Auszug Publikationen 2004

- **Estimation of Sector Weights from Real-World Data**

Wir denken nach, um vorzudenken

Estimation of Sector Weights from Real-World Data

Dr. Michael Lesko / Dr. Frank Schlottmann / Stephan Vorgrimler

Gundlach, M. / Lehrbass, F. (Eds.): CreditRisk+™ in the Banking Industry

Summary

We discuss four different approaches to the estimation of sector weights for the CreditRisk+ model from German real-world data. Using a sample loan portfolio, we compare these approaches in terms of the resulting unexpected loss risk figures.

Introduction

The CreditRisk+ model in its general sector analysis formulation assumes the presence of common independent systematic risk factors that influence the default rates of all obligors in a given credit portfolio in order to incorporate default dependencies. The levels of implicit default correlation depend on volatilities of the risk factors and on the factor loadings, also called risk factor weights, of the obligors. We discuss different estimation methods for these model parameters from historical macro-economic data and investigate the empirical differences concerning the resulting loss distribution and the risk measures (e.g. unexpected loss / credit-value-at-risk) using a sample portfolio.

To avoid unnecessary complexity, we assume a one-period calculation for a given portfolio of K corporate obligors using N systematic risk factors (sectors) and an idiosyncratic risk factor for a time period of one year as introduced in Chapter 2. For multiple-period calculations (hold-to-maturity) see [3].

In the following sections we assume given parameters α_k for each random scalar factor S_k, such that the respective second parameter β_k of the gamma distribution, see Chapter 2, for sector k is then determined by

$$\beta_k := \frac{1}{\alpha_k}.$$

When parameterizing from observable risk factors one could choose

$$\alpha_k := \frac{1}{var\,[S_k]}.$$

(recall that $\mathbb{E}\,[S_k] \equiv 1$). If the risk factors are not directly observable the choice of the α_k values leads to N additional degrees of freedom (since there are N systematic risk factors). This topic is addressed later in this chapter.

1. Methods for the Estimation of the Risk Factor Weights

1.1 Plain vanilla industry sectors as risk factors

An obvious approach that is motivated by the choice of countries as risk factors in [3, A.7] is to introduce a separate risk factor for each industry within a given country. This is a common approach in real-world applications that allows a straightforward choice of the risk factor weights: each obligor A is allocated to the risk factor k_A that matches his industry classification best, i. e.

$$w_{Ak_A} := 100\%, \qquad \forall k \neq k_A : w_{Ak} := 0\% .$$

Since the respective industry classification is quite easy to obtain either from the obligor's data stored in the usual databases of financial lending institutions or from publicly available sources, the simplicity of this approach with respect to an IT-based data providing process seems very appealing.

However, the CreditRisk+ model assumes the risk factors to be independent, and within most countries, this will most probably not be a valid assumption for the different industries. For instance, we have analysed macro-economic data provided by the German federal statistical office (Statistisches Bundesamt) and found correlation coefficients of about 80% between pairs of the annual historical default rates within different West German industries in the years 1962 – 1993.[1] In the empirical section the consequences of neglecting these industry default rate correlations will be analyzed. Therefore, such a simplified approach might lead to significant underestimation of unexpected losses. A sample result illustrating this fact is presented later in this chapter.

We now discuss different alternatives for the estimation of risk factor weights that incorporate correlations between industries.

1.2 The approach of Bürgisser, Kurth, Wagner and Wolf

Bürgisser et al. [2] developed an approach that incorporates correlations between the risk factors, see Chapter 9. The authors calculate the standard deviation of total portfolio losses assuming correlated risk factors (denoted by $\sigma_{X,corr}$ below) using the following formula (cf. [2, Eg. (12)]or 9.11)):

$$\sigma^2_{X,corr} = \sum_k \sigma_k^2 EL_k^2 + \sum_{k,l \wedge k \neq l} corr(S_k, S_l)\sigma_k\sigma_l EL_k EL_l + \sum_A p_A \nu^2_A \qquad (1)$$

where $corr(S_k, S_l)$ denotes the correlation between two distinct sector variables.

The total portfolio loss distribution is then calculated using the standard CreditRisk+ algorithm for the given portfolio data under the assumption of one systematic risk factor $\widehat{S_1}$. The parameters of the Gamma distribution for $\widehat{S_1}$ are calibrated such that the resulting standard deviation of portfolio loss σ_X obtained by application of the standard CreditRisk+ model matches the standard deviation specified in (1):

$$\sigma_X = \sigma_{X,corr.}$$

After calibrating $\widehat{S_1}$, the desired unexpected loss figures can be computed from the output of the standard CreditRisk+ algorithm, e. g. the unexpected loss at the 99th percentile level (denoted by $UL_{0.99}$).

For a sample portfolio, Bürgisser et al. reported a higher standard deviation of about 25% concerning the total portfolio loss, and the corresponding unexpected loss at the 99th percentile level was about 20% higher compared with the results obtained by using the approach from Section 1.1.

This approach is very interesting for real-world applications, since it directly incorporates the observed correlation parameters into a transparent and easy-to-implement calibration process, and no additional information or modelling assumption is required. However, from the perspective of the model-theoretic assumptions, it has to be kept in mind that the resulting distribution of total loss from the given portfolio is not obtained under the assumption of correlated risk factors, but merely a distribution that has the same standard deviation as the total losses that result from the portfolio if the risk factors are correlated. In other words, this avoids the problem of using correlated risk factors in the CreditRisk+ calculations in an appealing way, but e. g. the percentile figures calculated by the calibrated one risk factor approach might not be very precise, and nothing is known about the true level of unexpected loss.

1.3 The approach of Lesko, Schlottmann and Vorgrimler

In Lesko et al. [6] we have proposed the estimation of risk factor weights by principal component analysis (PCA), see e. g. [7] for a comprehensive introduction of this method. The basis for this approach are abstract, independent risk factors (as assumed by standard CreditRisk+) which explain the evolution of default rates within the different industries. All obligors that have the same industry classification are assumed to share the same risk factor weight within this industry.

The first goal for the risk factor weight estimation under these assumptions is to identify appropriate independent risk factors. Secondly, the influence of these risk factors on the respective industry (and therefore, on the obligors that belong to this industry) has to be estimated. The PCA method provides a framework that allows the simultaneous identification of independent risk factors and the calculation of the corresponding risk factor weights. Using this method, the abstract risk factors and their influences on the respective industry are obtained directly from the volatilities between default rates in the considered industries and the correlations between these default rates. In general, the number of systematic risk factors can be chosen arbitrarily between 1 and the number of given industries, which we denote by J below. When choosing this parameter, it should be kept in mind that for a higher number of abstract risk factors, the accuracy of the resulting risk factor weights is also higher in terms of the implicit correlation structure represented by these weights.[2]
Our algorithm works basically as follows:

Algorithm 1. *PCA approach for CreditRisk+*

Input: *positive semidefinite covariance matrix M_C of dimension J x J*
1: *Calculate the eigenvalues and the corresponding eigenvectors (column vectors) of the covariance matrix M_C*
2: *Build diagonal matrix M_D containing the eigenvalues*
3: *Build Matrix M_E containing the eigenvectors as columns*
4: *Choose the number N of systematic risk factors*
5: *Truncate the matrix of eigenvectors to the number of systematic risk factors by deletion of the eigenvector columns having the smallest associated eigenvalues, also delete the columns and rows of these eigenvalues from M_D (final result of this operation: J x N matrix $\widetilde{M_E}$ and N x N matrix $\widetilde{M_D}$)*
6: *Incorporate idiosyncractic risk factor by adding a column to $\widetilde{M_E}$ and a column of zeros as well as a row of zeros to $\widetilde{M_D}$ (result: J x $(N+1)$ matrix $\widehat{M_E}$, $(N+1)$ x $(N+1)$ matrix $\widehat{M_D}$)*
Output: *risk factor weight matrix $\widehat{M_E}$, risk factor variances as diagonal elements of $\widehat{M_D}$.*

The input covariance matrix M_C is estimated from the given correlations and the given volatilities of the industry default rates. The diagonal matrix M_D contains the eigenvalues of M_C, and each column of the matrix M_E contains the eigenvector associated with the eigenvalue in the corresponding column of M_D. These matrices satisfy the property

$$M_C = M_E \; x \; M_D \; x \; (M_E)^T \tag{2}$$

where $(M_E)^T$ is the transposed matrix of eigenvectors and x denotes the usual product of matrices.

The number of systematic abstract risk factors N is determined by analyzing the eigenvalues of M_C first and choosing N depending on these eigenvalues. For instance, a single large eigenvalue indicates that only one systematic risk factor explains most of the systematic risk in the economy specified by the given industries' data.

In the next step (cf. line 5) the risk factor allocation based on the chosen number of systematic risk factors is derived by deletion of the columns from the matrix of eigenvectors $\widehat{M_E}$ that correspond to the $J - N$ smallest eigenvalues in M_D. These eigenvalues are also removed from M_D by deletion of the respective row and column vectors.

After this construction of the matrices $\widetilde{M_D}$ and $\widetilde{M_E}$ the resulting covariance matrix $\widetilde{M_C}$ can be checked in analogy to (2) to validate the calculated risk factor allocation parameters:

$$\widetilde{M_C} = \widetilde{M_E} \times \widetilde{M_D} \times (\widetilde{M_E})^T \tag{3}$$

In line 6, an idiosyncratic risk factor is included by adding a column to the truncated matrix of eigenvectors $\widetilde{M_E}$ such that the sum for each row of the resulting matrix $\widehat{M_E}$ equals 1. Moreover, an additional row of zeros and an additional column of zeros for the specific sector have to be added to the truncated matrix $\widetilde{M_D}$. Then, the final output of Algorithm 1 is the resulting risk factor weight matrix $\widehat{M_E}$ (each row i contains the risk factor weights of an obligor belonging to industry $i \in \{1, \dots, J\}$) and the matrix $\widehat{M_D}$ containing the volatilities of the risk factors as diagonal elements.

We consider the following sample correlation matrix in our calculations below:

Both variants lead to the same standard deviation of total losses from the portfolio, but the resulting unexpected loss figures are different in general, cf. the corresponding remarks by Gordy [5]. We followed the first path and assumed fixed parameters of the S_k random scalar factors in our above example, which lead to the shown calibration of the risk factor weights.

1.4 Single risk factor for the whole economy

Compared with the second and the third approach, which are motivated by mathematical considerations, the following approach is suggested by empirical observation of historical macro-economic default rates. As already mentioned in the above sections, the statistical analysis of historical default rates within German industries yields correlations of about 80% and above. Using the CreditRisk+ approach of abstract background risk factors, this suggests that there is only one systematic background risk factor (e. g. the general state of the German economy) that influences the default rates of all German obligors uniformly.[3] Under this assumption, a single risk factor approach seems appropriate, the risk factor weights of all obligors are assumed to be $w_{A1} \equiv 100\%$, and the free parameter of the single risk factor is obtained e. g. by calibration of β_1 according to an observable variance of the German all-industries corporate default rate.

This is a straightforward approach that is somewhat similar to the approach described in section 1.1, but this time, the implicit error moves the unexpected loss results into the opposite direction: the above choice of the risk factor weights and the calculation of a standard deviation for the single risk factor is similar to an im-

Table1: Sample correlations between industry sectors in Germany

Industry	Agriculture	Manufacturing	Construction	Trade	Trasportion	Services
Agriculture	100%	70%	95%	94%	50%	96%
Manufacturing	70%	100%	72%	84%	90%	78%
Construction	95%	72%	100%	95%	45%	98%
Trade	94%	84%	95%	100%	64%	96%
Transportion	50%	90%	45%	64%	100%	51%
Services	96%	78%	98%	96%	51%	100%

The values in Table 1 reflect typical correlations between historical default rates of selected industries in West Germany. Obviously, the historical default rates are highly correlated, which underlines the criticism concerning the approach described in section 1.1 that neglects the correlation between the risk factors.

Assuming three abstract risk factors, an application of our approach based on principle component analysis on a small portfolio consisting of three obligors yields the results shown in Table 2. There are different possible variants of our approach, since the given volatility of the historical default rates within the industries can be respected either by the calibration of the risk factor weights or by the calibration of the abstract risk factors' parameters α_k, β_k.

Table 2: Input data for a sample portfolio using PCA output

A	Industry	p_A	σ_A	v_A	w_{A0}	w_{A1}	w_{A2}	w_{A3}
1	Construction	1.2%	0.9%	1,200,000	5.0%	67.7%	22.4%	4.9%
2	Trade	0.4%	0.3%	3,200,000	23.4%	70.9%	6.1%	-0.4%
3	Construction	2.8%	2.0%	400,000	5.0%	67.7%	22.4%	4.9%

plicit correlation between the default rates of 100% for each pair of industries, so this leads to an overestimation of the portfolio risk if the true correlation is lower, and this is e. g. the case for the historical West German data. In other words, this approach ignores possible diversification by investing in different industries. Moreover, if the correlation structure between industries is inhomogeneous, the accuracy of the CreditRisk+ calculations under this approach might not be appropriate.

In the next section, we compare the empirical results for a sample portfolio by application of the different approaches discussed so far.

2. Empirical Comparison

We now apply the four approaches of sector weight estimation for real-world data described in the previous sections to a typical German middle-market loan portfolio consisting of 1000 obligors within the six industry sectors shown in Table 1. The sum of net exposures of all obligors is approximately 1,000,000,000 EUR, whereas the individual net exposures (ν_A) are between 1,000 EUR and 35,000,000 EUR. The mean default probability of each obligor is in the range of $0.1\% \leq p_A \leq 8\%$.

The results of the CreditRisk+ model calculations depending on the chosen risk factor weight estimation approach are shown in Table 3. We use *EL* as an abbreviation for expected loss, *SD* for the standard deviation of total portfolio loss, and $UL\varepsilon$ is again the unexpected loss at the confidence level ε.

Looking at the different results depending on the chosen method of risk factor weight estimation, the first observation is that the choice of industry sectors as risk factors yields significantly lower measures of risk. This is due to the fact that the correlations between the industry sectors are ignored in this first approach of risk factor weight estimation, while the correlations are in fact significantly positive as shown in our example from Table 1. Hence, this approach truly underestimates the risk in our example and also in any other case where the correlations between industry sectors are all non-negative.

In contrast to that, the results of the other three approaches are quite similar concerning the standard deviation as a first risk measure. The standard deviation of the second and third approach is approximately the same, while the corresponding risk measure of the fourth approach is slightly higher. This is a result that is consistent with our considerations in the preceding sections, since the Bürgisser et al. and the PCA approach respect the pairwise correlations between the industry sectors that are <100% between different sectors in Table 1 while the single systematic sector approach ignores this diversification possibility. Hence, the latter approach overestimates the standard deviation.

The results concerning the UL_ε values are somewhat more different: the PCA yields the highest values. Comparing the second and the third approach, the result is not surprising, since there is a difference between the parameter estimation based on calibration of the standard deviation by Bürgisser et al. compared with the PCA transformation of the risk factor weights. The more interesting observation is that the UL_ε value of the single systematic sector approach (which overestimates the risk in general) is lower than the corresponding UL_ε value of the PCA approach.

A more detailed analysis reveals that the difference lies in the distinct risk factor weight estimation procedure. To achieve a higher default rate volatility (and therefore, higher default correlations) an important choice can be made within both the estimation of the risk factor weights and the estimation of the default rate volatilities: either higher default rate volatilities or higher risk factor weights can be chosen.[4] In our sample calculations, we have chosen a standard deviation of $\sigma_1 := 0.72$ for the single systematic risk factor (fourth approach) while we assumed a default rate standard deviation equal to *1* in the PCA approach. The choice of a higher default rate volatility leads to a higher UL_ε in many cases, and this effect even increases for $\varepsilon \rightarrow 1$, see also the corresponding results obtained by Gordy [5]. Hence, this is the main reason for the difference between both approaches concerning the percentiles computed by CreditRisk+.

Finally, to illustrate the differences between the two simplest methods of risk factor estimations presented in the previous sections, the two probability functions of portfolio loss are displayed in Figure 1, and their tails are shown in Figure 2. All loss values on the *x*-axis are in millions of euros.

Table3: Comparison of CreditRisk+ results for real-world portfolio

Approach	Description	EL	SD	$UL_{0.99}$	$UL_{0.999}$
1	Indurstry sector -> risk factor	21,365	15,080	49,370	75,894
2	Bürgisser et al.	21,365	19,441	66,195	104,231
3	PCA	21,365	19,445	68,963	112,450
4	Single systematic sector	21,365	19,844	67,936	107,328

Figure 1: Comparison of probability function of portfolio loss for the first approach (solid line) and the fourth approach (dotted line)

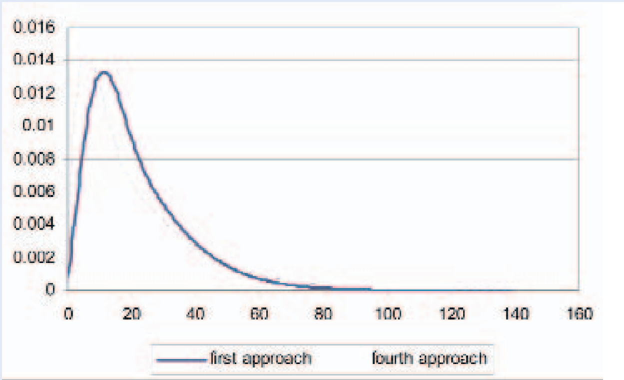
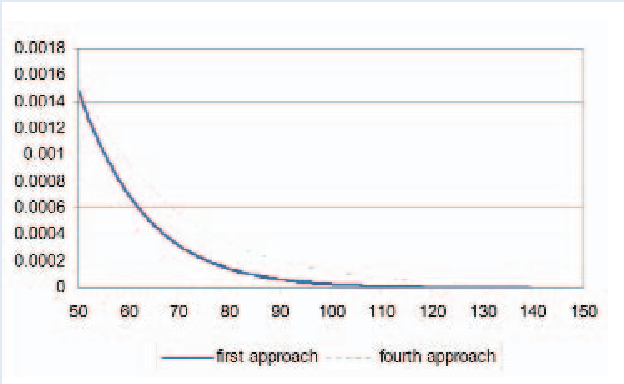


Figure 2: Comparison of the tail probability function of portfolio loss for the first approach (solid line) and the fourth approach (dotted line)



Clearly, the probabilities of large losses are higher for the fourth approach (cf. the tail of the dotted line in Figure 2). Hence, the corresponding risk measures are higher than its counterparts calculated using the first approach of risk factor weight estimation (cf. the solid line in Figure 2).

3. Conclusion

In the previous sections, we have discussed different alternative methods of risk factor weight estimation for CreditRisk+.[5] Summarizing our remarks and the empirical results obtained for a West German middle-market loan portfolio, the influence of the correlations between industry sectors should not be neglected. Moreover, we have pointed out that there is a degree of freedom in the joint estimation of risk factor weights and parameters of the systematic risk factor random variables, which can lead to significant differences concerning the tails of the resulting portfolio loss probability distributions. This degree of freedom and its consequences are an interesting point for future research activities.

From an application-oriented point of view, the use of a single systematic sector approach of risk factor weight estimation seems to be a good choice in many cases where banks hold portfolios of obligors that mainly depend on the state of a single economy (e. g. firms operating in one country). Besides general economic considerations, this is supported by our analysis of historical default rates in West Germany.

Moreover, from our point of view, the approach by Bürgisser et al. and the PCA approach are well-suited for credit portfolios that are diversified at least over two different countries not depending strongly upon each other concerning their state of the economy.

Autoren:

Dipl. Math. oec. Dr. Michael Lesko
Leiter Research Gesamtbanksteuerung bei GILLARDON.
Studium der Wirtschaftsmathematik und Promotion an der Universität Ulm. Begleitend zur Promotion Mitarbeiter am Institut für Finanz- und Aktuarwissenschaften (IFA), Ulm. Seit 1998 bei GILLARDON tätig mit dem Schwerpunkt Kreditrisikomodellierung und -systeme. Diverse Veröffentlichungen sowie Seminar- und Referententätigkeiten zu dieser Thematik.

Dipl. Wi.-Ing. Dr. Frank Schlottmann
Studium und Promotion an der Universität Karlsruhe (TH). Parallel dazu unternehmerische Tätigkeit im Bereich Informationstechnologie-Consulting und -Training. Seit 1994 bei GILLARDON tätig in den Bereichen Entwicklung, Beratung und Projekte mit aktuellem Schwerpunkt Kreditrisiko und Research. Zahlreiche internationale Publikationen und Vorträge im Bereich des finanziellen Risikomanagements.

Dipl. Math. Dipl.-Wirtsch.-Inform. Stephan Vorgrimler
Studium der Mathematik und Wirtschaftsinformatik an der Universität Mannheim. Seit 1996 bei GILLARDON tätig mit den Schwerpunkten Kreditrisiko und Research.

References:

[1] P. Bürgisser, A. Kurth and A. Wagner. Incorporating servity variations into credit risk. Journal of Risk, 3(4): 5 – 31, 2001

[2] P. Bürgisser, A. Kurth, A. Wagner, and M. Wolf: Integrating correlations. Risk, 12(7): 57 – 60, 1999.

[3] CreditSuisse Financial Products: CreditRisk+™ – A credit risk management framework. London, 1997.
Available at http:/www.csfp..com/creditrisk.

[4] M. B. Gordy. Saddlepoint approximation of CreditRisk+. Journal of Banking & Finance, 24: 119 – 149, 2000.

[5] M. B. Gordy. A comparative anatomy of credit risk models. Journal of Banking and Finance, 24(1-2):119–149, 2000.

[6] M. Lesko, F. Schlottmann, and S. Vorgrimler. Fortschritte bei der Schätzung von Risikofaktorgewichten für CreditRisk+. Die Bank 6: 436 – 441, 2001.

[7] I. Jolliffe: Principal Component Analysis, 2nd edition. Springer-Verlag, Heidelberg, 2002.

[8] T. Wilde: CreditRisk+. In S. Das, editor: Credit Derivatives and Credit Linked Notes, 2nd edition. J. Wiley & Sons, New York, 2000.

1 These coefficients are estimated in a simplified manner by measuring the default rate correlations. A more refined estimation procedure would constist of estimating correlations of forecast errors. In this case the estimatied correlations are lower, nonetheless they still have values significantly above zero.

2 Note that the numerical stability of the caluculations using the Panjer recursion in CreditRisk+ has to be watched carefully in the number of abstract risk factors in raised, cf. [8, 611 ff.]. When using a large number of abstract risk factors, one of the numerically more stable caluclation schemes presented in the second part of this volume is preferable.

3 The empirical data considered in the above analysis contains only defaults of corporate obligors. Nonetheless, the single risk factor approach is also straightforward approach for retail portfolios or mixed corporate/retail portfolios consisting solely of obligors the are assumed to depend on a single systematic background risk factor.

4 This has already been pointed out by Gordy [5].

5 Cf. also the alternative approach by Boegelein et al. in Chapter 14.

GILLARDON – innovative Lösungen für die Finanzwirtschaft

Die Lösungen

Unsere Kernkompetenzen umfassen die Bereiche Kundenberatung, Produktkalkulation und Gesamtbanksteuerung.

Kundenberatung

evenit™ ist das themenorientierte Beratungssystem für alle Vertriebskanäle für die Themen Altersvorsorge, Baufinanzierung, Vermögensanalyse und Financial Planning.

Produktkalkulation

MARZIPAN™ ist die Lösung zur Produktberatung und -kalkulation von Aktiv- und Passivgeschäften auf Basis der Marktzins- und Barwertmethode.

FinanceFactory™ ist das regelbasierte Kalkulationssystem für die Absatzfinanzierung, das alle Darlehensvarianten der Absatzfinanzierung inklusive Restkreditversicherung und Subventionsrechnung abdeckt.

Gesamtbanksteuerung

THINC™ ist die integrierte Softwarelösung zur wertorientierten Gesamtbanksteuerung und deckt die Themen Markt- und Vertriebssteuerung, Bilanzstrukturmanagement, Risikocontrolling, Treasury, Adressrisikosteuerung, Basel II und IAS / IFRS ab. THINC unterstützt Sie bei der Erfüllung der Anforderungen aus den MaRisk.

GILLARDON ist Branchenspezialist für Softwarelösungen, Consulting und Seminare in den Themenbereichen Kundenberatung, Produktkalkulation und Gesamtbanksteuerung.