. We fitted a new one-factor model to approximate the variance contributions for the portfolio loss distribution as closely as possible, measured by our distance function $\mathbf{D}_{CF}$. However, as mentioned earlier, we hope that in a normal copula world this should also allow us to approximate the whole loss distributions fairly well. From this one-factor model it would easily be possible to very efficiently calculate the loss distribution via numerical integration and all statistics of interest, e.g. expected shortfalls or tranche risk contributions. This would be far more laborious in a multi-factor setting. It might sound a bit preposterous, but there are a couple of points having to be borne in mind. Surely enough we are no longer in a position to describe accurately pairwise correlations, however, that's also not been our main goal. We want to model the loss distribution and for this the variance contributions should allow for a good approximation. And as we have to approximate only N risk contributions with our N free parameters we would expect to get far better results than if we would want to match all $N \times (N-1)/2$ pairwise correlations. Further, as the core factor's calibration includes every variable affecting the loss distribution, we would argue that most of the important information enters the calibration process. Unlike in the standard PCA, big variance contributors will also have a heavy impact in the calibration, which is a very desired property. The quality and degree of accuracy of this approach will be demonstrated in the next section, where for some synthetic and realistic datasets the new core factor will be calibrated.

In terms of implementation, it should be mentioned that the calibration could be performed with a simple gradient or Gauß method. Because the underlying functions are smooth and monotonous, convergence is rarely an issue. As we will demonstrate in the next section, the calibration works very well for portfolios with up to 400-500 instruments. However, especially for optimisation algorithms some thought should be put into an efficient calibration algorithm, as a number of costly bivariate normal distributions have to be evaluated repeatedly. It is also worth mentioning, as the asset exposures change during an optimisation, the calibration to the (new) risk contributions would have to be performed anew in every iteration. In any case, the most time consuming first calibration (in further optimisation steps the old core factors can be used as starting point) takes in our C++ implementation a couple of minutes for the above portfolios. Further, to avoid an early termination of the algorithm, one also should ensure, that the approximation error in the core factor model w.r.t. a desired statistic is smaller than the improvement steps of the optimisation.

3. Applications

To illustrate the performance of our approach, we apply it in several examples to two different portfolios. One portfolio is synthetically generated and shall us allow to analyse some of the properties and capabilities of the approach. The second portfolio shall allow us to examine the core factor model under more realistic circumstances, and should represent a large well diversified portfolio, as can be found as underlying of, e.g., standard balance sheet SME CLOs.

The Portfolios

The first portfolio consists of 100 assets, each with the same maturity, PD, coupon and loss given default. The exposures are allocated equally to 10 buckets of 10 assets with 2mln to 18mln volume. Further, we varied industry and country specifications according to the scheme outlined in table 4. In the full factor model there are 1 global factor, 4 industry factors and 4 country factors. A 5% global correlation has been assigned together with 20% intra-industry correlation and another 5% intra-country correlation. Two assets from the same country and industry would therefore have a 30% asset correlation, whereas assets from different countries and industries would be correlated by 5%. In a variation of this portfolio, for the analysis how concentrations are handled by the new model, we raised the volume of one single assets to 110mln, representing 10% of the portfolio then. Further, we erased industry and country factor loadings for one asset, thus reducing its R^2 to 5%, and increased factor loadings to $\sqrt{15\%}, \sqrt{35\%}$, and $\sqrt{20\%}$ for another asset to generate an R^2 of 70% for it. The second portfolio shall represent a large well diversified standard balance sheet SME CLO portfolio. It consists of 400 assets in more than 25 industries and countries, with an average rating of BB-/Ba3. The total pool size is 2.000mln, the maximum single obligor concentration is approx. 1%.

Risk Contributions and Loss Distributions

Firstly, for both portfolios we compare variance contributions and loss distributions calculated by the different approaches. As a gold standard, we calculate these quantities with the full model. Additional to this, we use a model based on an approximation of the covariance matrix by three principal components, and finally our core factor model. Though generally for these applications a factor reduction is not necessary, it allows us to get an idea about the quality of the approximations. For all models variance contributions can be calculated analytically from their respective covariance matrices. This means the true covariance matrix for the full model is used and its approximations for the two reduced factors solutions. The loss distributions can be determined efficiently using Monte Carlo methods for the full model and with Fourier transformations (see Merino and Nyfeler, 2002, for the general idea) for the core factor and PCA approximations, using numerical integrations over the factors. Figure 1 shows the results for the synthetic dataset. In Subfigure 1a a very good fit from the core factor model for the variance contributions can be seen, simply superimposing the true contributions. We plotted the results of the different approaches against the true variance contributions, so the main diagonal represents a perfect fit. This can be seen in the light grey points for the results of the full model results. The earlier mentioned systematic bias in the PCA approximation is also quite apparent, where most of the contributions are approximated very well. However, the results for a handful of assets to the left are relatively poor. The approximations of the true loss distribution and its tail are displayed in Figures 1b and c, and are very good for both the core factor model as well as the PCA approach (for reasons of clarity not displayed). Thereby, the calculation with the core factor is substantially faster than with the three PCAs, despite the time needed for the calibration. As the calculation time increases exponentially with the factor dimension, the lower core factor dimension more than makes up for the time spent on calibration. Even for very few nodes in the numerical integration, e.g. 12, calculations were 10-15 times faster than with the higher dimensional model. To present a more quantitative view on the approximation, we further calculated the expected losses for a virtual tranching of this loss distribution, with the results shown in Table 1d. For seven tranches the attachment and detachment points are given, and we calculated the expected loss of these tranches with the full model and the core factor approach. With a relative error in the 2%-3% range the core factor approach matches the Monte Carlo results astonishingly well, bearing in mind that we use only one factor instead of nine in the original model. For the most senior tranche we would argue, that due to the finite number of Monte Carlo simulations on one side and the continuous character of the Fourier transformations on the other, the bigger relative error in the results should have more numerical than theoretical reasons.

Figure 2 shows the variance contributions for the synthetic data set with concentrations in terms of volumes as well as correlations. Here the same degree of accuracy can be achieved, even the 10% volume concentration in the upper right corner is captured very accurately.

For the realistic data set, we performed the same set of analyses as for the synthetic data and the results are displayed in Figure 3. Again we can show a very good approximation with the core factor approach to the true variance contributions. However, for few assets a suboptimal fit can be observed. After a bit of fine tunig of the algorithm, these could be brought to match the true values as well. But as we wanted to use the same standard algorithm for all applications in this paper, we decided to display these slightly worse results, to properly illustrate the capabilities of the approach in general realistic conditions. For this application the results for the PCA build a noticeable 'cloud' around the true contributions. This could be a result of the dramatically increased number of factors which determine the covariance matrix. Nevertheless, the resulting loss distributions are again for both approximations very accurate. The PCA approach has again been omitted from Figures 3b and c for reasons of clarity. For another virtual tranching for this data set, the relative errors also show a great degree of accuracy, when using the core factor approach.

In conclusion, we can state that where we have a gold standard, the core factor allows to approximate the true variance contributions and the whole loss distributions to a remarkable degree of accuracy, even though it is using only one factor in the copula approach.
In a next step, we also want to calculate the expected shortfall contributions to the 99% quantile. These statistics are displayed in Figure 4, calculated for the core factor and the PCA approach. Such figures can be of great importance when, e.g., calculcating hedge ratios for CDO tranches, and are easily accessible numerically for the Fourier transfomation approaches. With the Monte Carlo algorithm in the full model a calculation is in general too time consuming in order to get stable results. It should be mentioned that in this case, as we do not have the results for the full model, the results are plotted against the core factor results. Therefore, the latter naturally appear on the diagonal. The results in Figure 4 show a bit more variation for the synthetic dataset compared to the variance contributions, whereas the results for the realistic data appear very similar. The increased variation for the synthetic data show a similar systematic pattern as earlier. This could be explained once by the different references in the plots, but also by the slightly worse fit in the quantiles by the PCA approach. For example for the latter, the relative error in the virtual tranching for tranches B, C, and D, would be -5.12%, -4.84%, and -4.20%, respectively, instead of the results shown in Table 1d. However, as we can not provide a gold standard in this case, we leave the final judgement to the interested reader.

Optimisation

In a next step we illustrate the performance of the core factor model when using it in optimisation routines. As a target we use once an unexpected loss minimisation and later an expected shortfall minimisation. To correct slightly for the strong effects the ratings have in the realistic data, we decided to keep constant not only the overall volume, but also the pool income. As lower rated assets in general are higher yielding, this has the effect that in the above optimisations all volume is not simply put into the highest rated assets. Individual allocations are also deterministically limited to a maximum of 5% of the pool volume. The results for the unexpected loss minimisation are for both data sets displayed in Figures 5a and b. As for the risk contributions, in the unexpected loss optimisation we are able to provide a gold standard, calculated with the full model. It is again depicted by the light grey line, and is superimposed with a very accurate match from the core factor results. The PCA approach shows not too bad results for most of the assets, however has quite an error margin in some allocations. It appears that the repeated redistribution of volumes according to variance contributions emphasised even more the small systematic differences as depicted in Figure 1a. For the realistic data set the differences in the results are less systematic, however, also here we can see variations around the true results in a comparable magnitude. Similar results are displayed for the expected shortfall optimisation in Figures 5c and d. When looking at the expected shortfall results for the synthetic data, one difference to the unexpected loss results is a slightly more accentuated separation between higher and lower risk contributors. To minimise an expected shortfall in the tail, it is apparently advisable to shift more volume in less correlated assets and by this to reduce the joint default probability for larger parts of the portfolio. In contrast to this, the unexpected loss minimisation appears to penalise concentrations in general a bit more. The results also demonstrate what substantial effect the systematic bias in the PCA approach can have. For example for the synthetic data set, as we have differences in the expected shortfall contributions for a handfull of assets (see Figure 4a), in an expected shortfall minimisation these assets get in the PCA approach allocations which are upto 20% higher than with the core factor. For illustration, we calculated the expected shortfall for both 'optimal' portfolios using the full model and the Monte Carlo approach. Starting at an expected shortfall of 9.14% for the original dataset, it showed an expected shortfall of 8.61% for the optimal portfolio from the core factor approach, and 8.90% for the PCA result, proving that the PCA result is indeed sub-optimal compared to the core factor results.

4. Discussion

In summary, in this paper we introduced a new factor reduction approach, allowing us to approximate the portfolio loss distribution to an astonishing degree of accuracy with just one core factor. A new one-factor model was calibrated to replicate a given set of portfolio loss variance contributions and could then be used to calculate the whole loss distribution. We could show, that besides the loss distribution, we can achieve good results using the core factor approach when determining other risk contributions, or when using it in optimisation algorithms. Thereby, the applications comprised synthetic as well as realistic data, as found in structured finance transactions. In addition, by cutting the number of dimensions in the numerical integration to just one, we simultaneously could reduce the computational burden dramatically.

Despite the above results, a number of questions still remained unanswered and are in need to be addressed. As we discard in the core factor approach a substantial amount of pairwise correlation information, the first question would be, what kind of information is needed to characterise a loss distribution uniquely. Usually we are given the asset volumes, the PDs, the LGDs, and the covariance matrix (e.g. in form of a factor model), which define the loss distribution. The above analyses suggest that instead of the covariance matrix also the assets' variance contributions might be sufficient to do this. This would mean that the very good fit we saw has not been just coincidence. It would justify to use a method, which was calibrated to variance contributions, to calculate quantiles and other loss distribution statistics. However, we can neither provide any mathematical proof for this assumption, nor are we able to quantify the amount of information we loose in case we do. It would be interesting to know, whether or to which extent variance contributions really are sufficient in a mathematical sense. Further, one observation we made was that the covariance matrix resulting from the single core factor is fairly different to the original one, yet both yield very similar loss distributions. This is also in line with observations form the structured finance market, where a plethora of public and proprietorial factor models are in use, which have very different factorisations, but nevertheless yield very similar loss distributions, i.e. tranchings. It would mean that as long as the variance contributions are the same, we would have a very subjective choice on how to populate the covariance matrix. However, it has to be stated unambiguously that this approach is by no means a supplement for a thorough calibrated factor model and should not be seen as such. It is calibrated to an 'existing' covariance matrix, and should be regarded as an information extract to facilitate the loss distribution calculation.

The second open question is a direct consequence of the first. If variance contributions are sufficient, or at least define the characteristics of interest for us, how good a fit can we achieve with just one factor. And, is there a direct analytical solution to calculate the core factors. Because of the two-dimensional normal distributions within the algorithm, we tend to answer the second part with 'no'. For the first part, in our research we experienced that the calibration of the core factor yield without exception an astonishingly accurate fit for all data sets used. As stated, the differences for the realistic data could have been avoided by a more finely tuned version of our general algorithm. Together with the fact that we have N free parameter to calibrate to N contributions, and that the partial derivatives of the normal distributions are smooth and monotonous functions in our parameters, we therefore are inclined to answer the first question with a not overly confident 'a perfect fit'. Nevertheless, if we are proven wrong, how many factors would we need to achieve the perfect fit? Due to the reduced number of parameters, in any case it should be a number far lower than N.

All above questions become even more difficult to answer, when we would want to leave the Gaussian copula world. If heavier tailed t-distributions or others are to be applied, the defined target function is probably no longer suited to yield accurate approximation results. Perhaps higher order moments would have to be integrated. Nevertheless, the central question remains the same: 'How many parameters do I really need to define the loss distribution uniquely?'

Apart from the open theoretical questions, there are also some practical problems in need to be addressed. So far we calculated the core factor for a static portfolio of bullet assets with fixed maturity, PD, and LGD. Naturally, the target function (9) would have to be adjusted if we were to allow for some more flexibility. The first that comes to mind would be a term structure for the PDs, potentially together with amortising assets. Time varying PDs or amortising assets would mean the variance contribution of an asset might change over time. Currently the core factor would have to be calibrated for every time point anew. A more integrated approach would be desirable. Another line of development would concern the calibration for larger data sets, i.e. more than 1000 assets. We hadn't convergence problems for the examples we showed so far, however, by the definition of the calibration, there is no doubt that at some point they will arise. Again, a direct approach should be far more robust and time efficient.

References

Andersen, L., Sidenius, J. and Basu, S. (2003). All your hedges in one basket. *Risk,* **November 2003**, 67–72.

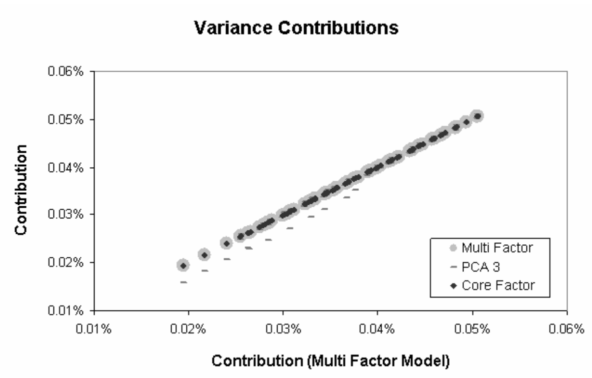
Merino, S. and Nyfeler, M. (2002). Calculating portfolio loss. *Risk,* **August 2002**, 82–86.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* **29**, 449–470.

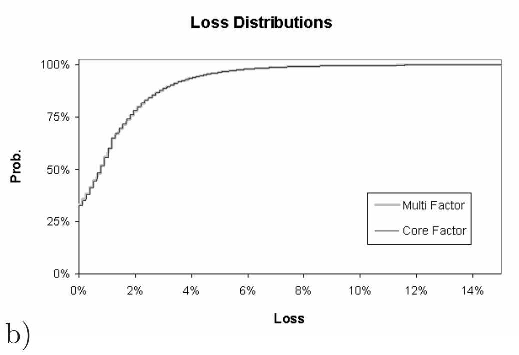
Vasicek, O. A. (1991). The loan loss distribution. Technical report, KMV Corporation.

	Country 1	Country 2	Country 3	Country 4	Total
Industry 1	13	11	8	5	37
Industry 2	11	9	7	3	30
Industry 3	8	7	5	2	22
Industry 4	5	3	2	1	11
Total	37	30	22	11	100

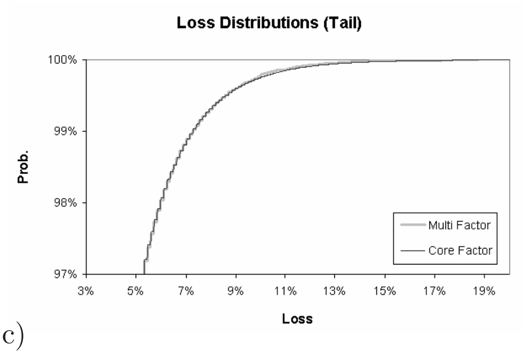
Table 1: Synthetic test portfolio.



a)



b)



c)

Tranche	A	B	C	D	E	F	G
Detachm. Point	100.00%	8.50%	7.50%	6.50%	5.50%	4.50%	3.50%
Attachm. Point	8.50%	7.50%	6.50%	5.50%	4.50%	3.50%	0.00%
EL Multi Factor	0.011%	0.690%	1.196%	2.086%	3.645%	6.494%	31.136%
EL Core Factor	0.011%	0.689%	1.169%	2.018%	3.542%	6.351%	31.131%
Rel. Error	9.09%	-0.12%	-2.30%	-3.28%	-2.83%	-2.20%	-0.01%

d)

Figure 1. Variance contributions and loss distributions for the synthetic dataset.

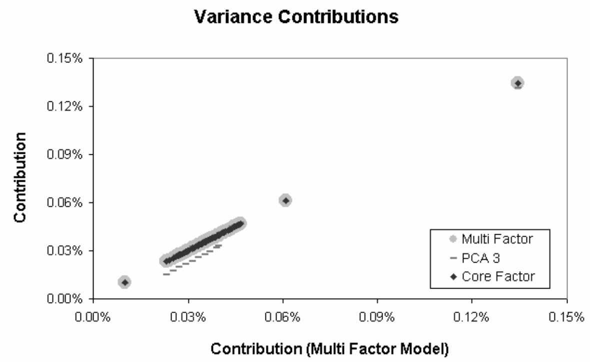
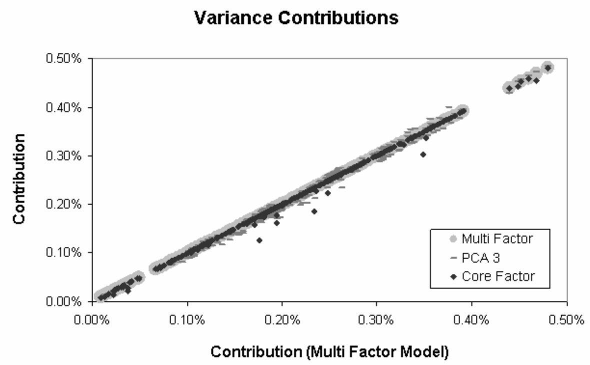
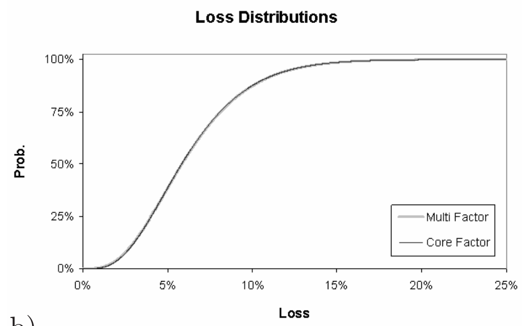


Figure 2. Variance contributions for the synthetic dataset with concentrations.



a)



b)

c)

Tranche	A	B	C	D	E	F	G
Detachm. Point	100.00%	18.50%	16.50%	14.50%	12.50%	10.50%	8.50%
Attachm. Point	18.50%	16.50%	14.50%	12.50%	10.50%	8.50%	0.00%
EL Multi Factor	0.004%	0.389%	1.116%	2.918%	7.005%	15.601%	68.253%
EL Core Factor	0.004%	0.395%	1.117%	2.910%	6.988%	15.415%	68.168%
Rel. Error	3.21%	1.62%	0.14%	-0.29%	-0.24%	-1.20%	-0.12%

d)

Figure 3. Variance contributions and loss distributions for the realistic CLO dataset.

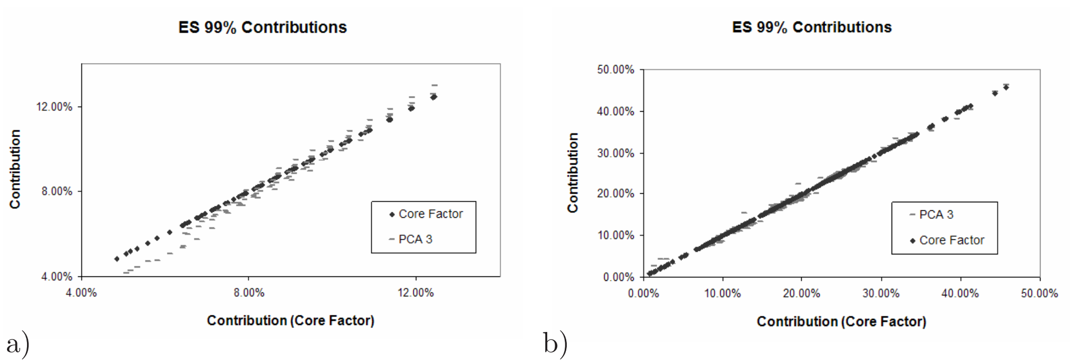


Figure 4. Expected Shortfall 99% Contributions for the synthetic (a) and realistic CLO (b) dataset.

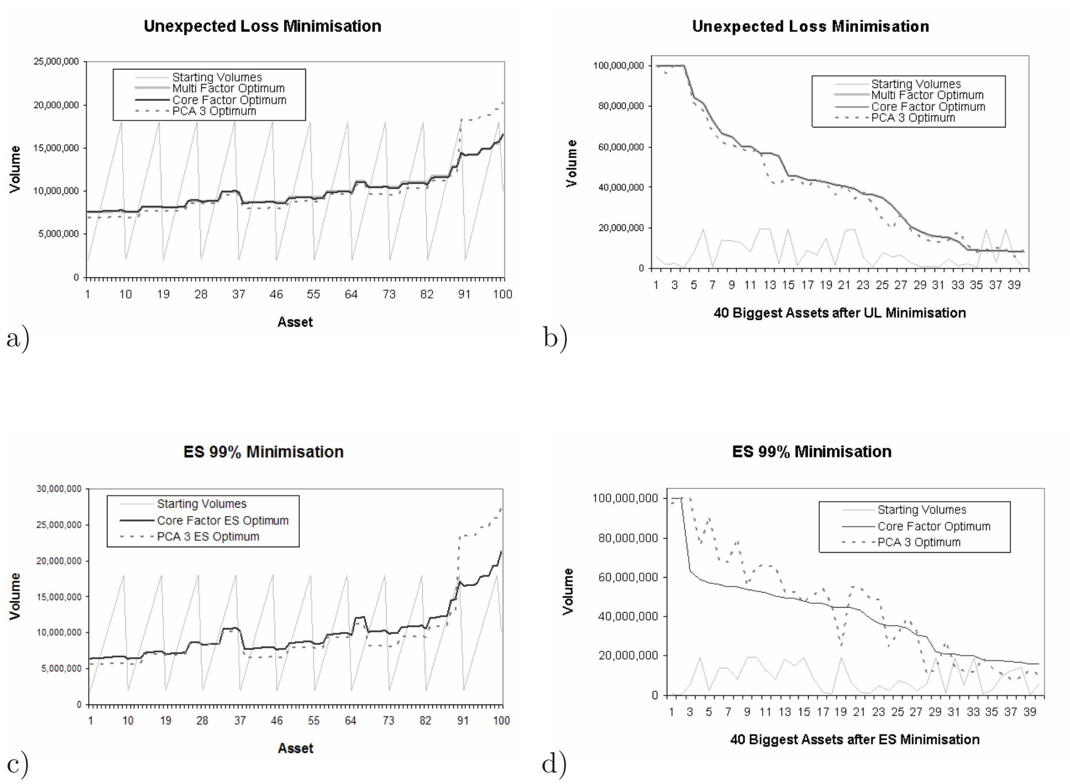


Figure 5. Optimisation results for the synthetic (a,c) and realistic CLO (b,d) dataset.