Measuring Default Correlation

Krishan Nagpal and Reza Bahar

This paper examines the structure of the correlation between defaults of US corporates. This study is based on historical default data over a period of 18 years. The industry sector and the credit rating of the firm are shown to be major determinants of the correlations.

Observation of historical default rates supports the idea that default events (and, more generally, all indicators of credit quality and transition) are correlated. Default correlations are caused by similar economic conditions and, within a sector, by industry-specific reasons. However, incorporating default correlation in any portfolio credit risk analysis is difficult because of the lack of good data on default correlation, and the complexity of developing realistic models of default correlations that capture its dependence on credit quality, region, industry and time horizon.

The main objective of this paper is to provide a historical study of default correlation. Default correlation describes the relationship between default probabilities and joint default probabilities. Since a common economic environment is an important contributor of default correlation, we limit this study to US corporates because of the scarcity of default data outside the United States. The study is based on default data from 1981 to 1999 of US companies rated by Standard & Poor's (S & P). Based on the observed correlation, we also propose a compact two-factor parametrization of default correlation that is intuitive and captures the important observed features of the correlation. The second part of the paper builds upon the proposed correlation model and develops an analytic approach for credit risk analysis.

Different approaches have been proposed to model default correlation. Wilson (1997a, 1997b) has proposed a model where rating transition probabilities are driven by some economic factors that are themselves modelled as dynamic processes. In Gupton, Finger and Bhatia (1997), the underlying source of correlation is not modelled. Instead, default correlations are assumed to be exogenously given and modelled by incorporating correlation in the random variables that are used to model rating (or credit quality) changes of different companies. Whatever the mechanism for generating correlations, a general approach to represent correlation is through Copula functions (Li

2000). Li proposed using Copula functions to describe distribution of survival times of a portfolio of assets. Once the Copula function is known, one can easily extract all the joint default probabilities of any combination of assets over any combination of time intervals.

One of the difficulties in generating Copula functions is the need for various conditional default probabilities, which may be difficult to obtain. For this reason, in this paper, we try to describe and parametrize default correlation for only two assets at a time. Moreover, correlations are not modelled using any market observable indexes or economic factors. The only conditioning here is in terms of rating and sector. The default correlations, as defined here, describe the relationship between default probabilities and conditional default probabilities for any two assets based on their ratings and industry classifications. A more realistic model on default correlation should also incorporate conditioning in terms of the prevailing economic or industry-specific environment (for example, spreads, GDP growth rate, and the like). Since the impact of market factors is ignored here, these results should be viewed as empirical results on default correlation for the average economic environment of the past 20 years (the data used is from 1981 to 1999).

Historical data

All the results presented in this section are based on US corporate default data from January 1981 to December 1999. Our two objectives were to determine default correlation between companies in different industry sectors (default correlation attributable to common macroeconomic conditions) and companies within an industry sector (attributable to both economic conditions as well as industry-specific reasons). The S & P corporate data was divided into 12 industry categories. However, the analysis here is limited to 11 industry groups because telecommunications were excluded as there were too few defaults (seven) to draw any meaningful conclusions. Moreover, due to deregulation, technological changes and licence costs, the future of the telecommunications sector is unlikely to be similar to the past.

For any sector and/or region of interest, ideally, one would like to know default correlation between all rating combinations and all possible time horizons. Thus, if the rated universe is divided into seven categories, then the default correlation information of interest would be a seven-by-seven matrix that depends on the horizon. However, within any of the industry sectors there were not enough defaults to draw such detailed conclusions about default correlation. Instead of the usual seven rated categories, we chose to divide the rated universe into two categories—investment grade and non-investment grade. Investment-grade rating corresponds to an S & P rating from AAA to BBB–, while non-investment-grade ratings are BB+ or lower. This aggregation of ratings into two groups does have some drawbacks—the impact of the change in rating distribution over time is ignored and comparison of different industries is made harder because some of the discrepancies may be attributable to different rating distributions in different sectors.

Unless one has very long time series of default data, it is difficult to establish whether historical variations in default frequency are caused by changes in default rates or due to the correlations. Here, all the analysis is based on the assumption that individual and pair-wise default probabilities do not change over time. Once they are determined based on the empirical data, the default correlation can be uniquely determined (the relationship between individual and pair-wise default probabilities and correlations is described in Equation 1).

We briefly describe the procedure used for determining the default correlations. The first step was to determine the default probability for the relevant time horizon. Here, the rated universe in the beginning of each year from 1981 to 1999 was considered an independent data set. In any given data set, corporations that had not defaulted but had become unrated within the relevant time horizon were excluded from the analysis.[1] In what follows, let $N_i^j(T)$ and $N_n^j(T)$ be the total number of investment-grade and non-investment-grade companies respectively that were rated in the beginning of the *j-th* year and were either rated or had defaulted after *T* years and $D_i^j(T)$ and $D_n^j(T)$ be the total number of companies that had defaulted within T years that were investment-grade and non-investment-grade, respectively, at the beginning of the *j-th* year.

Investment-grade and non-investment-grade default probabilities over T years denoted by $p_i(T)$ and $p_n(T)$ were obtained as follows:

$$p_i(T) = \frac{\Sigma_j D_i^j(T)}{\Sigma_j N_i^j(T)} \qquad p_n(T) = \frac{\Sigma_j D_n^j(T)}{\Sigma_j N_n^j(T)}$$

Let $q_{ii}(T)$, $q_{nn}(T)$ and $q_{in}(T)$ be the pair-wise default probabilities for the three pair combinations—two investment-grade companies, two non-investment-grade companies and a pair with one investment-grade and one non-investment-grade company, respectively. They were obtained by determining how many pairs from the total number defaulted over the relevant horizon:

$$q_{ii}(T) = \frac{\Sigma_j D_i^j(T)(D_i^j(T)-1)}{\Sigma_j N_i^j(T)(N_i^j(T)-1)}$$

$$q_{nn}(T) = \frac{\Sigma_j D_n^j(T)(D_n^j(T)-1)}{\Sigma_j N_n^j(T)(N_n^j(T)-1)}$$

$$q_{in}(T) = \frac{\Sigma_j D_i^j(T) D_n^j(T)}{\Sigma_j N_i^j(T) N_n^j(T)}$$

As well as default probabilities, one may also be interested in the uncertainty associated with these default rates. Here, we assume that default rate observation from each year provides an independent observation of default rate. Instead of weighing the default rate for data from each year equally, each yearly observation is weighted by the number of the rated companies that year. More precisely, the variance (or the square of standard deviation) of investment-grade default probabilities over T years was obtained as:

$$\sigma_i^2(T) = \sum_j \frac{N_i^j(T)}{\Sigma_j N_i^j(T)} \left[\frac{D_i^j(T)}{N_i^j(T)} - p_i(T) \right]^2$$

The standard deviation of non-investment-grade default rates was obtained similarly. Figures 1 and 2 show the average default rates and their standard deviation for investment-grade and non-investment-grade companies, respectively.

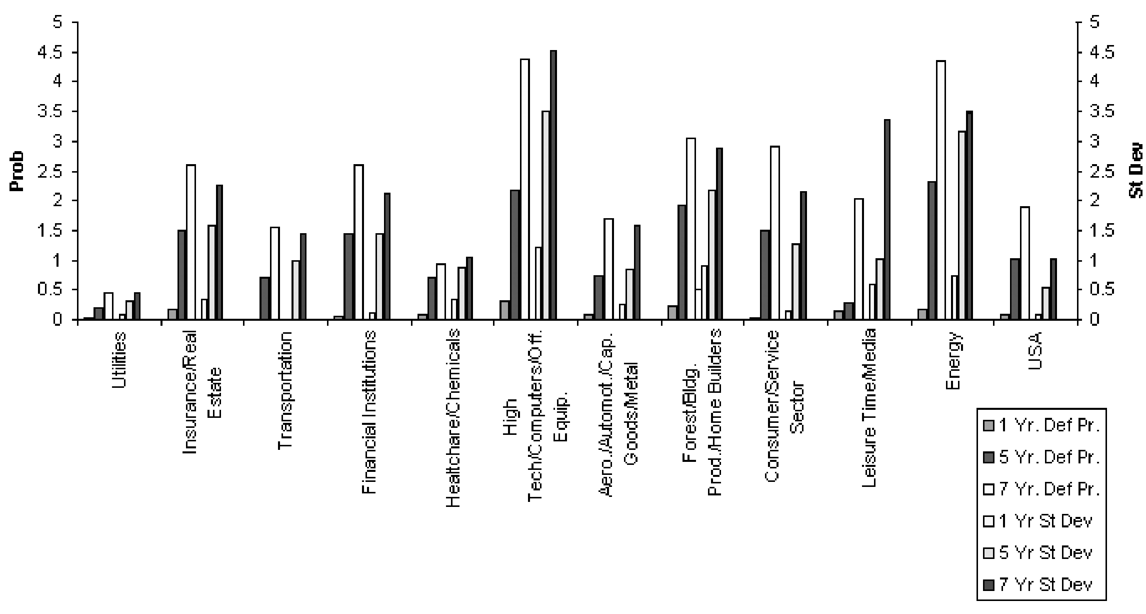

Figure 1: Investment-grade default probabilities and standard deviation

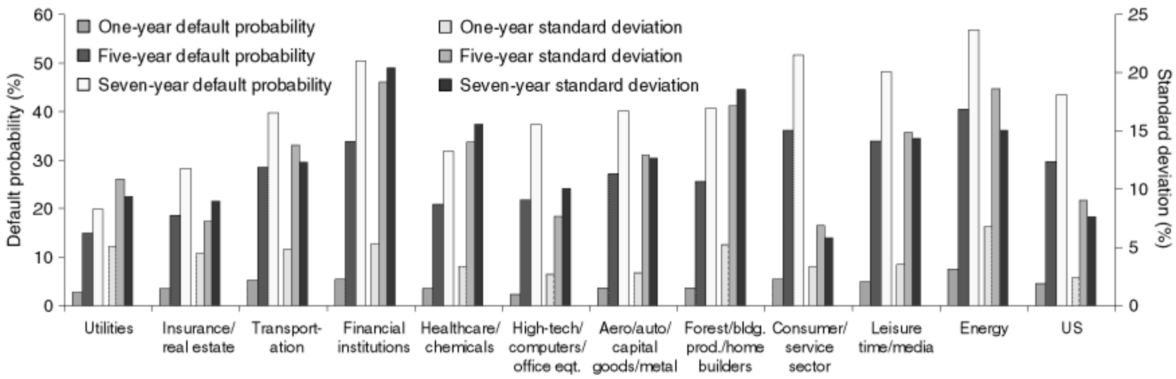

Figure 2: Non-investment grade default probabilities and standard deviation

If company i has a default probability of p, and x_i is a binomial random variable that is one if the company defaults and is zero otherwise, the standard deviation of x_i is:

$$\sigma(x_i) = \sqrt{p(1-p)}$$

If default correlation between two companies i and j is defined as the correlation between the binomial random variables x_i and x_j, which indicate the occurrence of default of companies i and j, respectively, the correlation can be obtained from the covariance of $x_i x_j$. Since $E\{x_i x_j\}$ is the joint default probability, the default correlations $c_{ii}(T)$, $c_{nn}(T)$ and $c_{in}(T)$, which denote correlations over T years for two investment-grade, two non-investment-grade and a pair of investment-grade and non-investment-grade companies, can be obtained from the following:

$$q_{ii}(T) = p_i(T)^2 + c_{ii}(T)p_i(T)(1-p_i(T)) \quad (1)$$

$$q_{nn}(T) = p_n(T)^2 + c_{nn}(T)p_n(T)(1-p_n(T))$$

$$q_{in}(T) = p_i(T)p_n(T) + c_{in}(T)\sqrt{p_i(T)(1-p_i(T))p_n(T)(1-p_n(T))}$$

Table 1 provides the default correlation obtained using Equation 1 over three horizons for different sectors and for all companies in the United States.

The default correlation for all companies in the United States can be used as a proxy for correlation between two companies that are not in the same industry, because this primarily captures the effect of a common macroeconomic environment.

The following summarizes some of the conclusions that can be made from Table 1:

- Default correlations for investment-grade companies are lower than those of non-investment-grade companies. This is consistent with previous studies by Erturk (2000), which show that over a short horizon, default correlations between investment-grade companies are generally quite low.
- Default correlation between non-investment-grade companies first increases with time, and then plateaus out or gradually declines. This has previously been observed by Lucas (1995). The reason suggested by Lucas is that, over a short horizon, defaults arise due to company-specific reasons and, thus, correlations are low. Over the medium term, economic conditions and industry-specific reasons have a significant impact. Over longer horizons, average conditions prevail since one sees both peaks and valleys of economic cycles and, thus, correlations are low. The correlations for non-investment-grade companies are highest at around five years, which also represents the approximate average duration of economic cycles. The default correlation between investment-grade companies, on the other hand, usually continues to increase all the way up to seven years.

		One year		Five years		Seven years	
		Inv grade	NIG grade	Inv grade	NIG grade	Inv grade	NIG grade
United States all	**Inv grade**	0.02	0.18	0.11	0.53	0.29	–0.08
	Non-inv grade	0.18	1.08	0.53	5.72	–0.08	5.50
Utilities	**Inv grade**	–0.02	0.17	0.05	1.21	–0.04	0.68
	Non-inv grade	0.17	1.34	1.21	–0.75	0.68	–5.35
Insurance/real estate	**Inv grade**	–0.05	0.37	–0.77	–0.02	–1.34	–0.96
	Non-inv grade	0.37	0.30	–0.02	–3.99	–0.96	–6.10
Transportation	**Inv grade**	0.00	0.00	–0.12	–0.61	–0.35	–2.27
	Non-inv grade	0.00	1.74	–0.61	4.78	–2.27	–0.24
Financial institutions	**Inv grade**	0.06	0.61	0.42	2.93	0.25	0.00
	Non-inv grade	0.61	6.17	2.93	31.76	0.00	39.35
Health care/ chemicals	**Inv grade**	–0.09	–0.19	–0.72	–1.84	–0.94	–1.99
	Non-inv grade	–0.19	2.17	–1.84	2.88	–1.99	8.33
Hi-tech/computers/ office eqt.	**Inv grade**	2.48	–0.85	2.03	–1.37	0.60	–0.61
	Non-inv grade	–0.85	1.12	–1.37	–2.96	–0.61	–4.90
Aero./auto./capital goods	**Inv grade**	–0.09	0.24	0.14	–0.13	0.62	–0.46
	Non-inv grade	0.24	0.73	–0.13	2.25	–0.46	–0.74
Forest/building products/home builders	**Inv grade**	2.02	–0.93	0.51	–0.02	1.06	4.98
	Non-inv grade	–0.93	6.49	–0.02	24.80	4.98	23.30
Consumer/service sector	**Inv grade**	–0.03	0.35	0.18	1.12	1.00	–3.54
	Non-inv grade	0.35	1.64	1.12	6.77	–3.54	2.97
Leisure time/media	**Inv grade**	–0.14	1.29	–0.27	2.97	0.07	–5.20
	Non-inv grade	1.29	2.16	2.97	19.96	–5.20	26.11
Energy	**Inv grade**	1.49	2.83	0.67	3.56	0.06	9.57
	Non-inv grade	2.83	6.40	3.56	19.08	9.57	18.26

Table 1: Default correlations in different sectors of the United States (correlations x 100)

- Sectors with high default rates (compared with the US average), such as financial institutions, forest/building products and energy also exhibit high default correlations.
- The causes of volatility in default rates in highly correlated sectors range from macroeconomic (recession) to industry-specific events. The default rates in the forest/building products sector are high during recessionary periods and, overall, show a strong correlation to the macroeconomic environment. The defaults in the energy sector, on the other hand, are primarily driven by industry-specific

factors—most notably, the price of oil—and are high during periods of low oil prices. The default rates in financial institutions show some dependence on the prevailing economic environment, but are also high in periods of industry-specific shocks, such as the savings and loans crisis in the United States in the 1980s.

Table 2 provides a qualitative summary of the observed numbers. The conclusions are based on comparisons of industry-specific numbers with those of the US average.

Sector	Default rates		Default corrs.	
	Inv grade	Non-inv grade	Inv grade	Non-inv grade
Utilities	L	L	L	L
Insurance/real estate	H	M	L	L
Transportation	L	M	L	M
Financial institutions	M	H	M	H
Health care/ chemicals	L	M	L	M
High-tech/ computers/office eqt.	H	M	H	L
Aerospace/ automotive/capital goods/metal	M	M	M	L
Forest/building products/home builders	H	H	H	H
Consumers/service sector	H	H	M	M
Leisure time/media	M	M	L	H
Energy	H	H	H	H
L = Low, M = Medium, H = High				

Table 2: Comparison of default rates and correlations to US averages

Parametrization for default correlation

A realistic model for correlation must incorporate default correlation within a region arising from common economic factors as well as the correlation within an industry due to industry-specific reasons. Moreover, it should also reflect the above observed characteristics that the default correlation is higher for weaker credits. Instead of defining default correlations between all possible combinations of rating categories, industries, regions and tenors, consider the following two-parameter model for default correlation: If X and Y are two non-defaulted assets, then:

$$Pr\{Default\ of\ X\ before\ T|Y\ defaults\ before\ T\} = (1+REG_{XY}(T)+IND_{XY}(T))\times Pr\{Default\ of\ X\ before\ time\ T\} \quad (2)$$

where:

$$REG_{XY}(T)\equiv\begin{cases} reg_j(T) & if\ X\ and\ Y\ are\ in\ the\ same\ region \\ 0 & otherwise \end{cases}$$

$$IND_{XY}(T)\equiv\begin{cases} ind_k(T) & if\ X and\ Y\ are\ in\ the\ same\ industry \\ 0 & otherwise \end{cases}$$

The above model describes a relationship between conditional and unconditional default probabilities. It can be viewed as describing the ratio of actual pair-wise default probability to the same under the assumption of independence of default events:

$$\frac{Pr\{Joint\ default\ of\ X\ and\ Y\ before\ T\ \}}{Pr\{X\ defaults\ before\ T\}\times Pr\{Y\ defaults\ before\ T\}} = (1+REG_{XY}(T)+IND_{XY}(T))$$

The factors REG_{XY} and IND_{XY} capture the default correlation contributions that arise from being in the same region and the same industry respectively. If REG_{XY} (respectively, IND_{XY}) is positive, it implies that the default correlation between two companies in the same region (respectively, industry/sector) is positive and vice versa. Moreover, the above model suggests that being in the same region as well as the same industry increases the default correlation if both of the factors are positive. For example, consider an asset X in the United States (which is defined as a region) that is in the insurance sector (which is an industry sector) where assets X and Y both have a tenor of three years. Let $reg_{US}(3) = 0.35$ and $ind_{ins}(3) = 0.5$ and the default probability of asset X in three years be 10%. Then the above model implies that if:

- Asset Y is neither in the United States nor in the insurance sector, then:
  $Pr\{X \text{ defaults in 3 years} | Y \text{ defaults in 3 years}\} = 10\%$
- Asset Y is in the United States but not in the insurance sector, then:
  $Pr\{X \text{ defaults in 3 years} | Y \text{ defaults in 3 years}\} = (1+0.35)\times 10\% = 13.5\%$
- Asset Y is not in the United States but in the insurance sector, then:
  $Pr\{X \text{ defaults in 3 years} | Y \text{ defaults in 3 years}\} = (1+0.5)\times 10\% = 15\%$
- Asset Y is both in the United States and in the insurance sector, then:
  $Pr\{X \text{ defaults in 3 years} | Y \text{ defaults in 3 years}\} = (1+0.35+0.5)\times 10\% = 18.5\%$

Note that it is assumed that REG_{XY} and IND_{XY} depend only on the time horizon T and not on the ratings (credit quality) of X and Y. Even though they do not depend on the credit quality of X and Y, the proposed model is consistent with the observation that the default correlations are higher for weaker credits and almost zero for very strong credits. To see this, let p_X and p_Y be the default probabilities, $corr_{XY}$ be the default correlation and q_{XY} be the joint default probability of assets X and Y. Then:

$$corr_{XY} = \frac{q_{XY} - p_X p_Y}{\sqrt{p_X p_Y (1-p_X)(1-p_Y)}} = \frac{(REG_{XY} + IND_{XY})\sqrt{p_X p_Y}}{\sqrt{(1-p_X)(1-p_Y)}} \quad (3)$$

Thus, the proposed model is consistent with the observation that correlation is greater between weak credits (when p_X and p_Y are large) and almost zero between strong credits (when p_X and p_Y are small).

The proposed model for default correlation was calibrated to the default probabilities and default correlations described in Figures 1 and 2 and Table 1. For any given industry sector, the optimal ratio of conditional and unconditional default probability was obtained to minimize the error in predicting pair-wise default probabilities. The optimal ratios (which in the notation of Equation 2 for industry j would be $(1 + REG_{US} + IND_j)$) are described in Table 3. In most cases, the relative error in predicting the pair-wise default probability matrix is within 2%.[2] The proposed model also shows a good fit to the US data when the rated universe is divided into seven categories (instead of the two considered here).

Sector	1 year	5 years	7 years
Utilities	1.47	0.96	0.79
Insurance/real estate	1.08	0.83	0.85
Transportation	1.31	1.12	1.00
Financial institutions	2.08	1.62	1.38
Health care/chemicals	1.59	1.11	1.18
High-tech/computers/ office eqt.	1.42	0.89	0.92
Aero/auto/capital goods/metal	1.20	1.06	0.99
Forest/building products/home builders	2.71	1.71	1.34
Consumer/service sector	1.28	1.12	1.03
Leisure time/media	1.41	1.39	1.28
Energy	1.78	1.28	1.14
United States	1.23	1.14	1.07

Table 3: Ratio of conditional and unconditional default probabilities in different sectors

It is interesting to note that the ratio is a non-increasing function of horizon T even when correlation increases as a function of T. This is because in the proposed framework described in Equation 3, the default correlation depends also on the default probability, which increases as the horizon becomes longer.

In the second part of the paper (see the Appendix to this section), we develop an analytical approach to obtaining the portfolio loss distribution where the correlation structure is as described above.

Summary

Default events exhibit correlation due to economic factors and/or industry-specific factors. This paper describes the average default correlations observed in various sectors of the United States from 1981 to 1999. A simple and compact parametrization for default correlation is also proposed that captures the correlation both regionally as well as within an industry.

References

Erturk, E., 2000, "Default correlation among investment grade borrowers," *Journal of Fixed Income*, 9(4): 55–60.

Gupton, G., C. Finger and M. Bhatia, 1997, *CreditMetrics Technical Document*, New York, NY: Morgan Guaranty Trust Company.

Li, D., 2000, "On default correlation, a copula function approach," *Journal of Fixed Income*, 9(4): 43–54.

Lucas, D., 1995, "Default correlation and credit analysis," *Journal of Fixed Income*, 9(4): 76–87,.

Wilson, T., 1997a, "Portfolio credit risk I," *Risk*, 10(9): 111–117, September.

Wilson, T., 1997b, "Portfolio credit risk II," *Risk*, 10(10): 56–61, October.

Endnotes

1. Many of the transitions to the unrated category are due to the debt being retired. Excluding these transitions may result in higher default rates as well as correlations.

2. Relative error here is defined as the ratio of the norms of the error matrix and the pair-wise default matrix. For a matrix, the norm used is the square root of the sum of squares of all its elements.