Applying Portfolio Credit Risk Models to Retail Portfolios

Nisso Bucay and Dan Rosen

We present a simulation-based model to estimate the credit loss distribution of retail loan portfolios and apply the model to a sample credit card portfolio of a North American financial institution. Within the portfolio model, we test three default models that describe the joint behavior of default events. The first model is purely descriptive in nature while the other two models are causal models of portfolio credit risk, where the influence of the economic cycle is captured through the correlations of default rates to various macroeconomic factors. The results obtained using all three default models are very similar when they are calibrated to the same historical data. In addition to measuring expected and unexpected losses, we demonstrate how the model also allows risk to be decomposed into its various sources, provides an understanding of concentrations and can be used to test how various economic factors affect portfolio risk.

In recent years, several methodologies for measuring portfolio credit risk have been introduced that demonstrate the benefits of using internal models to measure credit risk in the banking book. These models measure economic credit capital and are specifically designed to capture portfolio effects and account for obligor default correlations.

Several portfolio credit risk models developed in the industry have been made public; e.g., CreditMetrics (Gupton et al. 1997), CreditRisk^{+} (Credit Suisse Financial Products 1997) and Credit Portfolio View (Wilson 1997a and 1997b). Others remain proprietary, such as KMV's Portfolio Manager (Kealhofer 1996). Although the models appear quite different on the surface, recent theoretical work has shown an underlying mathematical equivalence among them (Gordy 2000; Koyluoglu and Hickman 1998). However, the models differ in their distributional assumptions, restrictions, calibration and solution. Also, empirical work shows that all models yield similar results if the input data is consistent (Crouhy and Mark 1998; Gordy 2000).

A limitation these credit risk models share is the assumption that, during the period of analysis, market risk factors, such as interest rates, are constant. While this assumption is not a major obstacle when measuring credit risk for portfolios of loans or floating rate instruments, it is not acceptable when a portfolio contains derivatives or instruments with embedded optionality. An example of an integrated market and credit risk model that overcomes this limitation is given in Iscoe et al. (1999). The authors extend the framework outlined by Gordy (2000) and Koyluoglu and Hickman (1998) by generating scenarios that include explicit market risk factors and credit drivers and allowing for stochastic exposures in each scenario.

The general principles of portfolio credit risk models are equally applicable for both the commercial and the retail markets. However, most of the applications of these models in the literature have focussed on portfolios of bonds or

First published in the *Algo Research Quarterly*, Vol. 3, No. 1, March 2000, pp. 45–73. Also published in *The Journal of Risk Finance*, Vol. 2, No. 3, Spring 2001, pp. 35–61.

corporate loans (e.g., Carey 1998; Crouhy and Mark 1998; Bucay and Rosen 1999; Wilson 1997). The measurement of portfolio credit risk in retail loan portfolios has not received as much attention.

In this paper, we develop a methodology to measure the credit risk of a retail portfolio. The method is based on the general portfolio credit risk framework described in Iscoe et al. (1999). We discuss the practical estimation and implementation of the model and demonstrate its applicability with a case study based on the credit card portfolio of a North American financial institution. Finally, we analyze the sensitivity of the results to various assumptions.

An important part of the framework is the model that describes the joint behaviour of default events. We present and test three models to describe this joint default behaviour and calibrate them using the same historical data. The first model is a *sector-based model*, which is purely descriptive in nature and makes no attempt to explain economic causality of credit distress. The other two models are *factor-based models* of portfolio credit risk. Factor-based models are *causal models*, in which the influence of the economic cycle is captured through the correlations of default rates to various macroeconomic factors. Both causal models use a multi-factor model that captures the systemic component of credit risk due to a set of macroeconomic factors. Therefore, the factor-based models are useful for further stress testing and estimating conditional losses using economic scenarios. These two models differ in the mathematical function they use to relate the factors to the default probabilities.

The rest of the paper is organized as follows. The next section briefly reviews the general quantitative framework for portfolio credit risk models. Thereafter, the credit risk models used in the case study are described, as well as the methodology for their estimation. Then the case study is presented in the following manner: first, the portfolio and the data are described together with the assumptions made to measure the different inputs of the model; second, the estimation of the different parameters of the model are discussed; third, the results are presented and the models are compared; fourth, several stress tests are presented. Finally, some concluding remarks and directions for future research are discussed.

Portfolio credit risk modelling framework

Portfolio credit risk models can be understood within a general underlying framework (see Gordy (2000); Koyluoglu and Hickman (1998); Iscoe et al. (1999)). In this section, we introduce the basic components of the framework. Subsequently, we present a model to assess the credit risk of a credit card portfolio.

We focus on **default-mode portfolio credit risk models**, i.e., on models that measure exclusively the losses due to default events. The framework also applies, more generally, to **mark-to-market** models, where losses due to credit migration are also considered.

Portfolio credit risk models consist of five parts:

Part 1: **Description of the scenarios or states of the world**. This is a model of the evolution of the relevant “systemic” or sector-specific credit drivers that drive credit events, as well as those market factors driving obligor exposures, over the period of analysis.

Part 2: **Correlated default model**. Default probabilities vary as a result of changing economic conditions. At each point in time, an obligor’s default probabilities are conditioned on the state of the world. Default correlations among obligors are determined by how changes in credit drivers affect conditional default probabilities.

Part 3: **Obligor exposures, recoveries and losses in a scenario**. The **credit exposure** to an obligor is the amount the institution stands to lose should the obligor default. Recovery rates are generally expressed as the percentage of the exposure that is recovered through such processes as bankruptcy proceedings, the sale of assets or direct sale to default markets. Exposures can be assumed to be constant in all scenarios for banking instruments without optionality as well as bonds, but not for derivatives or banking book products with credit-related optionality such as prepayment options.

Part 4: **Conditional portfolio loss distribution in a scenario**. Conditional on a scenario, obligor defaults are independent. Based on this property, we can apply various techniques to obtain the conditional portfolio loss distribution (see, for example, Credit Suisse (1997); Finger (1999); Nagpal and Bahar (1999)).

Part 5: **Aggregation of losses in all scenarios**. The unconditional distribution of portfolio credit losses is obtained by averaging the conditional loss distributions over all scenarios.

Single-step portfolio credit risk model for a retail portfolio

We present a single-step, default-mode portfolio credit risk model. The model estimates the distribution of potential losses due to obligor defaults occurring during a single horizon. We assume that exposures and recovery rates at the end of the horizon are deterministic and do not vary with the state of the economy. This is a simplifying assumption that could be relaxed in future work.

Consider the single period $[t_0, t]$; specifically, assume $t = 1$ year. The portfolio contains N obligors or accounts; each obligor belongs to one of $N^S \leq N$ sectors. A sector is a group of obligors of similar characteristics and credit quality. Thus, it is assumed that obligors in a sector are statistically identical; i.e., they have the same probability of default, recovery and exposure in each scenario.

Three variants of the model are presented: a sector-based logit model, a factor-based logit model and a factor-based Merton model.

The scenarios in factor-based models are described by both systemic and sector-specific factors, while the scenarios in sector-based models are described by only sector-specific factors. **Sector-based models** are purely descriptive and make no attempt to explain the economic causality of credit distress. On the other hand, **factor-based models** are causal models of portfolio credit risk. In factor-based models, the influence of the economic cycle is captured through the use of multi-factor models, which capture the correlations of defaults to various systemic factors. Hence, factor-based credit risk models are useful for further stress testing and estimating portfolio losses conditional on economic forecast scenarios.

The factor-based logit model and Merton model differ in the mathematical function used to describe conditional default probabilities for each sector. While the logit model uses a functional form purely for mathematical convenience, the Merton model uses a functional relationship derived from financial principles and a microeconomic view of credit.

In the following sections, the specific parts of the models are described.

1. Scenarios or states of the world

A **scenario** or state of the world at t is defined by the outcome of q systemic and sector-specific factors that influence the creditworthiness of the obligors in the portfolio. We refer to these factors as the **credit drivers.** Of the q credit drivers, there are q^M systemic drivers that represent macroeconomic, country and industry factors; the remainder $q^S = q - q^M$ drivers are sector-specific factors.

Denote by $\boldsymbol{x}$ the vector of factor returns at time t; i.e., $\boldsymbol{x}$ has components $x_k = ln\{r_k(t)/r_k(t_0)\}$, where $r_k(t)$ is the value of the k-th factor at time t. At the horizon, assume that the returns are normally distributed: $\boldsymbol{x} \sim N(\mu, \mathbf{Q})$, where μ is a vector of mean returns, and $\mathbf{Q}$ is a covariance matrix. Denote by $\mathbf{Z}$ the vector of standardized factor returns with entries $Z = (x_k - \mu_k) / \sigma_k$. To distinguish between systemic macroeconomic factors and sector-specific factors, we write the standardized factor returns as the row vector $\mathbf{Z} = (\mathbf{Z}^M, \mathbf{Z}^S)$, where $\mathbf{Z}^M = (Z_1^M, \ldots, Z_{q^M}^M)$ represents the macroeconomic factors and $\mathbf{Z}^S = (Z_1^S, \ldots, Z_{q^S}^S)$ the sector-specific factors.

Factor-based models attempt to explain partially the economic causality of credit losses. Therefore, scenarios include the realization of both the macroeconomic and sector-specific factors; i.e., they are defined on the whole vector $\mathbf{Z}$. In this case, it is common to assume also that the sector-specific factor returns are uncorrelated.

The sector-based model, on the other hand, is descriptive only and assumes that the states of the world are described by levels of sector-specific factors. Therefore, scenarios are represented only by realizations of the vector of sector-specific drivers, $\mathbf{Z}^S$. In this case, these factors are assumed to be correlated.

For mathematical convenience, it is common practice before the analysis to transform the factor returns to a vector of independent factors. This can be achieved, for example, by applying principal component analysis (PCA) to the original macroeconomic factor returns and the sector-specific factors as required. Hence, for ease of exposition, and without loss of generality, we assume that the standardized factor returns, $\mathbf{Z}$, are independent.

2. Joint default model

For each obligor, the joint default model consists of three components. The first is the definition of the unconditional probability of default. The second is the definition of a creditworthiness index and the estimation of the multi-factor model that links the index to the credit drivers. The third component is a model of obligor default that links the creditworthiness index to the default probability; the default model is used to obtain conditional default probabilities in each scenario. Each of these components is explained below.

Denote by τ the time of default of an obligor in sector j, and by p_j its **unconditional probability of default** by time t:

$$p_j = P_r\{\tau \le t\}$$

It is assumed that unconditional probabilities for each sector are known. The method used to obtain these probabilities from historical default experience is described in Appendix 1.

The second component of the joint default model is the creditworthiness index. Consider a given obligor l in sector j. The **obligor creditworthiness index,** denoted by Y_l, is a continuous variable that determines an obligor's creditworthiness or financial health. The likelihood of the obligor being in default at time t can be determined directly by the value of its index. In general, Y_l is a standard normal variable (i.e., with zero mean and unit variance).

The creditworthiness index, Y_l, is related to the scenario, $\mathbf{Z}$, through a linear multi-factor model:

$$Y_l = \sum_{k=1}^{q} \beta_{lk} Z_k + \sigma_l \varepsilon_l \tag{1}$$

where

$$\sigma_l = \sqrt{1 - \sum_{k=1}^{q} \beta_{lk}^2}$$

β_{lk} is the sensitivity of the index l to factor k and the ε_l for each obligor index are independent and identically distributed standard normal variables (independent of Z_k) representing obligor-specific, or idiosyncratic, components.

All obligors in a sector share the first term in the right side of Equation 1. Also, all obligors in a sector, j, have a common σ_l, denoted by σ_j, where the subindex j denotes the sector to which obligor l belongs. However, each obligor has its own specific, uncorrelated component, ε_l. Thus, obligors in a given sector share the systemic component and have idiosyncratic components of similar magnitude. Hence, the index for any obligor in sector j is

$$Y_l = \sum_{k=1}^{q} \beta_{jk} Z_k + \sigma_j \varepsilon_l$$

An obligor creditworthiness index consists of three components: a systemic component driven by the macroeconomic factors, $\mathbf{Z}^M$; a sector-specific component driven by the sector-specific factors, $\mathbf{Z}^S$; and an obligor-idiosyncratic component. Thus, Equation 1 can be rewritten as

$$Y_l = \sum_{k=1}^{q^M} \beta^M_{jk} Z^M_k + \sum_{k=1}^{q^S} \beta^S_{jk} Z^S_k + \sigma_j \varepsilon_l \tag{2}$$

where

$$\sigma_j = \sqrt{1 - \left(\sum_{k=1}^{q^M} (\beta^M_{jk})^2 + \sum_{k=1}^{q^S} (\beta^S_{jk})^2 \right)}$$

It is useful to write the index as

$$Y_l = \beta^{S*}_j Y^S_j + \sigma_j \varepsilon_l \tag{3}$$

where

$$\sigma_j = \sqrt{1 - (\beta^{S*}_j)^2}$$

Y^S_j denotes the **sector creditworthiness index** common to all obligors in sector j and $(\beta^{S*}_j)^2$ represents the percent of variance of the obligor index explained by the sector index. (The sector creditworthiness index is standard normal.) The logit model requires only sector creditworthiness indices, while the Merton model requires explicitly the obligor creditworthiness indices.

The sector creditworthiness index in a factor-based model contains both the systemic macroeconomic component and a sector-specific component. Furthermore, the only sector-specific factor that contributes to the financial health of an obligor in sector j is Z^S_j. Then, from Equation 2 and Equation 3, the sector creditworthiness index for a factor-based model becomes

$$Y^S_j = \sum_{k=1}^{q^M} \beta^{M*}_{jk} Z^M_k + \beta^{S*}_{jj} Z^S_j$$

where

$$\beta^{M*}_{jk} = \beta^M_{jk} / \beta^{S*}_j, \; k = 1, \ldots q^M$$

and $\beta^{S*}_{jj} = \beta^S_{jj} / \beta^{S*}_j$.

On the other hand, in a sector-based model, the sector creditworthiness index does not include a macroeconomic component. Since, in this case, the sector-specific factors are correlated, the creditworthiness index for a sector-based model can be expressed as

$$Y^S_j = \sum_{k=1}^{q^S} \beta^{S*}_{jk} Z^S_k$$

The functional forms for the index used in each model are summarized in Table 1.

Model	Index	Default model
Sector-based logit	sector	$Y^S_j = \sum_{k=1}^{q^S} \beta^{S*}_{jk} Z^S_k$
Factor-based logit	sector	$Y^S_j = \sum_{k=1}^{q^M} \beta^{M*}_{jk} Z^M_k + \beta^{S*}_{jj} Z^S_j$
Factor-based Merton	obligor	$Y^S_j = \sum_{k=1}^{q^M} \beta^{M*}_{jk} Z^M_k + \beta^{S*}_{jj} Z^S_j$

Table 1: Creditworthiness indices in each model

The third and final component of the joint default model is a model of obligor default used to obtain conditional default probabilities in each scenario.

The **conditional probability of default** of an account is the probability that an obligor defaults conditional on a scenario. Given the definition of the sectors and the scenarios, all obligors in a given sector share the same conditional default probabilities. Formally, the conditional probability of default of an obligor in sector j, $p_j(\mathbf{Z})$, is the probability that an obligor defaults conditional on the state of the credit drivers, $\mathbf{Z}$:

$$p_j(\mathbf{Z}) = Pr\{\tau \le t | \mathbf{Z}\}$$

In factor-based models, default probabilities are conditioned on both macroeconomic and sector-specific drivers. In the sector-based model, the probability of default of an obligor in sector j is conditioned on the sector factors only:

$$p_j(\mathbf{Z}) = p_j(\mathbf{Z}^S) = Pr\{\tau \le t | \mathbf{Z}^S\}$$

The computation of conditional probabilities requires a model that describes the functional relationship between the creditworthiness index (and, hence, the systemic and sector-specific factors) and obligor default probabilities. The functional relationship is a map of the index to the range [0,1].

We consider two types of models: a logit model, as presented, for example, in Wilson (1997), and a Merton model, as used in CreditMetrics. From an econometric perspective, the latter is usually referred to as a probit model.

In the **logit model,** the probability of default of an obligor in sector j is related to the sector creditworthiness index, Y_j^S, through

$$p_j(Y_j^S) = \frac{1}{1 + a_j exp(b_j Y_j^S)} \tag{4}$$

where a_j and b_j are two strictly positive parameters of the model. Note that Equation 4 can alternatively be written as

$$p_j(y_j^S) = \frac{1}{1 + exp(y_j^S)}$$

where $y_j^S = ln a_j + b_j Y_j^S$. That is, the logit model can also be expressed in terms of the non-standardized version of the creditworthiness index, $y_j^S \sim N(ln(a_j), b_j^2)$. The variables y_j^S are referred to as the **logit variables**.

Based on Equation 4 and the definition of the sector indices, the conditional default probabilities, $p_j(\mathbf{Z})$, for the sector-based and the factor-based logit models are, respectively,

$$p_j(\mathbf{Z}) = \frac{1}{1 + a_j exp\left(b_j \sum_k \beta_{lk}^{S*} Z_k^S\right)}$$

$$p_j(\mathbf{Z}) = \frac{1}{1 + a_j exp\left(b_j\left(\sum_k \beta_{lk}^{M*} Z_k^M + \beta_{jj}^{S*} Z_j^S\right)\right)}$$

In the **Merton model** (Merton 1974), default occurs when the assets of a firm fall below a given boundary, generally defined by its liabilities. In this situation, an obligor's creditworthiness index, Y_l, can be considered to be the standardized returns of its asset levels. Since obligors in a sector are statistically identical, they share the same default boundary. Thus, default of an obligor l in sector j occurs when Y_l falls below a given sector boundary, α_j. Figure 1 provides a graphical representation of the model.

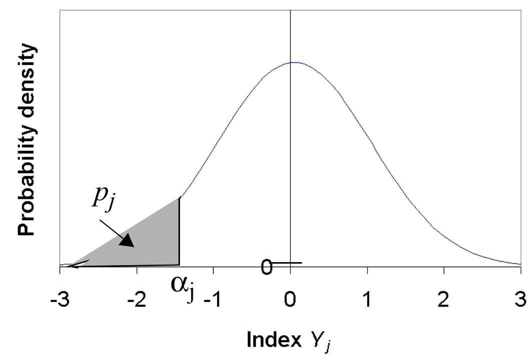

Figure 1: Merton model of default

Since the indices are standard normal variables, the unconditional probability of default of obligor l in sector j can be expressed as

$$p_j = Pr\{Y_l < \alpha_j\} = \Phi(\alpha_j) \tag{5}$$

where Φ denotes the normal cumulative density function and α_j is the unconditional sector threshold. The unconditional threshold is generally calculated by taking the inverse of Equation 5, $\alpha_j = \Phi^{-1}(p_j)$.

From Equation 5, the probability of default, conditional on the index itself, is either zero or one. However, given the factor model defined in Equations 1 to 3, the probability of default of an obligor in sector j, conditional on a scenario, $\mathbf{Z}$, is given by

$$p_j(\mathbf{Z}) = Pr\{Y_l < \alpha_j | \mathbf{Z}\} \tag{6}$$

$$= Pr\left\{ \sum_{k=1}^{q^M} \beta_{jk}^M Z_k^M + \beta_{jj}^S Z_j^S + \sigma_j \varepsilon_l < \alpha_j \middle| \mathbf{Z} \right\}$$

$$= Pr\left\{ \varepsilon_l < \frac{\alpha_j - \left(\sum_{k=1}^{q^M} \beta_{jk}^M Z_k^M + \beta_{jj}^S Z_j^S \right)}{\sigma_j} \right\}$$

$$= \Phi\left(\frac{\alpha_j - \left(\sum_{k=1}^{q^M} \beta_{jk}^M Z_k^M + \beta_{jj}^S Z_j^S \right)}{\sigma_j} \right)$$

$$= \Phi(\tilde{\alpha}_j)$$

The conditional threshold, $\tilde{\alpha}_j = \tilde{\alpha}_j(Z)$, is the threshold that the idiosyncratic component must cross for default to occur in the state of the world, **Z**. Note that the set of q credit drivers includes both macroeconomic and sector-specific credit drivers.

The three default models are summarized in Table 2.

The logit function, Equation 4, and the formula for conditional default probabilities in the Merton model, Equation 6, have similar functional forms. This can be seen in Figure 2, which graphs the conditional default probabilities obtained from both models as a function of the sector creditworthiness index.

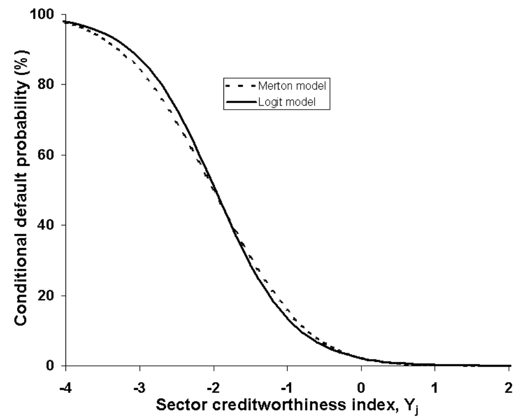

Figure 2: Logit model versus Merton model

Finally, in the case of factor-based models, note that the correlations of creditworthiness indices are uniquely determined by the multi-factor model (which links the indices to the credit driver returns). The correlations of obligor defaults are then obtained from the functional relationship between the indices and the default probabilities, as determined by the Merton or logit model. Similarly, in sector-based models, default correlations are fully defined by the sector index correlations and the default model.

Model	Conditional default probabilities
Sector-based logit model	$p_j(\mathbf{Z}^S) = \dfrac{1}{1 + a_j exp\left\{ b_j \left(\sum_k \beta_{jk}^{S*} Z_k^S \right) \right\}}$
Factor-based logit model	$p_j(\mathbf{Z}) = \dfrac{1}{1 + a_j exp\left\{ b_j \left(\sum_k \beta_{jk}^{M*} Z_k^M + \beta_{jj}^{S*} Z_j^S \right) \right\}}$
Factor-based Merton model	$p_j(\mathbf{Z}) = \Phi\left(\dfrac{\alpha_j - \left(\sum_k \beta_{jk}^M Z_k^M + \beta_{jj}^S Z_j^S \right)}{\sigma_j} \right)$

Table 2: Default models

3. Obligor exposures and recoveries in a scenario

The exposure to an obligor j at the horizon t, V_j, is the amount that can be lost in outstanding transactions with that obligor when default occurs (unadjusted for future recoveries). We assume that the amount that can be lost is deterministic and does not depend on the state of the world: $V_j \neq f(\mathbf{Z})$. Recoveries, in the event of default, are also assumed to be deterministic.

Therefore, the economic loss if an obligor in sector j defaults in any state of the world is

$$L_j(\mathbf{Z}) = \begin{cases} V_j \cdot (1-\gamma_j) & \text{with prob.} \quad p_j(\mathbf{Z}) \\ 0 & \text{with prob.} \quad 1-p_j(\mathbf{Z}) \end{cases}$$

where γ_j is the recovery rate expressed as a fraction of the obligor's exposure. This does not necessarily mean that recovery occurs precisely at default, only that it is expressed as a fraction of the exposed value at default.

4. Conditional loss distribution in a scenario

An important fact used for computation is that, conditional on a scenario, obligor defaults are independent. In the most general case, a Monte Carlo simulation can be applied to determine portfolio conditional losses; however, more effective computational tools exploit the property of obligor independence. For example, if a portfolio contains a very large number of obligors, each with a small marginal contribution, then the Law of Large Numbers (LLN) can be applied to estimate conditional portfolio losses. As the number of obligors approaches infinity, the conditional loss distribution converges to the mean losses over that scenario, and the conditional variance and higher moments become negligible. Hence, the conditional portfolio losses, $L(\mathbf{Z})$, are given by the sum of the expected losses of each obligor:

$$L(\mathbf{Z}) = \sum_j E\{L_j(\mathbf{Z})\} = \sum_j V_j \cdot (1-\gamma_j) \cdot p_j(\mathbf{Z}) \tag{7}$$

The number of accounts in each sector is assumed to be sufficiently large to apply Equation 7. This assumption is made for computational efficiency and can easily be relaxed.

Other efficient methods to compute conditional portfolio losses include the application of the Central Limit Theorem (which assumes that the number of obligors is large, but not necessarily as large as that required for the LLN), the application of moment generating functions with numerical integration, and the application of probability generating functions with a discretization of exposures.

5. Aggregation of losses in all scenarios

Unconditional portfolio losses are obtained by averaging the conditional losses over all states of the world. Mathematically, the loss distribution is given by

$$Pr\{L_p < \theta\} = \int_{\mathbf{Z}} Pr\{L(\mathbf{Z}) < \theta\} dF(\mathbf{Z})$$

where L_P denotes the unconditional portfolio losses, θ denotes the level of losses and $F(\mathbf{Z})$ is the distribution of $\mathbf{Z}$.

The integral is obtained by performing a Monte Carlo simulation. The Monte Carlo simulation process is performed by

- generating a set of joint scenarios on $\mathbf{Z}$. In the factor-based model, scenarios are generated on $\mathbf{Z} = (\mathbf{Z}^M, \mathbf{Z}^S)$, while in the sector-based model scenarios are generated only on $\mathbf{Z}^S$.
- computing, under each scenario
  - the conditional default probabilities for each sector (using either the logit or the Merton model)
  - the conditional portfolio losses (assuming that the LLN applies)
- obtaining the distribution of portfolio losses by averaging the distribution over all scenarios.

The methodology for calibrating each of these models is presented in Appendix 1.

Case study

In this case study we apply the credit risk models described in the previous section to a sample credit card portfolio of a North American financial institution. The analysis period is the first quarter of 1999.

The objective of the study is to compute the portfolio credit loss distribution of outstanding accounts over a one-year horizon and to analyze the various contributions to these losses. Credit losses are defined as those arising exclusively from the event of an obligor's default.

We describe the portfolio and historical data followed by some formal definitions and modelling assumptions. Thereafter, we present the macroeconomic factors and market data. For obvious reasons of confidentiality, the data presented has been normalized. However, this in no way affects the analysis or the conclusions that can be drawn from the results.

Portfolio description

The data consists of account information for credit cards issued between the last quarter of 1995 and the first quarter of 1999. Accounts are grouped in terms of their cohort and risk class. A **cohort** is formed by all the credit cards issued in a particular month.

Cards are scored at acquisition. The score is an internal rating of the creditworthiness of a particular cardholder and assesses a borrower's future repayment performance (see, for example, Mays (1998); Lewis (1992)). Scoring models have been commonly used to measure credit risk in retail portfolios (see, for example, Richeson et al. (1994)).

A **risk class** is formed by accounts with similar scores. Accounts in the portfolio were originally classified into 20 risk classes, 18 of which are scored. The other two classes correspond to two types of unscored accounts. They require special consideration because of their size and specific characteristics. The first class, "unscored_1", contains special accounts with cards issued directly to existing customers of the financial institution. The second class, "unscored_2", contains cards for which a reliable score is not available (e.g., cards that were not scored or cards for which the score was lost).

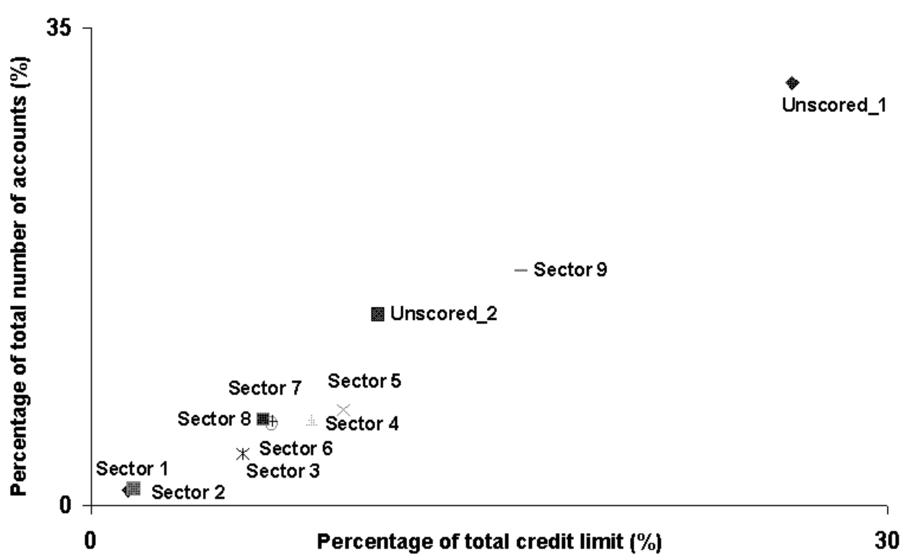

Figure 3: Composition of the portfolio

For modelling purposes, we construct sectors of similar accounts. Each sector must contain a large number of accounts in order to estimate the parameters of the model reliably. For this study, a **sector** is defined as all the accounts that belong to a specific risk class. Given that some risk classes in the low and high risk categories are very thinly populated, the number of sectors was consolidated from 20 to 11. With larger samples, it is possible to group accounts in sectors by adding, for example, geographical or demographic information.

The sample portfolio contains between half a million and a million cards. Figure 3 presents the percentage of the total number of accounts in each sector and the percentage of the total credit limit available to each sector. Sector 1 contains the highest risk accounts (low scores), while sector 9 contains the lowest risk accounts (high scores). Note that the unscored sectors represent a large percentage of the sample portfolio when measured by the percentage of total accounts or total credit limits. As is expected, on a per card basis, the average credit limit generally increases with the score. At any point in time, accounts in the portfolio are classified as performing, in default or closed. This data is used to estimate default rates. Figure 4 presents the cumulative percentage of accounts classified as either performing or in default, by sector.

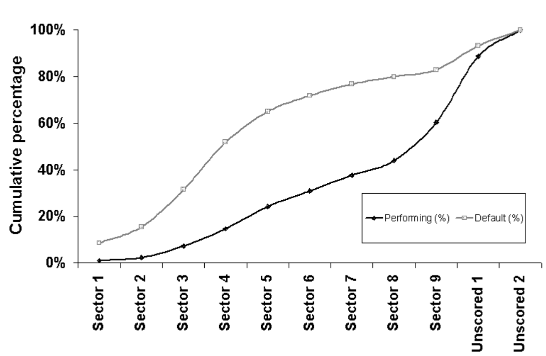

Figure 4: Cumulative percentage of performing and default accounts by sector

Portfolio as of first quarter, 1999

We measure the credit risk of all cards in the portfolio classified as performing at the end of the first quarter of 1999. Figure 5 shows the average balance per card and the utilization rate (given by the ratio of the current balance over the original credit limit). The average balance and the utilization rate are presented as deviations from the average balance per card and the utilization rate for the portfolio as a whole. For example, while the average balance per card in sector 1 is 30% above the average balance per card for the portfolio, the average balance per card in sector 9 is about 45% below.

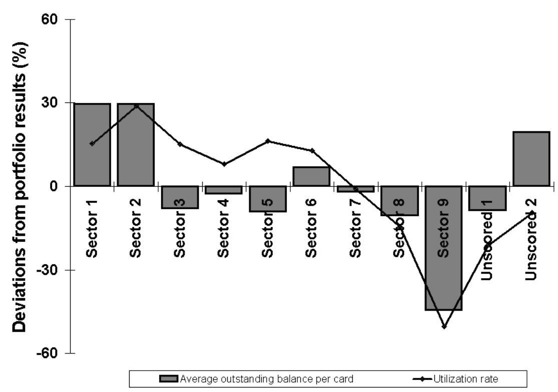

Figure 5: Average balance per card

Not surprisingly, high risk accounts carry a larger average balance since accounts with low scores are expected to be credit takers, while cards in the high score categories are expected to have a larger transactional component.

Modelling assumptions and data

All three models require

- current ratings (or scores) for all accounts
- definition of sectors for analysis
- definition of default events
- default probabilities for each sector
- credit exposures
- recovery rates.

In addition, the factor-based models also require the definition of macroeconomic factors or credit drivers and their joint distribution. Below, we describe the assumptions and data used in the model.

Account scores

The score of each account at the current time is required to classify accounts into sectors and to estimate the likelihood of each account defaulting in the following year. Since the sample data contains only the scores of the credit cards when they were issued, we assume that all outstanding accounts retain a score similar to the one they were originally assigned; hence, they remain in the same sector that they were assigned to at acquisition. More precisely, we assume that overall, the portfolio has the same proportion of accounts in each sector.

Sectors

As noted, 11 sectors are identified. The assumptions of the model are that a sector is homogeneous (i.e., accounts are approximately the same size) and all accounts within a sector are statistically identical.

Default events

A default event occurs when a particular cardholder fails to pay three minimum monthly payments on the credit card balance (and the loan is eventually charged-off by the bank) or when the cardholder declares bankruptcy. Given the method for estimating probabilities of default, we also assume that all accounts classified as performing at the time of the analysis were performing in the previous months. This assumption may require further validation.

Probabilities of default

In order to calibrate the joint default model, we require the one-year probabilities of default for all sectors and a description of how these default probabilities change through time.

For each month in the sample, default probabilities are estimated as the observed default rates over the following year of all new accounts issued for each sector. The one-year default rates for these cohorts are measured by the cumulative number of defaults (charge-offs and bankruptcies) one year after a specific cohort was formed, as a percentage of the number of accounts issued.

There are two reasons to estimate default rates from only the one-year default rates of accounts when they are issued. First, the only data available are the scores assigned to the cards on the issue date; second, using a new set of cards each month results in independent samples that can be used to estimate the distribution of default probabilities, even though the periods are overlapping. The latter is a subtle, yet important, point in the estimation of the joint default model.

The time series contains 28 one-year default rates for each sector. Table 3 summarizes the statistics for these series. The median and standard deviations are presented as a percent of the mean rate in each sector.

Average default probabilities decrease monotonically with the score of the sector. The average rates of sector unscored_1 and sector unscored_2 fall between the averages rates of sector 6 and sector 7.

Although it is difficult to assess accurately the distribution of default rates using only a few observations, the distribution of default rates in general appears to be close to normal. We test for the normality of the distribution of default rates by applying a Chi-square goodness of fit test with a 5% confidence level. This suggests that a direct simulation of joint default probabilities might be a simple alternative to other models.

Table A1 in Appendix 2 provides the sample correlations of default probabilities between sectors. Correlations between default rates in each sector are substantial (they range between –12% and 88%) and, hence, cannot be assumed to be independent.

Credit exposures

Credit exposure is the amount the bank stands to lose in the event that a cardholder defaults, unadjusted for any recovery. Generally, when an account that has been performing accumulates missed payments prior to default, the outstanding balance quickly approaches the current credit limit (which for long-standing accounts may be different from the original credit limit). With timely updates on the evolution of credit limits, the credit exposure to an account that defaults is generally close to the last authorized credit limit.

Sector	1	2	3	4	5	6	7	8	9	u_1	u_2
Median[a]	108	103	107	113	112	102	101	98	96	88	104
Standard deviation[b]	36	43	42	42	37	40	47	41	42	60	40
Excess Kurtosis	–0.55	–0.42	–0.95	–0.88	–0.90	–0.23	4.02	3.16	–0.37	0.82	–0.25
Skewness	–0.54	–0.50	–0.57	–0.54	–0.55	–0.13	1.39	1.31	0.24	1.08	–0.33
Chi-square statistic	6.5	2.5	11.5	5.5	5.0	2.0	2.5	7.5	2.5	10.1	2.0
Critical Chi-square	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5
Test result	Accept	Accept	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Reject	Accept

Table 3: Statistics for default probabilities time series

a. ratio of median / mean rate (%)
b. ratio of standard deviation / mean rate (%)

Sector	1	2	3	4	5	6	7	8	9	u_1	u_2
Mean	2.5	3.1	3.6	3.7	3.9	4.2	4.5	4.8	5.8	4.3	4.4
Median	2.4	2.9	3.4	3.5	3.8	4.1	4.4	4.7	5.8	4.3	4.3
Standard deviation	0.5	0.5	0.6	0.6	0.5	0.5	0.4	0.4	0.5	0.5	0.6
Excess Kurtosis	0.69	0.58	2.31	0.45	0.73	3.0	1.3	1.1	–0.09	–0.43	3.26
Skewness	1.21	1.1	1.5	1.23	1.2	1.53	0.63	0.26	0.72	–0.32	1.7
Chi-square statistic	12.5	1.3	22.0	16.5	15.5	6.0	5.5	5.5	6.5	6.4	10.0
Critical Chi-square	9.5	6.0	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5
Test result	Reject	Accept	Reject	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Reject

Table 4: Statistics of the time series of logit variables

We estimate the credit exposure of an account in default by the average utilization rate at the time of default (the product of the percentage by which the outstanding balance exceeds the original credit limit and the limit). Average exposures in each sector are shown in Figure 6 as deviations from average exposures for the portfolio.

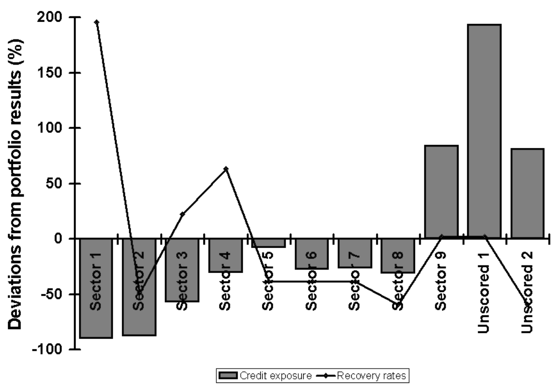

Figure 6: Credit exposure and average recovery rate per sector

Recovery rates

In the case of retail loans, like credit cards, factors such as the lack of collateral, the small size of loans and the expense incurred in court proceedings contribute to low recovery rates, if recovery occurs at all. In this study, recovery rates are deterministic and are based on the average loss rate in each sector. The recovery rates are presented in Figure 6 as deviations from the average recovery rate for the portfolio.

Credit drivers

For the factor-based models, nine macroeconomic variables are considered to be the credit drivers that systemically drive the default probabilities of each sector. These credit drivers are

1. industrial production
2. stock index
3. consumer price index
4. retail sales
5. unemployment level
6. three-month treasury bill at tender
7. short-term government bond yield
8. medium-term government bond yield
9. long-term government bond yield.

Monthly data for the credit drivers from December 1982 to March 1999 is obtained from the Standard & Poor's financial market and economic database (Standard & Poor's 1999).

Calibration of joint default probability models

The estimation techniques used to calibrate the joint default probability models are presented in Appendix 1. In this section, we present the results as applied to the data in this case study.

The data on default probabilities cover the period from the last quarter of 1995 to the first quarter of 1998. Therefore, a time series of 28 monthly overlapping one-year returns on each of the nine macroeconomic factors $(x_k(t_i), i = 1,\ldots,28, k = 1,\ldots,9)$ is used to estimate the parameters of the joint default model.

Sector-based logit model

The statistics for the logit variables in each sector are presented in Table 4. As can be seen in Table 4, the assumption of normality of the logit variables might not be accurate, particularly for sector 3, sector 4 and sector 5. However, additional data would be required to estimate this distribution with more confidence.

The correlation matrix of the logit variables is presented in Table A2 in Appendix 2. As is expected from the correlations of the default probabilities, correlations are substantial between the sector indices, ranging between –10% and 82%.

Factor-based models

The factor-based models are based on factors that are independent standard normal variables. (This assumption is made for computational convenience and does not restrict the analysis.) Independent factors are obtained from the original macroeconomic factors using principal component analysis (PCA).

Sector	1	2	3	4	5	6	7	8	9	u_1	u_2
Variance of logit variable	0.25	0.24	0.38	0.35	0.23	0.278	0.25	0.17	0.23	0.28	0.34
Percentage systemic	65.32	52.20	73.22	65.77	73.40	46.85	38.16	47.47	50.48	52.67	63.84
Percentage sector-specific	34.70	47.80	26.77	34.23	26.58	53.14	61.83	52.55	49.53	47.33	36.17

Table 5: Specific and systemic risk in the factor-based logit model

Sector	1	2	3	4	5	6	7	8	9	u_1	u_2
Variance of inverse normal variables	0.0545	0.0470	0.0606	0.0552	0.0347	0.0388	0.0339	0.0220	0.0238	0.0429	0.0439
Percentage systemic	65.49	52.72	74.85	66.81	74.60	48.99	39.38	47.29	50.62	53.33	63.49
Percentage sector-specific	34.51	47.28	25.15	33.19	25.40	51.01	60.62	52.71	49.38	46.67	36.51
Percentage variance of sector-specific factors, σ_j^2	0.97	0.98	0.97	0.98	0.98	0.98	0.98	0.99	0.99	0.98	0.98

Table 6: Specific and systemic risk in the factor-based Merton model

With PCA, the credit drivers are expressed as linear combinations of uncorrelated standardized random variables called principal components (PC). The use of principal components reduces the number of credit drivers used in the estimation of the model.

Figure 7 presents the percent variance that each of the principal components explains, as well as the cumulative variance. Note that five factors explain more than 95% of the joint movements of all nine macroeconomic factors. Figure 8 plots the weights in each factor for the first five principal components. For example, the first factor, which explains over half of the joint movements of the credit drivers, has positive weights on retail sales, stock index, unemployment level, consumer price index and industrial production, and a negative weight on all interest rate credit drivers.

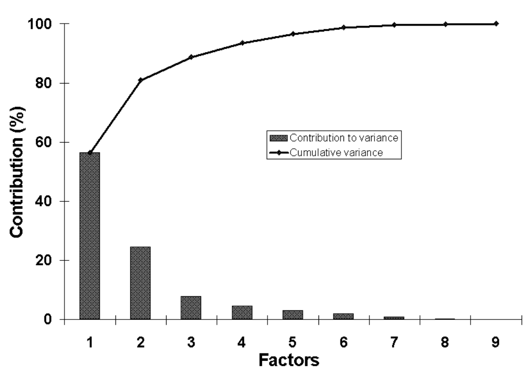

Figure 7: Percent of variance explained by PCA

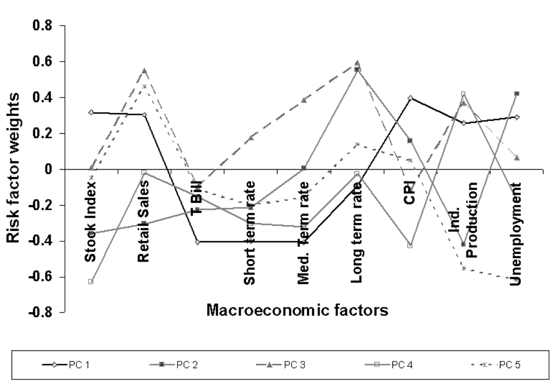

Figure 8: Factor weights for the first five principal components.

For each default model, the coefficients of the multi-factor model are estimated using regression as described in Appendix 1. In this exercise, the regressors are the principal components of the nine macroeconomic credit drivers. (Note that the factor-based models with nine principal components are likely to suffer from over-fitting given the small number of observations. However, they will likely result in more conservative losses since they build higher correlations. For predictive purposes, and further risk decomposition, a model with fewer systemic factors is likely to be better behaved and more robust.)

Logit model

The regression results are presented in Table A3 in Appendix 2. Note that some individual weights from the regression are not statistically significant, in part because of the small number of observations. Based on the regression, estimates for the systemic and idiosyncratic components of the indices are presented in Table 5.

The credit drivers explain between 38% and 73% of the variance of the sector creditworthiness indices. Figure 9 shows the explanatory power of the credit drivers by plotting the systemic component and the historical realizations of the index in the period of estimation for selected sectors. Clearly, the systemic component tracks the main tendencies of the indices. Only four sectors are shown, but the results are similar for the remaining indices.

Factor-based Merton model

The weights (β_{jk}) of the factor-based Merton model are summarized in Table A4 in Appendix 2. The systemic and sector-specific components of the indices are presented in the second and third row of Table 6. These are the relative sizes of the coefficients $\sum_k (\beta^M_{jk})^2$ and $(\beta^S_{jj})^2$ as defined in Appendix 1 (Equation and Equation). The sector-specific component, σ_j^2, is presented in the last row of Table 6.

Figure 9: Systemic component and index over period of estimation for factor-based logit model

In the following sections, we present the results of the analyses performed on the sample portfolio, considering first the portfolio loss distribution calculated according to each of the three models, followed by an analysis of risk contribution and marginal risk by sector and finally, stress testing of some parameters of the model.

Portfolio loss distribution

The portfolio loss distribution is estimated based on 5,000 Monte Carlo scenarios on the relevant credit drivers for each of the models. Scenarios on the logit variables for each sector, y_j^S, are used in the sector-based logit model; scenarios on the nine macroeconomic credit drivers and the 11 sector-specific credit drivers are used for the factor-based models. (Scenarios are generated directly on the standardized independent returns, or first on the macroeconomic factors and then transformed into standardized independent variables. Given the independence of the sector-specific credit drivers, simulation could have been restricted to scenarios on the systemic credit drivers only.)

Sector-based logit model

The portfolio loss distribution is presented in Figure 10. The distribution is presented as deviations from the expected losses. As expected, the distribution is skewed and has a long fat tail on the left due to the nature of credit risk.

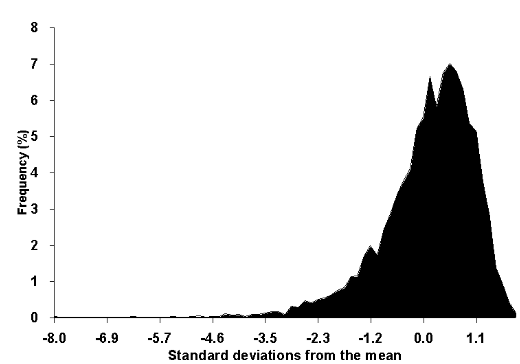

Figure 10: Loss distribution for the sector-based logit model

Table 7 presents the relevant statistics of the loss distribution, including the expected losses, standard deviation, maximum percentile losses, unexpected losses (Credit VaR) and expected shortfall at the 99th and 99.9th percentiles for each of the models. Expected losses have been normalized to 1.0 and all the statistics are scaled to expected losses. Numbers in parenthesis represent the number of standard deviation from the expected losses.

Credit VaR measures the capital required to cover unexpected losses (maximum percentile losses minus expected losses) at the chosen level (99% or 99.9%). In this case, capital is approximately 15% to 78% higher than the reserves, depending on the confidence level chosen. Note that Credit VaR (99%) is approximately three times the standard deviation; if the distribution were normal, Credit VaR would be only twice the standard deviation.

In addition to the most commonly known measures of risk, Table 7 also presents the expected shortfall (tail conditional loss). **Expected shortfall** measures the expected losses beyond a specified percentile of the distribution. By measuring the area under the tail of the distribution, expected shortfall provides a good measure of extreme losses, should they occur. On the other hand, maximum percentile losses are point estimates in the tail of the distribution and may present undesirable properties from a risk management perspective (see Artzner et al. 1998)

Factor-based logit model

The relevant statistics for the loss distribution using the factor-based logit model are presented in Table 7. The results for the factor-based loss distribution are similar to those for the sector-based loss distribution. The loss distribution looks qualitatively similar to that in Figure 10. This implies that when scenarios are defined using explicit macroeconomic and sector-specific credit drivers, the joint behavior of default probabilities for each sector is largely accounted for.

The main differences arise in the extreme tail of the distribution. For example, at the 99.9th percentile, expected shortfall is 5.6% lower in the factor-based model than in the sector-based model. This occurs because the sector-based model presents fewer correlations, since the model can only build correlations through the systemic credit drivers (the sector-specific factors are assumed independent).

	Sector-based logit model	Factor-based logit model	Factor-based Merton model
Expected losses	1.0	1.0	1.0
Standard deviation	0.4	0.3	0.3
Maximum losses (99%)	2.1	2.1	1.9
Credit VaR (99%)	1.1 (3.2)	1.1 (3.2)	0.9 (2.8)
Expected shortfall (99%)	2.4	2.3	2.2
Maximum losses (99.9%)	2.7	2.6	2.5
Credit VaR (99.9%)	1.8 (5.0)	1.6 (4.6)	1.6 (4.8)
Expected shortfall (99.9%)	3.1	3.0	2.8

Table 7: Statistics for one-year loss distribution

Factor-based Merton model

The relevant statistics for the loss distribution using the factor-based Merton model are reported in Table 7. Again, the results are very similar to those of the logit model, particularly to the factor-based logit model. The functional equivalence of the factor-based models is clear since the expected and unexpected losses (at the 99th and 99.9th percentiles) in the logit and Merton model are similar.

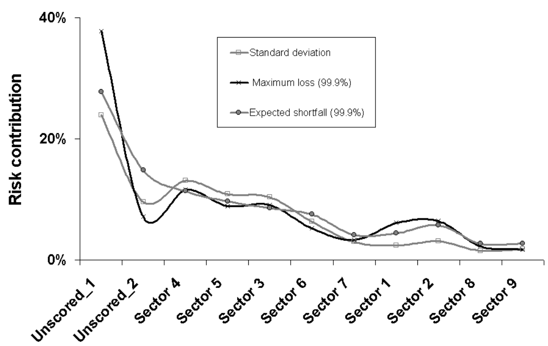

Figure 11: Risk contribution by sector (%)

Risk contributions and marginal risk

Figure 11 presents the risk contributions of each sector. The graphs present the percentage decrease in the specified statistic if the accounts of the corresponding sector are eliminated. The risk contribution of any sector is roughly the product of the size of the sector and the marginal risk of increasing the exposure to a particular sector by one unit. Therefore, it is useful to understand whether the risk contribution of a particular score arises because a large portion of the portfolio falls in that sector, or because the sector has a high marginal risk. Figure 12 plots the marginal risk of every sector (marginal standard deviation as a percentage of mean exposure) against the mean exposure.

From a risk management perspective, it is desirable to have a small exposure to sectors with high marginal risk and a large exposure to sectors with low marginal risk. Sectors with high exposure and high marginal risk are outliers, as is the case, for example, for sector unscored_1. One sector dominates another sector if it has both higher exposure and marginal risk. For example, sector 5 dominates sector 6, and sector unscored_1 dominates sectors 6 through 9 and unscored_2. Dominant portfolios are outliers that may point to opportunities for effective restructuring.

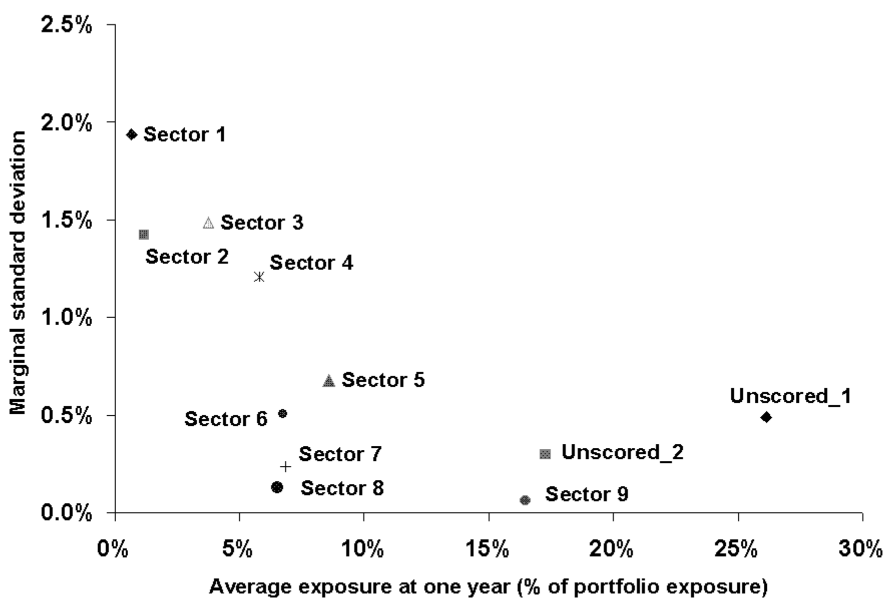

Figure 12: Marginal risk versus average exposure

Note in Figure 12 that, on a marginal basis, sectors 1 through 9 have progressively lower marginal risk, since, on average, accounts in these sectors have decreasing unconditional default probabilities. On a portfolio basis, however, correlations between obligor defaults may play a significant role. Therefore, one sector with a lower default probability than another may have a larger marginal contribution, since it has a higher correlation to the overall portfolio. This is the case, for example, for sector 2 and sector 3.

Stress testing

The results obtained using the factor-based logit model (Table 7) are designated as the **base case**. Comparisons to the base case are used to test the sensitivity of the loss distribution to changes in various parameters of the model. Given the functional equivalence of the factor-based logit and Merton models, and the similarity of the results, the comparison could have been based on the results of either model.

We perform four tests. First, we assess the appropriateness of the Monte Carlo simulation by computing sampling errors for the distribution estimates and comparing the results to a simulation with a larger number of scenarios. Second, we assess the impact of concentration risk by assuming all sectors are independent. Third, we estimate a model of the credit driver evolution using a larger data sample that better captures the impact of the business cycle. Finally, we apply a weaker definition of default that results in higher default rates, and estimate the portfolio losses that result from this larger set of events.

Sampling errors

The statistics in the base case are point estimates based on 5,000 Monte Carlo scenarios on both macroeconomic and sector-specific credit drivers. These estimates can be characterized using probabilistic confidence bounds. Confidence bounds on the mean and standard deviation are estimated using standard methods found in most statistics texts; the bounds on percentiles are estimated using rank statistics (Pritsker 1997).

Table 8 presents the 95% confidence bounds for the expected losses, standard deviation and maximum losses at the 99th and 99.9th percentiles. The statistics are presented relative to expected losses in the base case. The numbers in parenthesis indicate the percentage deviation from the estimate. While the point estimate of the Maximum losses (99%) is twice the level of the expected losses, with 95% confidence, the true losses are within 5% of this ratio. At higher percentiles, the confidence bounds widen. Hence, the certainty of the results diminishes.

Table 8 also summarizes the results of a simulation with 10,000 scenarios. Notice that the difference between the estimates of the two simulations is much smaller than the difference between the confidence bounds with 5,000 scenarios.

	Estimate with 5,000 scenarios			Estimate with 10,000 scenarios
	Lower bound	Estimate	Upper bound	
Expected losses	0.98 (1.5%)	1.00	1.02 (1.5%)	1.00
Standard deviation	0.34 (2.2%)	0.35	0.36 (4.4%)	0.35
Maximum losses (99%)	2.02 (4.3%)	2.11	2.22 (5.6%)	2.08
Maximum losses (99.9%)	2.46 (5.6%)	2.60	3.41 (31.2%)	2.67

Table 8: 95% confidence bounds for the estimates

For example, while the bounds for the Maximum losses (99%) scaled by expected losses are approximately 5% of the estimate, the difference between the two simulations is approximately 1%. In general, the non-parametric bounds on maximum losses are fairly conservative. The accuracy of the Monte Carlo simulation is inversely related to the square root of the number of scenarios. Therefore, using four times as many scenarios reduces the uncertainty in the results by a factor of about two.

Given the greater uncertainty in the estimation of the parameters of the model, these results suggest that increasing the number of scenarios may result in unnecessary additional computation.

Independent defaults

We estimate the loss distribution assuming that defaults across sectors are uncorrelated and determined only by the sector-specific factor.

Figure 13 illustrates the loss distribution assuming that sector defaults are independent. The distribution is graphed against deviations from expected losses in the base case. This loss distribution has a higher mass in the centre and a tail that is not as fat as the distribution of the base. The fact that the base case has a fatter tail can also be concluded by noticing that the standard deviation is smaller if defaults are independent. Therefore, even though extreme losses have the same distance from the mean, economic capital is higher if defaults are correlated.

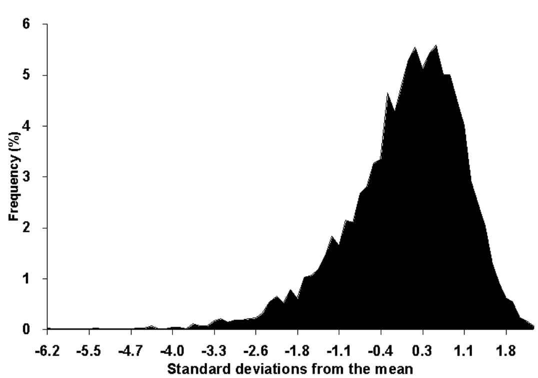

Figure 13: Loss distribution with independent defaults

Table 9 presents the statistics for the case of independent defaults and compares them to those of the base case. The results are presented relative to the expected losses in the base case. The numbers in parenthesis indicate the percentage deviation from the base case.

	Base case	Independent defaults
Expected losses	1.0	1.0
Standard deviation	0.3	0.2
Maximum losses (99%)	2.1	1.6
Credit VaR (99%)	1.1 (3.2)	0.6 (3.0)
Expected shortfall (99%)	2.3	1.8
Maximum losses (99.9%)	2.6	2.0
Credit VaR (99.9%)	1.6 (4.6)	1.0 (4.5)
Expected shortfall (99.9%)	3.0	2.1

Table 9: Statistics for independent defaults

Expected losses are not affected by correlations, therefore, credit reserves are not affected. However, economic capital is very sensitive to correlations. Assuming independent defaults, Credit VaR with independent defaults is about 60% lower, compared to the base case. The ratio of the expected shortfall in the base case to the expected shortfall with independent defaults is between 70% and 75% for the 99th and 99.9th percentiles, respectively.

Losses with correlated credit risk drivers

The parameters of the credit risk models are estimated based on data for defaults and credit drivers that spans a short, 28-month time period (from the last quarter of 1995 to the first quarter of 1998). Thus, the sample probably does not cover a full business cycle. One major advantage of factor-based models in credit risk measurement is that data for the credit drivers is available for longer horizons. Thus, more information about the business cycle can be incorporated using factor-based models.

Recall that the default model with credit drivers consists of two parts: a multi-factor model that links the creditworthiness index of each sector (and hence the default probabilities) to the credit drivers, and a model for the evolution of the credit drivers. The results in the base case are obtained using data from the 28-month period to estimate both parts of the model. We can refine the estimates of portfolio losses by estimating the joint behavior of the credit drivers using data for longer horizons.

In this example, we perform the PCA on the credit drivers using quarterly non-overlapping data from a period that extends from the first quarter of 1983 to the first quarter of 1999. The regression model is re-estimated with the newly transformed credit drivers, which are correlated during the estimation period of the default model. The simulation is performed using these new parameters. Figure 14 presents the resulting portfolio loss distribution. The distribution is graphed against deviations from expected losses in the base case.

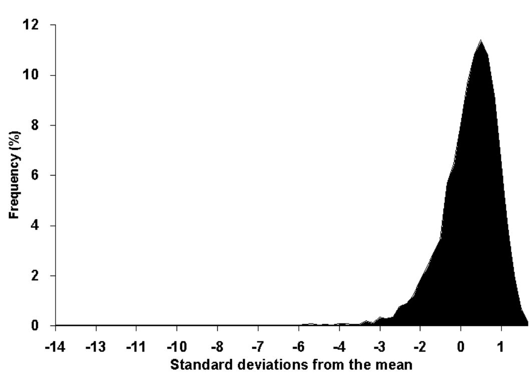

Figure 14: Loss distribution with correlated credit drivers

The statistics for the loss distribution for the case of correlated credit drivers are presented in Table 10. The results presented are relative to the expected losses in the base case. The numbers in parenthesis indicate the percentage deviation from the base case.

	Base case	Correlated credit drivers
Expected losses	1.0	1.1
Standard deviation	0.3	0.3
Maximum losses (99%)	2.1	2.1
Credit VaR (99%)	1.1 (3.2)	1.0 (3.0)
Expected shortfall (99%)	2.3	2.5
Maximum losses (99.9%)	2.6	3.2
Credit VaR (99.9%)	1.6 (4.6)	2.1 (6.7)
Expected shortfall (99.9%)	3.0	4.1

Table 10: Statistics conditioned on correlated credit drivers

The expected losses are higher and the standard deviation is somewhat smaller than in the base case. Given that, the Credit VaR at the 99% level in both cases are very similar. However, the Credit VaR at the 99.9% level is more than 25% higher for correlated credit drivers, due to the longer tail associated with the distribution. In general, the statistics at the 99.9% level are between 25% and 39% higher than those of the base case. One explanation for these results is that the scenarios span a broader range of market and economic changes that better capture the effect the economic cycle has on consumer finance.

Default losses with false-performing accounts

At the end of each month, some accounts are classified as performing, though they are actually in default. The default losses computed in the base case do not incorporate the "potential" losses of these *false-performing* accounts. Thus, it is useful to estimate the impact on the relevant default statistics of classifying these accounts as true defaults. To accomplish this, the parameters of the models must be estimated with new regressions.

The histogram of the loss distribution and the statistics of the distribution are shown in Figure 15 and Table 11, respectively. In both cases, the results are presented relative to the expected losses in the base case. The numbers in parenthesis indicate the percentage deviation from the base case.

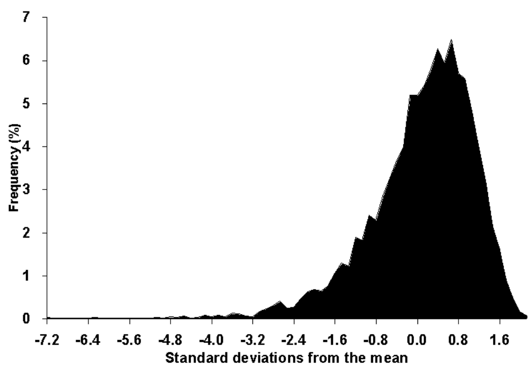

Figure 15: Loss distribution with false-performing accounts

	Base case	False-performing accounts
Expected losses	1.0	1.5
Standard deviation	0.3	0.5
Maximum losses (99%)	2.1	3.1
Credit VaR (99%)	1.1 (3.2)	1.6 (3.1)
Expected shortfall (99%)	2.3	3.5
Maximum losses (99.9%)	2.6	4.0
Credit VaR (99.9%)	1.6 (4.6)	2.5 (4.8)
Expected shortfall (99.9%)	3.0	4.5

Table 11: Statistics including false-performing accounts

Note that while changing correlations affects only dispersion statistics, changing default frequencies have a severe impact on the expected losses as well. Incorporating false-performing accounts increases the expected losses by 50% (because of the increase in estimated default probabilities) and increases economic capital by 45% to 55%, depending on the statistic and confidence level being used.

Concluding remarks

We have developed a simulation-based framework to estimate the one-period credit loss distribution of a retail loan portfolio, and demonstrated the usefulness of the model by estimating one-year credit losses for a credit card portfolio. The framework allows risk to be decomposed into its various sources, and provides further understanding of risk concentrations and the impact of economic factors on the portfolio.

We present and test three models to describe joint default behavior: a sector-based logit model, a factor-based logit model and a factor-based Merton model. While the sector-based model is purely descriptive, the factor-based models are causal models that capture the economic cycle through macroeconomic factors. Hence, the latter models are useful for further stress testing and estimating conditional losses using economic scenarios.

All three default models yield comparable results when the same data is used. The discrepancies that arise are due to different distributional assumptions. The discrepancies observed are small, and are probably amplified by the scarcity of data. More statistical analysis is required to explore the differences between the models.

Several reasonable and commonly used assumptions may influence the results of the models. First, sector indices and factor returns are assumed to be joint normally distributed. Although this is common practice, it may be unrealistic; more research is required to determine the effect of this assumption. Second, the exposure and recovery rates per sector are deterministic; although this is common to most portfolio models today, it may be useful to explore the implication of stochastic exposures using an integrated market and credit risk model, such as that introduced in Iscoe et al. (1999). Finally, each sector is assumed to have a large, fully diversified sub-portfolio; this is reasonable given the large number of accounts in each sector, but may lead to some underestimation of credit risk. This assumption could be relaxed easily during the simulation.

One of the main limitations in the case study is that of data availability. Although the results are useful, a large amount of historical data may be required to obtain more robust estimations of joint default probability distributions and factor models. The lack of sufficient data leads to two problems.

The first problem is that the uncertainty associated with the parameters entering the model is quite large. This does not mean that the parameters should be discarded but, rather, that they should be treated as "best guesses," given the current information. Consequently, more stress testing and sensitivity analysis must be performed. If conservative estimates of credit risk are required, one can apply the most conservative parameters obtained from the data.

The second problem is that given that the default data covers less than three years, the model may not capture the dependency of default frequencies on the economic cycle. In this case, data over a longer time horizon, which spans economic cycles, is required. We show how the impact of economic cycles can be addressed using multi-factor models. Of course, more work in this area is required.

Since data availability is often an issue, it is important, whenever possible, to complement the data used for estimation with external industry/agency data, as well as reasonable, conservative data acquired from internal experience.

The results from the case study clearly show that some refinement in the modelling of the portfolio may lead to the greatest improvements. For example, as expected, there is a large concentration of losses in the unscored sectors. These sectors are also the most likely to be inhomogeneous sectors; the use of historical data to estimate default rates for these sectors may be less accurate. Further refinement and classification of accounts in these sectors is likely to lead to substantial enhancements in the credit risk assessment of the portfolio.

In this study, the score of each card at acquisition is used as an indicator of default likelihood over the following year. This means that the results assume that the credit cards in the portfolio at the time of the analysis maintain the score they were assigned initially, or, alternatively, that migration across sectors is such that the net of migration across sectors is very small. This may be a strong assumption. A better way to address this problem is to use current, up-to-date scores and account characteristics to group obligors into sectors. However, for many institutions this data may not be readily available and acquiring it may be a costly proposition. An alternative in this case is to use Bayesian methods to refine the composition of the cards in each sector, given all historical default experience.

The methodology presented in this paper can also be used to obtain risk management reports that include further sensitivity analysis and stress testing, RAROC and risk-return analysis, systemic and idiosyncratic risk decomposition, risk contributions of economic factors and conditional credit risk calculation using factor-based models, and economic forecasts.

In conclusion, the application of portfolio credit risk models to retail portfolios is in its infancy and much more research is required. Of particular importance is the validation and backtesting of the models. This is an issue that was raised by the Basle Committee (1999) and work in this area will likely require extensive (external) data and intensive computation (see, for example, Lopez (2000)).

Acknowledgements

We would like to thank David Dallaire, Alex Kreinin and Leonid Merkoulovitch for fruitful discussions and help provided in the estimation of the different models and Michael Zerbs for constructive feedback on the analysis of the results.

References

Artzner, P., F. Delbaen, J.M. Eber and D. Heath, 1998, "Coherent measures of risk," Working Paper, Institut de Recherche Mathématique Avancée, Université Louis Pasteur et C.N.R.S.

Bank for International Settlements, *A New Capital Adequacy Framework: consultative paper,* Basle Committee on Banking Supervision, April 1999.

Bucay, N. and D. Rosen, 1999, "Credit risk of an international bond portfolio: a case study," *Algo Research Quarterly,* 2(1): 9–29.

Carey, M., 1998, "Credit risk in private debt portfolios," *Journal of Finance,* 53(4): 1363–1387.

Credite Suisse Financial Products, 1997, *CreditRisk+: A Credit Risk Management Framework*, New York, NY.

Crouhy, M., and Mark, R., 1998, "A comparative analysis of current credit risk models," Paper presented at the conference Credit Modeling and Regulatory Implications, London, September.

Finger, C., 1999, "Conditional approaches for CreditMetrics portfolio distributions," *CreditMetrics Monitor,* April: 14–33.

Gordy, M., 2000, "A comparative anatomy of credit risk models," *Journal of Banking and Finance,* 24(1/2): 119–149.

Gupton, G., C. Finger and M. Bhatia, *CreditMetrics: The Benchmark for Understanding Credit Risk,* Technical Document, 1997, New York, NY: JP Morgan Inc.

Iscoe, I., A. Kreinin and D. Rosen, 1999, "An integrated market and credit risk portfolio model", *Algo Research Quarterly,* 2(3): 21–37.

Kealhofer, S., 1996, "Managing default risk in portfolios of derivatives," in *Derivative Credit Risk,* London: Risk Publications.

Koyluoglu, H. and A. Hickman, 1998, "Reconcilable differences," *Risk,* 11(10): 56–62.

Kreinin, A., L. Merkoulovitch, D. Rosen and M. Zerbs, 1998, "Principal component analysis in quasi Monte Carlo simulation," *Algo Research Quarterly,* 2(2): 21–29.

Lewis, A., 1992, *An Introduction to Credit Scoring,* Fair Isaac & Co., San Rafael, CA.

Lopez, J. and M. Saidenberg, 2000, "Evaluating credit risk models," *Journal of Banking and Finance,* 24(1/2): 151–165.

Mays, E. ed., 1998, *Credit Risk Modeling: Design and Application,* Glenlake Publishing Co., Chicago, IL.

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

Nagpal, K. and R. Bahar, 1999, "An analytical approach for credit risk analysis under correlated defaults," *CreditMetrics Monitor,* April: 51–74.

Pritsker, M., 1997, "Evaluating value at risk methodologies: accuracy vs. computational time," *Journal of Financial Services Research,* 12(2/3): 201–242.

Richeson, L., R. Zimmerman and K. Barnett, 1994, "Predicting consumer credit performance: can neural networks outperform traditional statistical methods?," *International Journal of Applied Expert Systems,* 2(2): 116–130.

Standard & Poor's, 1999, *Financial Market and Economics Database*, Lexington, MA: DRI/McGRaw-Hill.

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

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

Appendix 1. Calibration of joint default probability models

The parameters of the joint default probability models are estimated as follows.

Sector-based logit model

Data input: for each sector j, a time series of equally-spaced observed default rates $p_j(t_i)$, $i = 1,\ldots,n$.

Output: joint distributions of the sector indices, Y_j^S.

Since the indices are assumed to be joint normal, we estimate the mean for each index and the covariance matrix. For each sector, the time series of default probabilities is transformed into a time series of logit variables

$$y_j^S(t_i) = ln\left(\frac{1 - p_j(t_i)}{p_j(t_i)}\right), \ i = 1,\ldots,n$$

The mean, volatility and correlations of the indices are obtained from the sample mean, volatility and correlations of this time series.

The indices, y_j^S, are assumed to be normal (not standard normal); they have a non-zero mean, μ_j^S, and non-unit standard deviation, σ_j^S. It is straightforward to express the model in terms of standard normal indices $Y_j^S = (y_j^S - \mu_j^S)/\sigma_j^S$.

Factor-based models

Data input: a time series of equally spaced observed default rates for each sector j and of macroeconomic factors returns,
$p_j(t_i)$, $x_k(t_i)$, $k = 1,\ldots, q^M$, $i = 1,\ldots,n$.

Output: the definition of the q independent credit drivers and the multi-factor model joining the indices to the credit drivers. Since the credit drivers include q^M macroeconomic factors and q^S sector-specific factors, the total number of credit drivers is $q = q^M + q^S$.

We write the (row) vector of uncorrelated standard credit drivers as
$\mathbf{Z} = (Z_1^M, \ldots, Z_{q^M}^M, Z_1^S, \ldots, Z_{q^S}^S)$ where the Z_j^M and Z_j^S denote macroeconomic and sector-specific credit drivers, respectively. A sector-specific factor affects only a single sector. Thus, for a given obligor l in sector j, the creditworthiness index in Equation 2 can be written as

$$Y_l = \sum_{k=1}^{q^M} \beta_{jk}^M Z_k^M + \beta_j^S Z_j^S + \sigma_j \varepsilon_l$$

where

$$\sigma_j = \sqrt{1 - \left(\sum_{k=1}^{q^M} \beta_{jk}^2 + (\beta_j^S)^2\right)}$$

A set of independent factors are constructed from the macroeconomic factors, as follows:

- First, obtain a time series of standardized factor returns for the macroeconomic credit drivers, $\hat{Z}_k(t_i) = (x_k(t_i) - \mu_k)/\sigma_k^x$, $k = 1,\ldots, q^M$ where μ_k and σ_k^x are the sample mean and standard deviation of each factor.
- Second, using principal component analysis (PCA), obtain a set of independent credit drivers as the linear combinations of the correlated macroeconomic factors. A brief overview of PCA is given in Kreinin et al. (1998).
- Third, construct a time series of the new uncorrelated standardized factors $Z_k(t_i) = \sum_{l=1}^{q} A_{kl}\hat{Z}_k(t_i)$, $i = 1,\ldots,n$; where the coefficients A_{kl} are determined from the eigenvalues and eigenvectors in the PCA.

The methods used to obtain the weights for the creditworthiness index of each model are outlined below.

From Tables 1 and 2, the model of the conditional probability of default for the **factor-based logit model** is

$$p_j(\mathbf{Z}) = \frac{1}{1 + a\ exp(bY_j^S)}$$

where

$$Y_j^S = \sum_{k=1} \beta_{jk}^{M*} Z_k^M + \beta_{jj}^{S*} Z_j^S$$

and

$$\sum_k (\beta_{jk}^{M*})^2 + (\beta_{jj}^{S*})^2 = 1$$

The factor weights are calculated as follows:

- Similar to the sector-based model, from the time series of default rates in each sector, we obtain a time series of logit variables:

$$y_j^S(t_i) = Ln\left(\frac{1-p_j(t_i)}{p_j(t_i)}\right),\ i = 1,\ldots,n$$

Each logit variable has mean μ_j^S and standard deviation σ_j^S.

- The component of the multi-factor model that depends on the macroeconomic factors is estimated by minimum least squares using

$$y_j^S(t_i) - \mu_j^S = \sum_{k=1}^{q^M} \tilde{\beta}_{jk}^M Z_k^M(t_i) + \xi_j(t_i)$$

where $\xi_j(t_i)$ are independent normal errors.

- The loading of the sector-specific factor, $\tilde{\beta}_{jj}^S$, is computed by matching the total volatility of the multi-factor model of the sector and the observed sample index volatility:

$$\tilde{\beta}_{jj}^S = \sqrt{(\sigma_j^S)^2 - \sum_k (\tilde{\beta}_{jk}^M)^2}$$

- Finally, the parameters of the model are given by

$$\beta_{jk}^{M*} = \tilde{\beta}_{jk}^M / \sigma_j^S, \quad \beta_{jj}^{S*} = \tilde{\beta}_{jj}^S / \sigma_j^S$$

$$a_j = exp(\mu_j^S), \quad b_j = \sigma_j^S$$

The process for the **factor-based Merton model** is similar to that described for the logit model. We estimate the parameters α_j, β_{jk}^M, β_{jj}^S and σ_j such that

$$p_j(\mathbf{Z}) = \Phi\left\{\frac{\alpha_j - \left(\sum_k \beta_{jk}^M Z_k^M + \beta_{jj}^S Z_j^S\right)}{\sigma_j}\right\}$$

and

$$\sigma_j = \sqrt{1 - \left(\sum_k (\beta_{jk}^M)^2 + (\beta_{jj}^S)^2\right)}$$

These parameters are obtained as follows:

- From the time series of default rates in each sector, we obtain a time series of inverse normal variables:
$\tilde{\alpha}_j(t_i) = \Phi^{-1}\{p_j(t_i)\},\ i = 1,\ldots,n.$

- For each sector j, we first apply minimum least squares to estimate $\hat{\alpha}_j$ and $\hat{\beta}_{jk}^M$ from the regression model

$$\Phi^{-1}(p_j) = \tilde{\alpha}_j = \hat{\alpha}_j + \sum_{k=1}^{q^M} \hat{\beta}_{jk}^M Z_k^M + \xi$$

- Then, we obtain the sensitivity of the inverse standard normal variable of sector j to the sector-specific credit driver by the following relationship

$$\hat{\beta}_{jj}^S = \sqrt{\sigma_{\alpha_j}^2 - \sum_{k=1}^{q^M} (\hat{\beta}_{jk}^M)^2}$$

where $\sigma_{\alpha_j}^2$ is the variance of $\tilde{\alpha}_j(t_i)$, which is the inverse standard normal variable for the sector over the period.

- The coefficients in the multi-factor model are then obtained by properly scaling the regression coefficients by the volatility of the idiosyncratic obligor component of the index

$$\beta_{jk}^M = \hat{\beta}_{jk}^M \cdot \sigma_j$$

$$\alpha_j = \hat{\alpha}_j \cdot \sigma_j$$

$$\beta_{jj}^S = \hat{\beta}_{jj}^S \cdot \sigma_j$$

where the volatility is finally given by

$$\sigma_j^2 = \frac{1}{1 + \sigma_{\alpha_j}^2}$$

The derivation of these formulae is presented below. Recall from Equation 6 that the conditional default probabilities in the Merton model are given by

$$p_j(Z) = \Phi\left(\frac{\alpha_j - \left(\sum_{k=1}^{q^M}\beta_{jk}Z_k + \beta_j^S Z^S\right)}{\sigma_j}\right) \qquad (A2)$$

From the time series of conditional default probabilities and macro-economic factors, we estimate the parameters α_j, β_{jk}, β_j^S and σ_j in Equation A2 with the extra constraint that

$$\sigma_j = \sqrt{1 - \left(\sum_{k=1}^{q^M}\beta_{jk}^2 + (\beta_j^S)^2\right)} \qquad (A3)$$

Applying an inverse normal transformation and making the change of variables

$$\hat{\alpha}_j = \alpha_j/\sigma_j \ , \quad \hat{\beta}_{jk} = \beta_{jk}/\sigma_j, \quad \hat{\beta}_j^S = \beta_j^S/\sigma_j \qquad (A4)$$

Equation A2 becomes the equation defined for the regression:

$$\Phi^{-1}(p_j) = \tilde{\alpha}_j = \hat{\alpha}_j + \sum_{k=1}^{q^M}\hat{\beta}_{jk}Z_k + \hat{\beta}_j^S Z^S$$

By regressing the inverse normal of the conditional default probabilities to the macro-economic factors we obtain the parameters $\hat{\beta}_{jk}$, and the residual volatility gives the sensitivity of the inverse to the sector-specific factor:

$$\hat{\beta}_j^S = \sqrt{\sigma_{\alpha_j}^2 - \sum_{k=1}^{n}\hat{\beta}_{jk}^2} \qquad (A5)$$

where $\sigma_{\alpha_j}^2$ is the variance of $\tilde{\alpha}_j(t_i)$.

The original parameters of the model are then obtained as follows. First the volatility σ_j can be estimated from Equation A3, Equation A4 and Equation A5:

$$\sigma_j^2 = 1 - \left(\sum_{k=1}^{q^M}\beta_{jk}^2 + (\beta_j^S)^2\right) = 1 - (\sigma_j^2 \cdot \sigma_{\alpha_j}^2)$$

Therefore, $\sigma_j^2 = \dfrac{1}{1+\sigma_{\alpha_j}^2}$.

Finally α_j, β_{jk} and β_j^S are obtained by simply substituting back in Equation A4.

Appendix 2. Estimation results

Sector	1	2	3	4	5	6	7	8	9	u_1	u_2
1	100										
2	37	100									
3	62	72	100								
4	76	57	88	100							
5	68	67	73	73	100						
6	37	35	57	62	56	100					
7	26	22	36	37	47	55	100				
8	13	43	41	19	33	8	9	100			
9	29	34	38	35	35	42	50	14	100		
u_1	35	14	35	45	44	52	–2	–6	8	100	
u_2	16	42	35	30	20	28	44	–12	16	–9	100

Table A1: Sample correlations of default rates per sector (%)

Sector	1	2	3	4	5	6	7	8	9	u_1	u_2
1	100										
2	32	100									
3	54	81	100								
4	69	721	82	100							
5	63	641	72	61	100						
6	38	25	37	58	40	100					
7	29	27	38	45	42	63	100				
8	30	52	45	34	37	-3	-10	100			
9	21	34	36	31	29	33	29	9	100		
u_1	49	21	41	44	48	48	4	1	8	100	
u_2	22	37	27	47	14	35	389	10	7	-1	100

Table A2: Sample correlations of logit variables (%)

	Sector										
	1	2	3	4	5	6	7	8	9	u_1	u_2
Constant	2.550	3.024	3.614	3.740	3.977	4.222	4.541	4.823	5.820	4.244	4.420
Factor 1	0.183	-0.049	-0.016	0.004	0.129	0.002	0.081	0.033	0.208	-0.054	-0.136
Factor 2	0.002	0.114	0.085	0.069	0.100	0.049	0.122	-0.031	-0.005	-0.060	-0.044
Factor 3	-0.096	0.028	0.049	-0.115	0.082	-0.125	0.041	-0.005	-0.058	0.047	-0.117
Factor 4	0.063	0.101	0.048	0.062	0.020	0.068	0.075	-0.043	0.089	-0.002	0.175
Factor 5	0.007	-0.077	0.142	0.043	0.008	0.012	-0.007	-0.049	0.129	0.126	-0.091
Factor 6	-0.061	0.143	0.154	0.177	0.087	0.108	0.111	-0.044	0.121	0.001	0.260
Factor 7	-0.055	-0.102	-0.078	-0.111	-0.066	-0.077	-0.134	-0.209	-0.055	-0.021	-0.122
Factor 8	0.133	0.144	0.289	0.222	0.179	0.231	0.171	0.134	0.135	0.013	0.241
Factor 9	-0.300	-0.208	-0.364	-0.337	-0.302	-0.191	-0.081	-0.100	-0.091	-0.349	0.014
Adjusted R^2	47.96	53.50	59.84	48.66	60.12	20.27	7.24	21.18	25.71	25.42	45.75
F-statistic	3.765	4.196	5.470	3.843	5.523	1.763	1.234	1.806	2.038	1.985	3.530

Table A3: Regression results for factor-based logit model and F-statistics

	Sector										
	1	2	3	4	5	6	7	8	9	u_1	u_2
Constant	-1.45	-1.68	-1.93	-1.98	-2.08	-2.18	-2.30	-2.41	-2.75	-2.19	-2.25
Factor 1	-0.09	0.02	0.01	0.00	-0.05	0.00	-0.03	-0.01	-0.07	0.02	0.05
Factor 2	0.00	-0.05	-0.03	-0.03	-0.04	-0.02	-0.04	0.01	0.00	0.02	0.02
Factor 3	0.04	-0.01	-0.02	0.04	-0.03	0.05	-0.02	0.00	0.02	-0.02	0.04
Factor 4	-0.03	-0.04	-0.02	-0.02	-0.01	-0.02	-0.03	0.02	-0.03	0.00	-0.06
Factor 5	0.00	0.03	-0.06	-0.02	0.00	0.00	0.00	0.02	-0.04	-0.05	0.03
Factor 6	0.02	-0.06	-0.06	-0.07	-0.03	-0.04	-0.04	0.02	-0.04	0.00	-0.10
Factor 7	0.03	0.04	0.03	0.05	0.03	0.03	0.05	0.08	0.02	0.01	0.04
Factor 8	-0.06	-0.06	-0.12	-0.09	-0.07	-0.09	-0.07	-0.05	-0.04	0.00	-0.09
Factor 9	0.14	0.09	0.15	0.14	0.12	0.07	0.03	0.04	0.03	0.14	-0.01
Sector 1	0.14										
Sector 2		0.15									
Sector 3			0.12								
Sector 4				0.14							
Sector 5					0.09						
Sector 6						0.14					
Sector 7							0.14				
Sector 8								0.11			
Sector 9									0.11		
Sector u_1										0.14	
Sector u_2											0.13

Table A4: Weights of factor-based Merton model