Is Firm Interdependence within Industries Important for Portfolio Credit Risk?

Kenneth Carling*, Lars Rönnegård†, and Kasper Roszbach‡§

First version: 8 August 2004
This version: 22 January 2007 (under revision)

Abstract

A drawback of available portfolio credit risk models is that they fail to allow for default risk dependency across loans other than through common risk factors. Thereby, these models ignore that close ties can exist between companies due to legal, financial and business relations. In this paper, we integrate the insights from theoretical models of default correlation into a commonly used model of default and portfolio credit risk by explicitly allowing for dependencies between firm defaults through both common factors and industry specific disturbances in a duration model. An application using pooled data from two Swedish banks' business loan portfolios over the period 1994-2000 shows that estimates of individual default risk are little affected by including industry specific errors. However, accounting for the within-industry correlation of defaults increases estimates of VaR by 50-200 percent. We show that the model we propose manages to follow both the trend in credit losses and produce industry driven, time-varying, fluctuations in losses around that trend. A conventional model that contains only systematic factors as drivers of default correlation, although able to fit the broad trends in credit losses, cannot match these fluctuations because it fails to capture credit losses in bad times, when banks are typically hit by large unexpected credit losses. The model developed here should thus aid banks and supervisors in determining the appropriate size of economic capital requirements. Our estimations show that it is likely that banks need larger capital buffers than conventional models indicate.

JEL classification: C34, C35, D61, D81, G21
Keywords: correlation, default, value-at-risk, credit risk, portfolio credit risk, duration model, industry dependency, cluster.

*IFAU, Uppsala, Sweden, and Dalarna University, SE 781 88 Borlange, Sweden; kca@du.se
†Uppsala University, Sweden; lars.ronnegard@lcb.uu.se
‡Research Division, Sveriges Riksbank, SE 103 37 Stockholm, Sweden; kasper.roszbach@riksbank.se
§We are grateful to comments and suggestions from Mark Carey, Tor Jacobson, Til Schuermann, Christian Wolff and seminar participants at Norges Bank, Erasmus University, CREDIT, the Probanker workshop in Maastricht, SITE/Stockholm School of Economics, the Econometric Society world congress 2005, the FMA 2006 European meeting, the OCC, the Board of Governors and the Federal Reserve Bank of San Francisco. The views expressed in this paper are solely the responsibility of the authors and should not interpreted as reflecting the opinon of the Executive Board of Sveriges Riksbank.

1 Introduction

Banks play an important role in the economy as savings institutions and as providers of credit and capital. In addition to supervision, deposit insurance and other regulatory conditions, capital requirements are used by governments to limit risks for depositors, to reduce the likelihood of both individual bank defaults and to keep down systemic risk. However, these capital requirements are likely to form restrictions on the workings of banks. Inadequate capital buffers may lead to unacceptable levels of portfolio credit losses, undesirable bank defaults and greater systemic risk, but excessive capital requirements will restrain credit provision needlessly. Hence striving for both stability and efficiency in financial markets and the economy at large will involve trading-of pursuing conflicting policies.

A common way for banks to determine the required amount of economic capital is to employ a portfolio credit risk model and calculate the loss distribution for a specific time horizon. After choosing an acceptable level of insolvency risk, a bank then pins down the capital requirement by subtracting expected credit losses from the loss at the percentile corresponding to the selected level of insolvency risk. Unfortunately, no real consensus consists yet among academics, practictioners and regulators on which models are most suitable for this purpose. As a result, estimates of required capital buffers are likely to vary, depending on the modeling approach that was chosen. Gordy [25] has noted that even very frequently used models of portfolio credit risk can vary widely in their estimates of economic capital, due to parameter sensitivity. According to Mingo [41], such variation will incite regulatory capital arbitrage by banking institutions. Such tendencies will make it even harder for regulators and supervisors to both achieve stability goals and improve financial market efficiency.[1]

Portfolio credit risk models have predominantly been developed during the last two decades. In recent years, supervisory and regulatory authorities have furthered the development of statistical models to measure portfolio credit risk, primarily by means of the new Basel II Capital Accord [11]. This Accord includes both more sophisticated rules for determining minimum capital requirements and guidelines for the type of models that banks can employ internally for purposes of credit risk analysis. In part due to these guidelines, research on portfolio credit risk modeling has made substantial advances. Broadly viewed, there are four different groups of modeling approaches: 'structural', econometric factor risk, 'top-down' actuarial, and non-parametric models. Gordy [24] compared two frequently used models and found that they were highly sensitive to the default correlations.

Not surprisingly, the modeling of asset correlations and default correlations has been receiving special attention. A variety of methods has been used to incorporate default correlations into

[1] Jacobson, Lindé and Roszbach [31] find evidence for Swedish banks that supports Gordy's and Mingo's conclusions.

portfolio models and to estimate them. Common for most approaches is, however, that any correlation is assumed to be fully captured by a set of common risk factors. Adding factors, for example, has thus been a way to increase the degree of comovement between business defaults. One of the exceptions is the nonparametric resampling method proposed by Carey [16]. It does not need any assumptions about asset correlations, since all comovement is captured by in the data that is resampled. This method is thus very useful for deriving loss distributions, but has as a downside in that it gives no insight into the driving forces behind extreme events. Das, Duffie, Kapadia and Saita [19] follow yet another approach. Das et al. test, and reject, the doubly stochastic assumption under which firms' defaults are correlated only by the factors determining their default intensities. They find evidence of default clustering exceeding that implied by the doubly stochastic model. To gauge the degree of correlation that is not captured by a Poisson model, they then calibrate an intensity-conditional (flat Gaussian) copula model. This copula correlation parameter measures of the amount of additional default correlation that must be added, on top of the default correlation implied by the co-movement of the factors, to match the upper-quartile moments of the empirical distribution of defaults. Their estimated incremental copula correlation is moderately small and ranges from 1% to 4%.

On the theory side Giesecke [22] approaches the correlation between firm defaults quite differently. He argues that defaults are correlated between firms because the default thresholds, the points where assets exceed debts, are not exogenously given, but endogenous and linked between firms. Except for being exposed to common (national, regional and sectorial) factor fluctuations, firms are tied to each other through both legal (e.g. parent-subsidiary), financial (e.g. trade credit) and business relations (e.g. buyer-supplier). Therefore, the economic distress of one firm can have an adverse effect on the financial health of other businesses, beyond the measurable effect of macroeconomic and other common factors. Default thresholds are thus also thought to be (partly) driven by common or systematic factors. Zhou [55] also analyzes the occurrence of multiple defaults and produces some insights that are relevant for risk management. Among other things, he finds that default correlations are generally very small over short horizons - they first increase and then slowly decrease over time; for given horizons, however, high credit quality is associated with a low default correlation. Zhou concludes that diversification can generate substantial reductions in risk over short horizons, because defaults are very weakly correlated within such time ranges. But for long term investments (5-10 years), default correlations can become an important factor to take into account if the underlying values are highly correlated. Concentration in one specific industry or region can therefore be very risky. Also, changes in the underlying values that lead to movements in credit quality can lead to substantial changes in credit risk, thereby requiring appropriate adjustments of the capital buffer.

Unfortunately, most commonly used models of portfolio credit risk do not account for the type of interactions that Giesecke describes. More importantly, these models are not able to account for the low frequency peaks in default rates and credit losses that we observe in the data. Das et al.[19] confirm this. At the same time, the work by Das et al is maybe the only and most successful approach to quantify this failure of portfolio models. Despite this achievement, their copula approach is less conclusive in helping us understand what drives or interpret the interdependence between business defaults.

Clearly, restrictive assumptions about the dependency between firms, such as assuming common factors and independent disturbances, paired with an inability to match peaks in default rates, form an unattractive and undesirable property for portfolio credit risk models. In addition it is likely to lead to the underestimation of portfolio credit losses and the required (marginal) capital charges. Policies based on statistics derived from such portfolio credit risk models may therefore inadvertently indicate too low a level of insolvency risk and, at the industry level, too high a level of financial stability. This in its turn could result in inefficiencies due to unexpected bankruptcies and bring about unnnecessary bankruptcy costs.

What do we ideally want from a good model?

- *Good prediction of individual PD*
- *Good prediction of the total # defaults*
- *Good ability to predict portfolio credit losses*
- *Capture, model and quantify the interaction between individual firms' probability of default*
- *Get a statistic that quantifies to what extent the model is better than a model that only predicts individual PD's*

Include story about Duffie Eckner et al ...

This paper aims at integrating some of the insights from default correlation theory into a more "conventional" model of default risk. We do so by allowing for dependencies between defaults not only through common factors but also through dependent error terms. The "conventional" model we use is a duration model as the one used by Shumway [48] and Carling et al. [17]. However, to increase the model's ability to match peaks in default rates, we allow for both idiosyncratic shocks and industry specific shocks. The industry shocks can be interpreted as within-industry correlations of defaults. The approach we choose is thus much like a non-linear panel model with random effects. The paper starts with a theoretical section that outlines the basics of the panel duration approach. We also show how the model can be applied in practice. For this purpose, we pool data from two Swedish banks, estimate the model, calculate the implied loss distributions and compare them to actual credit losses.

Our results show that significant and substantial differences exist between industries in the degree of inter-firm correlation of defaults. Retail businesses and firms in the public sector display stronger than average comovement while companies in the real estate and finance industry exhibit by far the largest correlation of defaults. An implication of this intra-industry comovement of defaults is that estimates of VaR increase by 50-200 percent, depending on the point in time and the aggregate economic conditions. The model we propose thus suggests that larger economic capital buffers are required for banks.

The rest of this paper is organized as follows. In Section 2 we summarize the main contributions of the literature, and show how a standard duration model of default risk can be extended to include within-industry correlation of defaults. Section 3 briefly discuss the data. In Section 4 we display the estimation results. Section 5 summarizes the paper and discusses potential extensions.

2 Modelling default risk

Broadly viewed, there are four groups of portfolio credit risk models.[2] The first group is 'structural' and based on Merton's [40] model of firm capital structure: individual firms default when their assets' value fall below the value of their liabilities. Examples of such a microeconomic causal model are CreditMetrics and KMV's PortfolioManager. The second group consists of econometric factor risk models, like McKinsey's CreditPortfolioView and Carling, Jacobson, Lindé and Roszbach [17]. The former employs a logistic (cross-sectional) model where a macroeconomic index and a number of idiosyncratic factors together determine default risk in 'homogeneous' subgroups, while the latter builds on a duration (panel data) model of default risk. Both groups apply similar Monte Carlo simulations to calculate portfolio risk, as both are 'bottom-up' models that compute default rates at either the individual firm level or at sub-portfolio level.[3] Both thus require a similar kind of aggregation. The third group contains 'top-down' actuarial models, like Credit Suisse's CreditRisk+, that make no assumptions with regard to causality. The last group consists of non-parametric approaches, as in Carey [16], who estimates credit loss distributions by means of a Monte Carlo resampling method. Calem and LaCour-Little [14] combine the second and fourth method and estimate a survival time model for mortgage loan data and apply Carey's method to simulate default risk distributions.

Few models incorporate both microeconomic and macroeconomic effects on default or credit risk. Carling et al. [17], Nickell, Perraudin and Varotto [44] and Wilson [52] are exceptions.[4] CreditPortfolioView incorporates a set of macroeconomic variables in a multifactor

[2] This is by no means the only or an exhaustive way to describe the directions in which this literature has evolved.

[3] Default risk modeling started in the late 1960s with work by Altman [4] [5] [6] [7]. More recently Frydman, Altman and Kao [21], Li [36], and Shumway [48] have contributed to this literature. These papers all analyze default risk at the firm level and assume that default correlations are adequately captured by a set of common risk factors.

[4] Wilson describes the principles behind McKinsey's portfolio credit risk model CreditPortfolioView.

country/sector logit model, with each factor modeled as an ARIMA process. Nickell et al. include macro dummies in an ordered probit model of rating changes. Carling et al. include firm specific panel data and macroeconomic time series as explanatory variables in a duration model of bank loan survival time. Koopman, Kräussl, Lucas and Monteiro [34] use U.S. rating transition and default data to relate default and rating dynamics to the business cycle, bank lending conditions, and financial market variables. Other approaches involve transition matrices to generate a linkage between underlying macroeconomic conditions and asset quality (e.g. Bangia, Diebold, Kronimus, Schagen and Schuermann [10]) or Monte Carlo resampling methods to generate unconditional portfolio credit loss distributions (e.g. Carey [16]). Pesaran, Schuermann, Treutler and Weiner [45] and Jacobson, Lindé and Roszbach [32] follow a novel, somewhat similar, approach. Both link a model of default risk and a model of the macroeconomy in order to let credit losses react to macro shocks and to allow for stress testing excercises. However, Jacobson et al. use default data of non-quoted firms, while Pesaran et al. link asset value changes of publicly quoted businesses to a dynamic global macroeconometric model.

Despite some differences, the first three of the above approaches to estimating portfolio loss distributions build on three more or less general components (see Koyluoglu and Hickman [35]). First, they contain some process that generates conditional default rates for each borrower in each state of nature and a measure of covariation between borrowers in different states of nature. Second, their set-up allows for the calculation of conditional default rate distributions for sets of homogeneous subportfolios (e.g., rating classes) as if individual borrower defaults are independent, since all joint behavior is accounted for in generating conditional default rates. Third, unconditional portfolio default distributions are obtained by aggregating homogeneous subportfolios' conditional distributions in each state of nature; then conditional distributions are averaged using the probability of a state of nature as the weighting factor. Gordy's [24] comparison of two commercially developed and frequently used models of portfolio credit risk, CreditMetrics and CreditRisk+, confirms the general insights of Koyluoglu and Hickman. He concludes that both models have very similar mathematical structures and that discrepancies in their predictions originate in different distributional assumptions and functional forms. Among other things, he concludes that the models are highly sensitive to both the average default correlations in the model - that in turn determine default rate volatility - and the shape of the implied distribution of default probabilities.[5] [6]

[5] Gordy [25] and Szegö [50] discuss the properties of Value-at-Risk (VaR) and compare it with other summary statistics of distributions as an instrument to evaluate capital requirements. Portfolio invariance for individual exposures' contribution to VaR is obtained only if the dependence across exposures is driven by a single (systematic) risk factor.

[6] Lucas, Klaassen, Spreij and Straetmans [?][?] derive an analytical approximation to the credit loss distribution in a CreditMetrics/Merton model, and show that such a model can even be used when one only has portfolios with relatively small numbers of counterparts at ones disposal.

Our objective in this paper is to develop a model of business default risk that can be used not only to adequately predict individual firms' default but also to accurately estimate expected and unexpected loan losses for a portfolio of credit exposures. Given the problems of earlier work in modeling correlations, a key property of such a model will have to be an improved ability to capture the comovement of default risk among businesses. Typically, this kind of reasoning leads to a binary regression model, of which the logistic regression model and the probit model are two examples with frequent use in credit-risk modelling. In extending the model to allow for potential interdependence in the portfolio, we incorporate random effect components in the binary regression model. Such a model is commonly referred to as GLMM (Generalised Linear Mixed Models) in the statistical literature. We begin however by outlining the model in the framework of linear models and postpone the complication that the binary variable default implies a non-linear model to the last subsection that discusses estimation issues.

2.1 Independent disturbances

Conceptually, we think of the loans as being governed by a duration process where loans eventually default. This way of modelling default duration differs from the approach in Shumway [48], since we allow the probability of default to depend on the length of the time period that the loan has been "active" and default has been avoided. Nevertheless, to keep focus we assume for now that the duration process has no memory.

Consider the latent-variable y_i^τ the propensity, for the i:th loan in the portfolio at calendar time τ, to default in a pre-determined time-unit ahead of τ. That is y_i^τ is an index taking on any real-value. Clearly, the propensity might vary between loans in the portfolio and over time and the question is what explains this variation. With each exposure we associate a set of characteristics x_i (at the time-point τ), being the unique properties of the firm at this time point as well as common properties to all firms, or a subset of them. We will denote the characteristics $x_i(\tau)$ to make it explicit that they are unique to the firm and/or to the calendar time τ. The model can be expressed as a relationship between the propensity to default and a set of characteristics (the systematic part) x_i as

$$y_i^\tau = x_i(\tau)\beta + \varepsilon_i^\tau \tag{1}$$

where β is a parameter-vector measuring the effect of the explanatory variables on the propensity to default and ε_i^τ is an unsystematic exposure specific component. The unsystematic component is perhaps best thought of as the bundle of all characteristics, not in the systematic part, that might influence the propensity of default for a loan. The model is much simplified if it is assumed that $x_i(\tau)$ and ε_i^τ are stochastically independent. Such assumption will have to be verified for the empirical specification that is chosen, but in general a richer specification of the systematic part is likely to make the assumption more realistic and tolerable. We also assume independence of the errors between both loans and time points, i.e. $f\left(\varepsilon_i^\tau, \varepsilon_j^\tau\right) = f\left(\varepsilon_i^\tau\right) f\left(\varepsilon_j^\tau\right)$ for $i \neq j$ and $f\left(\varepsilon_i^\tau, \varepsilon_i^{\tau+l}\right) = f\left(\varepsilon_i^\tau\right) f\left(\varepsilon_i^{\tau+l}\right)$ for $l \neq 0$.

2.2 Disturbances with dependency across industries

In the model in equation (1), the outcome of a single loan is uncertain because of the loan specific error-term. A portfolio that consist of just one loan would have a potential loss that exceeds greatly the expected loss. However, for large portfolios with loans of equal size, the potential loss would be roughly equal to the expected loss since the error-terms would average out, even if the unexplained cause of default for a single loan would be substantial. This property of the model and the implied assumption of independence across loans in the portfolio is not particularly appealing considering the likely interdependency of some companies.

As an extension of the model, we think of the companies in the portfolio as belonging to only one out of K clusters. As Giesecke [22] mentioned, it is likely that there are strong business relations between companies within clusters and that some common shock may hit all companies in a cluster. Typically, firms are tied to each other through both legal, financial and business relations, as a result of which the economic distress of one firm can have an adverse effect on the financial health of other businesses beyond the measurable effect of macroeconomic conditions and other systematic fluctuations. Alternatively, both can be the result of simultaneous exposure to a common shock. We denote the cluster specific shock by ε_k and assume it follows the normal distribution with expectation equal to zero, while the clusters are labeled $k = 1, ..., K$. We now take the unexplained cause of default to depend on the loan specific shock, ε_i^τ, and the common shock, ε_k^τ, to the cluster the loan belongs to, where, by construction, ε_i^τ and ε_k^τ are independent. In addition, we assume a cluster-specific shock is independent of the cluster-specific shocks for other clusters, i.e. ε_k^τ is independent of ε_{k+j}^τ for all $j \neq 0$.[7] The model in equation (1) is thus extended to

$$y_i^\tau = x_i(\tau)\beta + \varepsilon_i^\tau + \varepsilon_k^\tau. \tag{2}$$

The variance in the loan-specific shock is denoted σ_i^2, whereas σ_k^2 represents the variance in the cluster-specific shocks for the k:th cluster, where the latter variance will be zero if there is no dependency between loans within the cluster.

The empirical identification of σ_k^2 can make use of two dimensions. The first dimension, which is applicable if data are only cross-sectional, is the variation in ε_k^τ for the K clusters in the portfolio. This identification strategy works only if one assumes equal cluster variances for all

[7] As in other empirical work on portfolio credit losses, we do not model the loan loss given default (LGD), s_i, because of a lack of data on recovery rates. Acharya, Barath and Srinivasan [2] is one of the few papers that studies the determinants of recovery rates (1-LGD) - for a sample of corporate bonds. It does not model default behavior, however. Whether loan-specific probabilities of default, π_i and LGD are governed by the same factors, and thus should be modeled simultaneously, is an emprical question that remains to be answered.

clusters, and hence for cross-sectional data one can only hope to identify σ_k^2 when it is assumed (by necessity) to be the same for all clusters. The second dimension is the variation in ε_k^τ over time, which of course requires longitudinal data, i.e. loans are followed over time. In this case σ_k^2 is identifiable for each cluster and one can allow for unequal strength in the cluster dependence. This property makes using longitudinal data a superior means of identification. Given the panel nature of our data set, we assume the availability of such data below.

If a common shock hits a cluster of loans, there are good reasons to suspect that the shock may persist over time. For instance, a shock might hit a cluster of steel-producing companies because of an unanticipated increase in energy taxes. It will take the companies some time to adjust to this shock and during this time period they will be more vulnerable than they were before the increase. Such a pattern can be expected to show up by means of serially correlated errors and it may therefore be appropriate to consider an autoregressive process for the cluster-specific errors. For example, in the empirical application that follows we are expecting a positive first-order auto-correlation. Introducing an AR(1) error structure into the model is relatively easy to do. If one extends the model by expanding (2) as

$$y_i^\tau = x_i(\tau)\beta + \varepsilon_i^\tau + \varepsilon_k^\tau + \delta_k \varepsilon_k^{\tau-1}, \tag{3}$$

then no conceptual problem arise as long as $|\delta| < 1$.

If ε_i^τ follows the normal distribution, then the dependence between two loans in the same cluster can be expressed in terms of a correlation measure, $\rho_k = \left[\sigma_k^2 / \left(\sigma_k^2 + \sigma_i^2\right)\right]$. The specification of the model in equation (2) and the definition of ρ_k above make it clear that only positive dependence is considered. This is not a necessary restriction and the estimation procedure we discuss below and apply in the empirical example can be used to estimate models with negative types of interdependence. There are reasons why the dependence between firms can be negative. For instance, if the failure of a company within the cluster makes the competition among surviving companies less harsh, then it is sensible to expect a negative dependence. It is likely that such an effect would show up in the data with some delay and the negative dependence would thus be captured by a negative autocorrelation (i.e. $\delta_k < 0$).[8]

We consider it counterintuitive and unlikely that a cluster specific shock could lead to negative correlation (i.e. $\rho_k < 0$) within the cluster, if well defined. However, if the data has a longitudinal character and the time-span between observations is long, for example annual, then pooling events from different months or quarters may possibly raise the issue of negative correlation. For the type of applications we work here, i.e. credit risk modelling, we believe this will

[8] A study of credit default swaps and stock markets by Jorion and Zhang [33] concludes that the sign of the correlation depends, among other things, on the type of credit event one studies. They find dominant contagion effects (positive correlations) for Chapter 11 bankruptcies but competition effects (negative correlations) for Chapter 7 bankruptcies.

not occur. Moreover, based on simulations we did, we established that the estimated correlation will be zero if the actual dependence is negative and the "misspecified" model in equation (2) is estimated. Hence, because a zero correlation estimate may imply either independence or a negative dependence within a cluster, we recommend that the specification be modified to allow for negative dependence whenever a zero correlation is found, in particular when the data has a low frequency.

2.3 Estimation

For the model in equation (2) estimation of the parameter vector β and σ_k^2 is standard practice in econometrics if the response variable is continuous and the (iterative) Generalized Least Squares procedure can be used (see e.g. Greene [26]). Unfortunately, when studying default risk, one only observes the events that a loan or firm defaults or survives, and never the continuous index y_i^τ. Therefore, we need threshold model for which we define an indicator variable D_i^τ that takes on unity if default occurs for the i:th loan at τ and equals zero otherwise:

$$D_i^\tau = \begin{cases} 1, & \text{if } y_i^\tau \geq 0 \\ 0, & \text{if } y_i^\tau < 0 \end{cases} \tag{4}$$

If the error is assumed to follow the standard normal distribution, then equations (1) - (4) describe a simple probit model. The predicted probability of default, $\widehat{\pi}_i^\tau = \Pr\left[D_i^\tau = 1 \mid x_i(\tau)\right]$, is $\Phi\left[x_i(\tau)\beta\right]$ for the probit model, where Φ denotes the cumulative standard normal distribution, i.e. the link-function for the probit model. If the non-systematic part is assumed to follow the logistic distribution then $\widehat{\pi}_i^\tau = \exp\left[x_i(\tau)\beta\right]\left(1+\exp\left[x_i(\tau)\beta\right]\right)^{-1}$. An unattractive feature of these specifications is the fact that they do not easily transform into a hazard rate. In fact, the shape of the hazard rate depends on the systematic part. This makes it hard to empirically examine the presence of duration dependence, i.e. whether the time a loan or firm has survived affects the probability of default.

If the unsystematic part is assumed to follow the extreme value distribution, then the specification (1) - (4) is equivalent to Cox's proportional hazard model (Narendranathan and Stewart [42]). In this case, the predicted probability of default is obtained from the threshold model as

$$\widehat{\pi}_i^\tau = 1 - \exp\left[-\exp\left(x_i(\tau)\beta + \gamma_t\right)\right] \tag{5}$$

where γ_t measures the effect of a $t-1$ duration of the loan without defaulting. In the estimation of the model, in addition to assuming some distribution for the error term, one also needs to assume that the random-effect component (i.e. the cluster-specific shock) follows the normal distribution. The validity of this normality assumption can, however, be examined (see Figure 3).

The threshold model turns the specification into a non-linear one that makes standard econometric procedures for estimating linear random effects models inadequate. Much progress has been made, however, in the development of statistical procedures for the estimation of these types of models. Probit models with random effects have been widely used in econometrics. See Heckman and Willis [29] for an early example and Guilkey and Murphy [27] for details. This line of research has mainly considered panel data with a small number of repeated measurements on the same subject, often less than 10. The parameters in such models with small cluster sizes can be estimated by maximizing the likelihood of the multivariate normal distribution using numerical integration (See Butler and Moffitt [13] and Greene [26]). Because numerical integration has a pivotal role in this approach, the methods are badly suited for problems where the size of the clusters is large and application is unfortunately infeasible for the data we work with.[9]

Another approach, that is more useful in our setting has been developed in the work on Generalized Linear Models (GLM). McCullagh and Nelder [39] contains an introduction. The random-effect probit model as well as the model in (5) is a special case of the Generalized Linear Mixed Models (GLMM) and can thus be estimated with such techniques. The general idea in GLM is to linearize the non-linear models such that conventional least-squares estimation can be conducted. The penalized quasi-likelihood (PQL) technique that we use first takes the first order Taylor expansion of the non-linear function, i.e. $\ln\left(\ln\left(1-\pi_i^\tau\right)\right)$, and then regresses it on $x_i\left(\tau\right)$ with a procedure similar to (iterative) Generalized Least Squares. Details of the procedure are given in Zhou, Perkins and Hui [54].[10]

The iterative scheme of the PQL-approach makes the properties of the estimator inherently dependent on the actual computer code in use. But it is well-known that the approach suffers from small-sample bias while an extensive simulation study of the estimator in various program packages, which would contribute greatly to the understanding of the estimator's properties, is still lacking. Nevertheless, Zhou, Perkins and Hui [54] contains an extensive discussion of the pros and cons of various software packages' implementation of PQL. Moreover, at least to our knowledge, PQL has not yet been rigorously proved to give consistent estimators. But, estimation of parameters in a GLM (i.e. without random effects) is consistent (McCullagh and Nelder [39]). As estimation of linear models with random effects is equivalent to estimation of fixed effects when the number of observations goes to infinity[11], and thus the PQL estimators ought to be consistent.

To show how the model can be applied in practice, we investigate in the next section whether

[9] The problem of numerical integration is overcome in Bayesian approach by use of importance sampling or Gibbs sampling techniques, and Bayesian methods thus present an alternative (Breslow and Clayton [12]).

[10] In the actual implementation we have choosen to use the GLIMMIX macro (Wolfinger and O'Connell [53]) that is provided by SAS. In this macro, one can implement different link functions including the probit and complementary log-log, which is the relevant in the case of the model in (5). See for example Olsson [43].

[11] In a mixed linear model, the fixed and random effects are estimated using iterative GLS (Greene [26]). The estimation of the random effects is corrected for the uncertainty caused by a finite number of observations within each cluster in the equations of the iterative GLS. When the number of observations within clusters goes to infinity the correction factor is flooded and the estimations are equivalent to those obtained from a fixed effects model.

a pattern of cluster correlation can be found in the non-systematic variation of default probabilities, that has been disregarded in other, earlier, models of bank loan default.

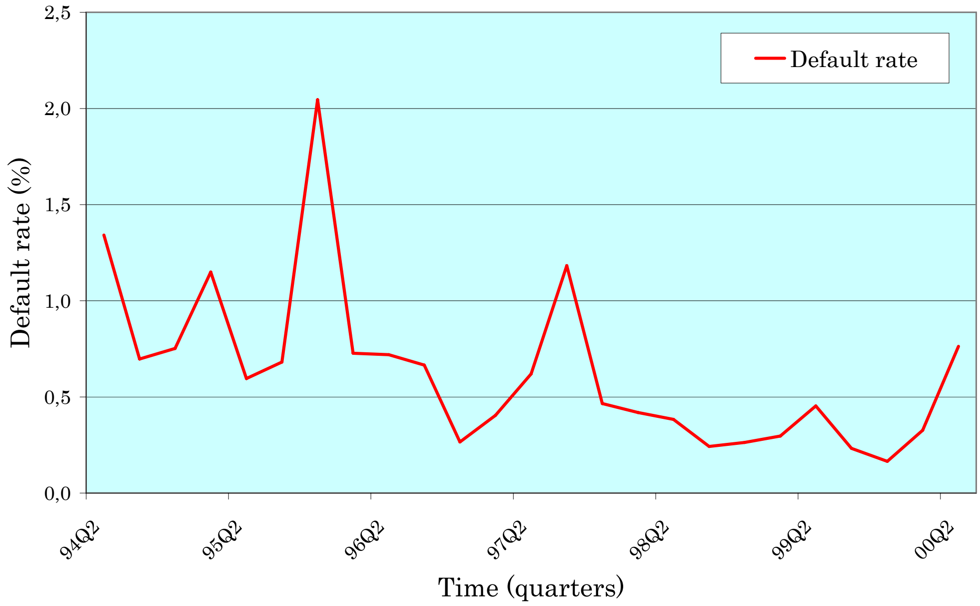
3 Data

In this paper we analyze loan survival and firm data from two Swedish banks. Each bank has a nationwide branch network serving both private and business customers and none of them has any well-known specialization profile. The data set has been examined before in Carling et al. [17], that use the data from bank A, and Jacobson et al. [31] that employ data from bank B. In order to focus on the modeling of the error terms, we will formulate the systematic part in the same way as in Carling et al [17] and abstract from issues like choosing the explanatory variables.

For the first bank, bank A say, we have detailed loan information for all 54,603 *aktiebolag,* the approximate Swedish equivalent of US corporations and UK limited businesses, that had one or more loans outstanding at the bank on the last day of at least one quarter between March 31, 1994 and March 31, 2000. For the second bank, bank B say, we have quite similar information for 46,323 companies. However the data comes from a slightly different time-period, between Jan 1, 1996 and June 30, 2000. Although we have annual report data on small firms such as general partnerships, limited partnerships and sole proprietors, these will be disregarded because we could not dispose of the relevant credit histories. However, a large part of the sample still consists of relatively small enterprises: respectively 65% and 53% of the banks' observations concern businesses with 5 or fewer employees. The average (censored) duration of a counterpart's presence in a bank's portfolio is almost identical in both banks: 8.6 quarters in bank A versus 8.7 in B. The average exposure is, however, over 50% larger in B: SEK 11.4 mn., compared with SEK 6.8 mn. in A. The median credit line in bank A has a size between SEK 250k and SEK 500k, while bank B has a median credit facility between SEK 1 mn. and SEK 2.5 mn. Bank B generally has a bigger share of its counterparts in industries with bigger credit lines, such as real estate, energy & water, and forestry & paper.

Figure 1 shows the quarterly default rates among the counterparts of banks A and B over the time-period 1994Q2 - 2000Q2. In our data a defaults is defined to take place whenever (i) payments on the principal or interest are at least 60 days overdue and (ii) a bank official has made a judgement and concluded that any such payment is unlikely to occur in the future. The general trend over time is one of a decreasing default rate, a pattern that is largely consistent with the overall improving state of the Swedish economy. Throughout the 1990s, the Swedish economy was slowly recovering from the deep recession that struck the country in 1991, with two notable downturns in economy activity, in 1995 and 1997. At the end of the sample period, we can observe the beginning of the persistent slump in the new millennium.

Figure 1: Quarterly default rates among counterparts in banks' portfolios during 1994Q2-2000Q2
A defaults is defined as the event where (i) payments on the principal or interest are at least 60 days overdue and (ii) a bank official has made a judgement and concluded that any such payment is unlikely to occur in the future.



The annual accounting data we use, is obtained by Upplysningscentralen (UC), the main credit bureau in Sweden, from the National Patent Office, Patent och Registreringsverket (PRV), to which firms are required to submit their annual report, and includes all typical balance sheet and income statement data, such as inventories, short and long term debt, total assets and a whole range of earnings variables. Payment remarks data are reported by banks and other businesses and stored by UC and comprise information on the events related to the remarks and payment behavior for both the company and its principals. The data provided by UC was available at different frequencies, varying from daily for payment remarks to annually for accounting data. We will discuss the specifics of both data sources in greater detail below.

The banks have provided us with the complete credit history and the unique, government provided, company identification number of each company. From this data we have constructed a number of credit variables, such as the size of the loan, actual exposure, and loan types. The second part of the data has been supplied by UC and contains information on most standard balance sheet and income statement variables. Some examples of balance sheet entries are cash, accounts receivable and payable, current assets and liabilities, fixed and total assets, total liabilities and total equity. Some examples of the income statement entries that were available are total turnover, earnings before interest, depreciation and amortization, depreciation, financial income, extraordinary income, and taxes. By means of the company identification number, we have been able to match the banks' data with UC's database. In addition to these annual report entries that are collected by the Swedish Patent and Registration Office and stored by UC, we have information on the companies' track records regarding payment behavior as summarized through remarks for 61 different credit and tax related events.[12] The third group of data consists of macro-variables. Earlier work [**?**] has shown that the output gap, the yield curve and consumers' expectations about the economy are important variables.

[12] Broadly, remarks belong to one of two categories: non-payment remarks or bank remarks. Storage and usage of the first group are regulated by the Credit Information Act, the Personal Data Act and overseen by the Swedish Data Inspection Board. Examples of events that are registered are: delays in tax payments, the repossession of delivered goods, seizure of property, resettlement of loans and actual bankruptcy. In practice, with a record of non-payment remarks individuals will not be granted any new loans and small businesses will find it very difficult to open new lines of credit. The second type of remarks provide information on firms' payment behavior at banks. All Swedish banks participate in this scheme and report any abuse of a bank account or a credit card and slow loans (repayment is considered questionable) to the credit bureau, that maintains these records. Storage and usage of these remarks is regulated merely by the Personal Data Act. Whereas a bank remark may have the same consequences as having a non-payment remark, this is not the case in general. Their effect on individual applications for credit presumably works mainly through the accumulation of negative indicators.

Table 1: Descriptive statistics for variables used in the model.
Table shows some selected descriptive statistics for merged data from bank A and B. All variables are obatained from the credit bureau. Statistics are split up for defaulted and (censored) surviving counterparts. All original variables are in nominal terms and in million SEK. Ratios and the remark variables are expressed as fractions.

	Defaulted				Surviving			
Variable	Mean	StDev	Min	Max	Mean	StDev	Min	Max
Bankdata								
Bank A	0.69	0.46	0	1	0.61	0.49	0	1
Short-term credit	0.65				0.42			
Long-term credit	0.14				0.23			
Mixed credits	0.21				0.35			
Credit bureau data								
Sales (log)	13.94	3.32	0	24.20	14.79	2.87	0	25.10
Sales missing	0.14	0.34	0	1	0.02	0.15	0	1
Earnings/Assets	-0.29	0.80	-2	2	0.06	0.42	-2	2
Inventory/Sales	0.12	0.15	0	0.50	0.10	0.13	0	0.50
Loans/Assets	0.86	0.52	0	2	0.73	0.31	0	2
Remarks 8,11,16,25	0.23	0.42	0	1	0.01	0.10	0	1
Remark 25	0.12	0.33	0	1	0.00	0.04	0	1
Accounting data								
Complete	0.50				0.96			
Reported previously	0.42				0.02			
Reported afterwards	0.00				0.01			
Not reported	0.08				0.02			
Macrodata								
Output gap	-4.90	1.89	-7.78	-0.50	-4.20	2.04	-7.78	-0.50
Yield curve	1.21	0.81	-0.41	2.80	1.25	0.74	-0.41	2.80
Nobs	5196				945497			

Notes: Earnings are before interest, taxes, depreciation and amortization. Remarks indicate that the firm had a payment remark due to one or more of the following events in the preceding four quarters: a bankruptcy petition, issuance of a court order to pay a debt, seizure of property. Earnings/Assets are truncated at -2 and 2, any value below and above these limits is set to the limit. Truncation was also done for Inventory/Sales and Loans/Assets with limits given by Min/Max in the table.

Table 1 contains descriptive statistics for the variables that enter the credit-risk model and provides us with some further insight into the counterparts of both banks. This simple univariate analysis is worth some comments. The type of credit seems to matter for default behavior as short-term credits, as opposed to exposure with longer maturities, make up a larger share of defaulted loans than of surviving loans. As regards the accounting data, we focus on four financial ratios, that are commonly used to study bankruptcy risk [13]. One may note some differences between defaulting and surviving counterparts. On average, surviving firms have lower average debt ratios, i.e. smaller debts relative to their assets, and higher earnings. Defaulting firms are persistently worse off than healthy businesses when measured by the earnings ratio, while their inventory turnover is higher. Total sales are also generally, but not always, lower for defaulting

[13] Examples of authors that employ financial ratios are Altman [5] [6][7], Frydman, Altman and Kao [21], Li [36], Shumway [48] and Carling, Jacobson, Lindé and Roszbach [?].

firms. Another interesting difference is that we do not have complete accounting data for a large fraction all defaulting companies, due to their failure to submit an annual report: about half of them had either never supplied a report or done so only for a preceding year. Payment remarks are also much more common amongst the defaulting companies.

For our specific empirical application, the way in which the clusters are identified is an important step. We use the Statistics Sweden's (SCB) Standard Industrial Classification (SNI) [47]. The SNI classification is based on the EU's recommended classification standard NACE (Nomenclature Generale des Activités Economiques dans les Communautes Européennes). SNI is primary an activity classification. Production units as companies and local units are classified after the activity which is carried out. The SNI system allows one company or a local unit to have several activities (SNI-codes). In the data base available to us, however, only the most important SNI-code was stored. There are more than a thousand sectors, although the coding is organized in such a manner that it is easy to see how similar the activity of two companies are. We have collapsed the sectors into seven main clusters: public companies, wholesale, retail, transport and communication, manufacturing and construction, forest and other industries, and real estate and finance.[14] The exact mapping is displayed in the table in the Appendix.

For further details about the data we refer to Carling et al. [17] or Jacobson et al. [31].

4 Applying the model

In this section we discuss how the duration model with cluster specific errors is implemented in practice. First, in Section 4.1 we estimate a model of business default or loan default. Next, in Section 4.2 we apply the default risk model to generate credit loss distributions and present some relevant VaR percentiles. We also discuss how adding the cluster error structure affects our estimates of credit losses.

4.1 Estimating a duration model of default risk with cluster specific errors

We begin with a short discussion of the systematic part of the default risk model. Because estimation of the full model in equation (3) did not reveal any significant autoregression in the

[14]We also considered splitting the seven clusters into 14 subclusters. This gave slightly greater correlations, but essentially no difference is the VaR-measure. On the one hand, a too broad classification of the cluster might lead to the interdependency of some companies within a cluster being obscured by the fact that no relation exists with other companies in the broader cluster. On the other hand, too many clusters will lead to a single cluster having very little effect on the VaR-measure. Our choice of seven clusters seemed to adequatelly balance the two conflicting interests.

Table 2. Estimates of the systematic part of the default risk duration model.
Model 1 is estimated with cluster specific errors, model 2 with a common error for all clusters. Correlations may differ for clusters but no auto-correlation is assumed. The link function is the complementary log-log.

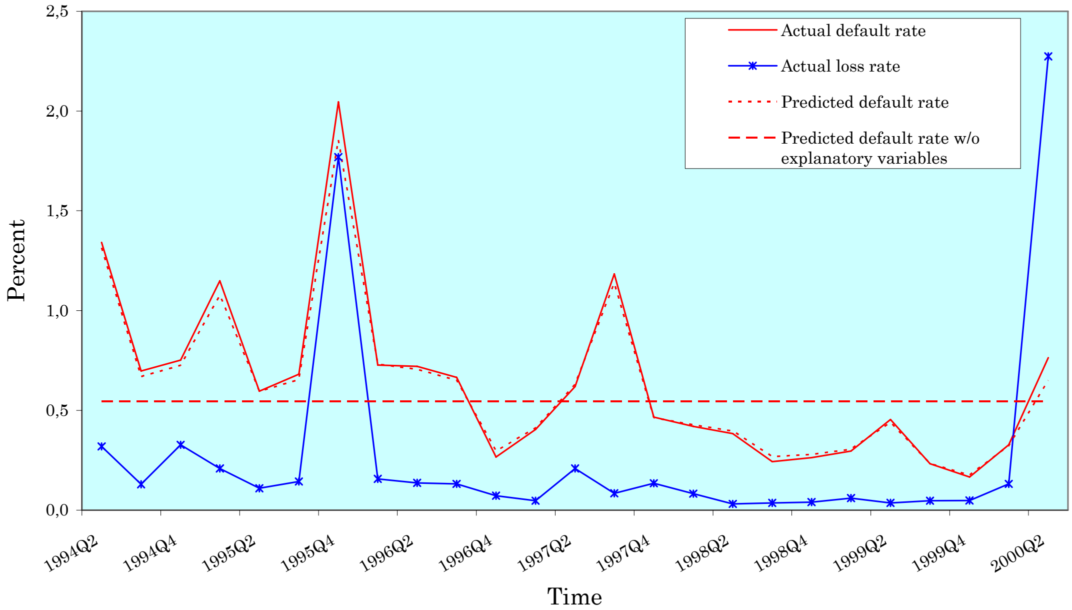
	Model 1		Model 2	
Variable	Estimate	StdError	Estimate	StdError
Bankdata				
Bank A	-0.111	0.037	-0.191	0.034
Short-term credit	**0.514**	0.037	**0.488**	0.038
Long-term credit	0.000	-	0.000	-
Mixed credits	**-0.295**	0.048	**-0.289**	0.049
Credit bureau data				
Sales (log)	**-0.033**	0.005	**-0.033**	0.005
Sales data missing	**0.765**	0.112	**0.788**	0.111
Earnings/Assets	**-0.207**	0.036	**-0.200**	0.037
Inventory/Sales	**0.545**	0.097	**0.526**	0.098
Loans/Assets	**0.896**	0.038	**0.911**	0.039
Remarks 8,11,16,25	**0.960**	0.049	**0.964**	0.051
Remark 25	**0.826**	0.064	**0.786**	0.064
Accounting data				
Complete	**-2.624**	0.082	**-2.538**	0.083
Reported previously	**0.595**	0.079	**0.637**	0.080
Not reported	**-3.547**	0.234	**-3.449**	0.241
Partially missing	0.000	-	0.000	-
Macro data				
Output gap	**-0.170**	0.022	**-0.187**	0.010
Yield curve	**-0.216**	0.054	**-0.246**	0.020
Duration				
1:st year	0.000	0.162	0.000	-
2:nd year	0.231	0.128	0.332	0.128
3:rd year	**0.327**	0.129	**0.141**	0.131
4:th year	**0.465**	0.127	**0.516**	0.128
5:th year	**0.479**	0.138	**0.393**	0.129
> 5 years	0.176	0.146	0.287	0.169
Intercept	**-4.290**	0.197	**-4.250**	0.170

Bold style coefficients are at least significant at the 5% level.

errors, we report only parameter estimates for the specification without such an autoregressive component. The first and second column of Table 2 contain the parameter estimates and standard errors for the model in (2), with cluster specific errors, while the third and fourth column display the parameters and their standard errors for the "conventional" model, with a common error for all clusters, from equation (1). A comparison of the significant parameters in columns 1 and 3 clearly shows that adding the industry specific common error structure only marginally

Figure 2: Default rates (predicted vs. actual) and average loss rate.

The figure compares actual default and loss rates during the sample period with the default rate implied by our preferred model and a model without explananatory variables



affects the size of the coefficients and standard errors for the systematic part - as one would anticipate.[15] The signs of all parameters are in line with what one would expect from the descriptive statistics in Table 1. Short-term credits, any sort of incomplete provision of accounting data, having any kind of "payment" remark, higher debt ratios and large inventories relative to turnover are all associated with a higher risk of default. Conversely, longer maturities, higher sales, and higher earnings relative to total assets reduce the likelihood of default. As one would expect favorable economic conditions, like a positive output gap and a steeper slope of the yield curve, reflecting an anticipated strengthening of the economy by the market, reduce counterparty risk. The increasing size and the significance level of the duration coefficients also suggest there is some support for the existence of a positive duration dependence, that is: of default risk increasing with the age of the loan.[16]

Before starting with a discussion of the cluster variances, it is useful to inspect the basic

[15] We also found that our parameter estimates from our complementary log-log specification were very similar to those in a probit specification. The correlation of the predicted probability derived from the two specifications was found to be 0.989.

[16] Because we only have information on the age of exposures within the limits of the sample period, and not for the period preceding the sample period, this result should be interpreted with care.

Table 3. Definition of clusters and their distribution in the data.
Estimation of the cluster-variance. The second column with parameter estimates refers to our preferred specification and therefore 95% confidence intervals are presented for the parameters.

	Equally-sized cluster	Cluster specific errors		
Variable	error w. systematic part	with systematic part	only macro variables	w/o systematic part
Public or subsidized sector	0.254	0.289 (0.169--0.606)	0.919	1.101
Wholesale	" - "	0.227 (0.122--0.570)	0.269	0.225
Retail	" - "	0.292 (0.171--0.612)	0.309	0.421
Transport and communication	" - "	0.217 (0.127--0.457)	0.280	0.394
Manufacturing	" - "	0.111 (0.051--0.402)	0.123	0.254
Construction, forest and other industries	" - "	0.213 (0.119--0.487)	0.174	0.254
Estate and finance	" - "	0.411 (0.230--0.935)	0.400	0.613

Note: To convert the cluster-variance estimates to estimates of correlations, one takes the ratio of the cluster-variance and 1.645 (i.e. the residual-variance for the complementary log-log specification) plus the cluster-variance.

properties of the estimated model. For this purpose, we have plotted the actual default rate against the default rate predicted by model 1 in Figure 2. For reference, we have also added the average default rate and the actual loss rate over the sample period. By visual inspection we can see that the predicted default rate, shown as a broken line, closely tracks the actual average rate of firm default; the R^2 of a simple regression is 0.995. Disregarding for the meanwhile the impact of loan size on portfolio credit losses, our preferred model is thus well suited to replicate the portfolio's default behavior. As the bold line in Figure 2 illustrates, the default rate and the loss rate do broadly follow the same pattern, but idiosyncratic variation in the loan size does tend to create short-lived discrepancies between them. For this reason, we will discuss both portfolio default rate - and portfolio loss rate distributions in Section 4.2. In Table 3, we report the estimated cluster specific error variances and compare them with the error variances (i) from a model specification that assumes equally-sized cluster variances, and from "cluster" models that contain (ii) only macro explanatory variables or (ii) no systematic factors at all. As the second column shows, all industries ("manufacturing" being an exception) have error variances that differ significantly from zero. In "wholesale", "transport and communication" and "construction, forest and other industries" non-systematic, cluster specific changes in business default risk are of comparable size. Somewhat bigger cluster specific shocks to default risk are experienced in the retail industry and the "public and subsidized" sector. The industry with by far the largest cluster specific error variance is the "real estate and finance" cluster, with a variance of 0.411. Since the variance of the unsystematic part is fixed at $\sigma_i^2 = 1.645$ for our choice of specification in eq. (5), the implied correlation is 0.2. Column 1 shows that the estimated variance if equality across clusters is assumed, is 0.254. However, a likelihood-ratio test easily rejects the restriction of a common correlation across all clusters is strongly rejected: $LR = 122.7 > \alpha_{0.05}\left(\chi^2_{(6)}\right)$. The clusters should be thus be examined separately, and not as one group.

Although we have no formal explanation of the differences between clusters, there may be some heuristic explanations. The financial sector and real estate sector, were exposed to fundamental changes during the 1990s, after financial markets had been deregulated during the late 1980's and early 1990's. In the early 1990's Sweden also experiences a banking crisis, the effects of which may have been dissipated only slowly. The deep recession that struck Sweden more or less simultaneously also had a big impact on the government budget and its contributions to various (semi-)governmental organizations.

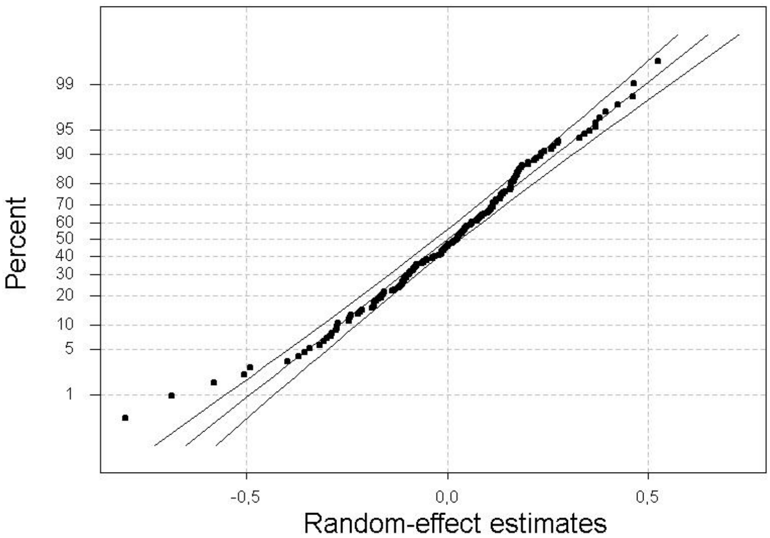
Although there are good theoretical reasons for the inclusion of close business relations within certain industries into models of default or credit risk [54][22], economic theory does not help us much in forming a prior for the size of the industry specific error variances. We presume that an (unconditional) correlation of 0.5 - the equivalent to a cluster variance of 1.645 - or even higher are plausible and that most of such correlation will be related to variation in macroeconomic variables.

When we compare the three models, the most substantial increase in cluster variances comes about when the systematic part is excluded completely (column 4) - although the exact magnitude of the effect varies somewhat between industries. The industry errors now capture part of the variation in the default rate that macro - and firm specific covariates can no longer explain. For obvious reasons, we would like to know if and to what extent this result is solely a consequence of a common impact that the macro-variables have on companies. To cast more light on this issue, we re-estimated the model and included the macro variables while keeping out the other explanatory variables. In the third column of Table 3, the estimates of the cluster variances with this specification are displayed. Once again, the effect varies between sectors, with the macro variables being able to explain more of the cluster variances for the "government" sector and "real estate and finance". Overall, including macro variables and excluding firm specific covariates reduces the estimated variances somewhat, but does not bring them back to the level of the model with a full systematic component. Hence, we find that a substantial part of the difference in the conditional and the unconditional correlations should be attributed to the firm specific variables in the systematic part.

In the estimation procedure, it was assumed that the random-effect component, i.e. ε_k^{τ}, follows the normal distribution with expectation equal to zero. To check the validity of this assumption, we re-scaled the estimated random-effects to let them have a unit standard deviation by dividing the estimates of the 175 random-effects components by the relevant cluster variance. Under the above mentioned assumption the standardized estimates should in fact fol-

Figure 3: A QQ plot for the estimated random effect components.

The figure compares the empirical distribution of estimated random-effect components to the normal distrbution. The scale of the vertical axis is determined by the dispersion of observations for the normal distribution The actual empirical distribution is depicted by dots. If the empirical distribution were to follow the normal curve exactly, the dots would overlap with the straight line in the plot. The concave/convex lines provide 95 % confidence limits.



low the standard Normal distribution. Figure 3 shows, by means of a Q-Q-plot, to what extent the distribution of the standardized estimates match the percentiles of the standard Normal distribution. If the correspondence between these two is perfect, the points will be placed along the straight line that lies in between the accompanying lines that reflect the upper and lower 95% confidence limits. The Q-Q-plot suggests that normality assumption is quite acceptable.

From the discussion in Section 2 we recall that the cluster specific shocks may well be persistent and that one thus may observe patterns of autocorrelation, for example an AR(1) structure as outlined in eq. (3). Therefore, we also used the estimated random-effects to check for the presence of any autoregressive structure in the correlations, by examining the autocorrelation function for the series of 25 observations we have available for each cluster. The strongest (and in fact only) sign of such a pattern was found for the "retail" industry, that has a first-order autocorrelation of about 0.25. However, even this estimate was insignificant and we thus conclude that no autoregressive structure is present in this data.[17]

4.2 Estimating portfolio credit risk distributions and VaR

Having estimated a duration model of default risk with cluster specific effects, we now turn to applying this model to the estimation of a portfolio credit loss distribution. In order to get a better understanding of the extent to which our modeling approach leads to improved estimates of credit loss distributions, we will compare the Value-at-Risk (VaR) estimates, the statistics we use to summarize the loss distribution, with those derived from two simple benchmark models.

We calculate VaR based on the portfolio at τ for a one quarter ahead horizon by mean of the following algorithm:

1. Let $\widehat{\beta}$ and $\widehat{\sigma}_{\varepsilon_k}$ be consistent estimates of β and σ_{ε_k} in the model in equation (2).[18]

2. Draw ε_k^τ from the normal distribution with expectation equal to zero and variance equal to $\widehat{\sigma}_k^2$

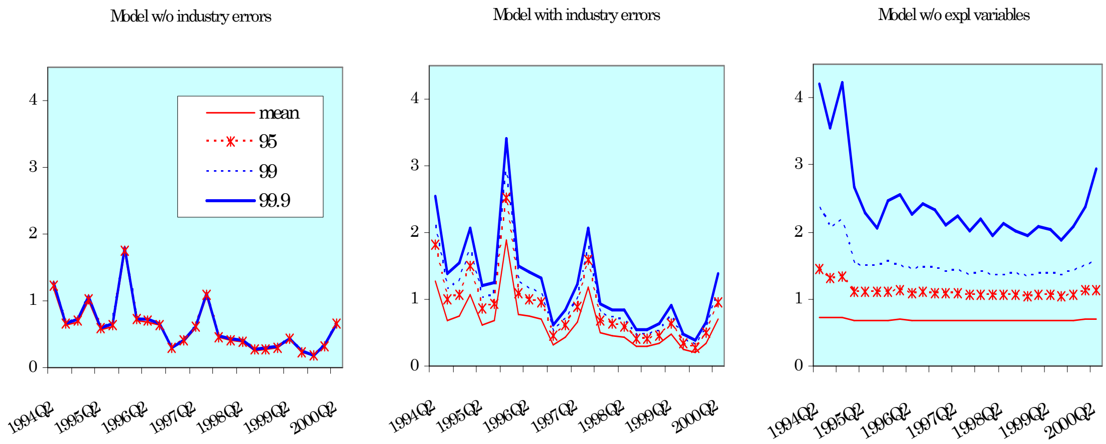
3. Draw ε_i^τ from either the standard normal distribution or the extreme-value distribution (depending on the estimated model)

[17] This approach does not estimate the autoregressive part jointly with other parameters in the model. The GLIMMIX macro in SAS provides this option but the implementation failed due to lack of memory. The procedure of estimating the $AR(1)$-part, conditional on the estimates for β, σ_i^2 and σ_k^2, gave no indication of an auto-regressive structure in the data. Most likely, a joint estimation procedure would have produced similar results, and thus we did not pursue the effort to solve this technical problem.

[18] Here we assume that the model has been estimated without an autoregressive part. In the case of unequal variance it is necessary to adjust (i) such that the draws from the normal-distribution is taken with the appropriate standard deviation for the cluster. If there is an autoregressive part in the model, then the draws in (i) should be taken from the corresponding autoregressive process.

Figure 4: Percentiles of the default rate distributions implied by three models.

The boxes display the mean and three percentiles from the credit loss distributions that have been generated using three different models of default risk. In Box 1, the default rate distribution has been generated with a conventional duration model. A cluster specific error structure has been added to arrive at the model used in Box 2 . In Box 3 all explanatory variables have been deleted from the cluster error model.



4. Repeat steps (2) and (3) for $k = 1, ..., K$

5. Calculate $\widehat{\pi}_i^\tau = \Phi\left[x_i(\tau)\widehat{\beta} + \varepsilon_i^\tau + \varepsilon_k^\tau[i \in k] + \gamma_t\right]$ or $\widehat{\pi}_i^\tau = 1 - \exp\left[-\exp\left(x_i(\tau)\widehat{\beta} + \varepsilon_i^\tau + \varepsilon_k^\tau[i \in k] + \gamma_t\right)\right]$ (depending on the estimated model)

6. Let $Loss^\tau = \sum_{i=1}^{n^\tau} \widehat{\pi}_i^\tau s_i^\tau / \sum_{i=1}^{n^\tau} s_i^\tau$, where s_i^τ is the size of loan i at time τ and n^τ is the number of loans in the loan portfolio at time τ.

7. Repeat (2)-(5) R times and order the R observations of $Loss^\tau$ by increasing size, where R should be large enough to guarantee that the loss distribution has converged.

8. Let VaR_z^τ equal the z-percentile in the distribution of $Loss^\tau$

Because our statistical model does not incorporate a description of (any possible relation between default risk and) the loan size or recovery rates, a comparison of *loss* distributions will not strictly give us an evaluation of the suitability of our statistical model for generating portfolio *credit loss* distributions. Sample specific, non-stochastic variation in the loan size may add noise to the portfolio *default rate* distribution generated by the model. For this reason, we start out by studying how the distribution of quarterly *default* rates is affected by using model (2) instead of a conventional specification. We can do this by following the procedure described in steps (1)-(7) above, but replacing the $Loss^\tau$ rate by a $defaultrate^\tau = \sum_{i=1}^{n^\tau} \widehat{\pi}_i^\tau$ in step (5). This will give us a purer picture of the merits of the model specification. Having done this, we

Table 4: The effect of including industry specific errors on the estimation of default risk distributions.
Table shows the mean and various percentiles of the portfolio default risk distribution that is generated by (i) the model with industry specific common errors, and (ii) the model without common industry specific errors. Normally styled figures refer to the model with industry specific errors, italics to the model witout industry errors.

		Simulated portfolio default rates at distribution percentiles							
Quarter	mean	*mean*	95	*95*	97.5	*97.5*	99.9	*99.9*
1994Q2	1,28	*1,19*	1,81	*1,23*	1,95	*1,23*	2,55	*1,27*
1994Q3	0,69	*0,63*	0,99	*0,65*	1,06	*0,66*	1,39	*0,67*
1994Q4	0,74	*0,68*	1,07	*0,70*	1,16	*0,71*	1,54	*0,73*
1995Q1	1,07	*0,99*	1,50	*1,03*	1,61	*1,03*	2,07	*1,06*
1995Q2	0,62	*0,56*	0,87	*0,58*	0,94	*0,59*	1,21	*0,60*
1995Q3	0,68	*0,62*	0,93	*0,64*	0,99	*0,65*	1,24	*0,66*
1995Q4	1,88	*1,72*	2,53	*1,76*	2,71	*1,77*	3,41	*1,79*
1996Q1	0,78	*0,70*	1,08	*0,72*	1,16	*0,72*	1,49	*0,73*
1996Q2	0,74	*0,68*	1,01	*0,70*	1,08	*0,70*	1,41	*0,72*
1996Q3	0,69	*0,63*	0,95	*0,64*	1,02	*0,64*	1,31	*0,65*
1996Q4	0,32	*0,29*	0,45	*0,30*	0,48	*0,30*	0,62	*0,30*
1997Q1	0,44	*0,39*	0,61	*0,41*	0,65	*0,41*	0,85	*0,42*
1997Q2	0,65	*0,59*	0,88	*0,61*	0,94	*0,61*	1,22	*0,62*
1997Q3	1,17	*1,07*	1,58	*1,09*	1,68	*1,10*	2,06	*1,11*
1997Q4	0,50	*0,45*	0,69	*0,46*	0,74	*0,46*	0,94	*0,47*
1998Q1	0,46	*0,41*	0,63	*0,42*	0,67	*0,43*	0,85	*0,43*
1998Q2	0,43	*0,38*	0,59	*0,39*	0,64	*0,40*	0,84	*0,40*
1998Q3	0,29	*0,26*	0,40	*0,27*	0,43	*0,27*	0,55	*0,27*
1998Q4	0,30	*0,27*	0,41	*0,28*	0,43	*0,28*	0,55	*0,29*
1999Q1	0,33	*0,30*	0,45	*0,30*	0,48	*0,31*	0,64	*0,31*
1999Q2	0,47	*0,42*	0,64	*0,43*	0,68	*0,44*	0,91	*0,44*
1999Q3	0,26	*0,23*	0,35	*0,23*	0,37	*0,24*	0,47	*0,24*
1999Q4	0,19	*0,17*	0,27	*0,18*	0,29	*0,18*	0,38	*0,18*
2000Q1	0,35	*0,31*	0,49	*0,32*	0,52	*0,32*	0,67	*0,33*
2000Q2	0,70	*0,63*	0,96	*0,65*	1,04	*0,65*	1,38	*0,66*

will calculate the credit loss distributions, to complete the illustration of how the model could be used by banks for in risk-management applications and by regulatory authorities for monitoring purposes.

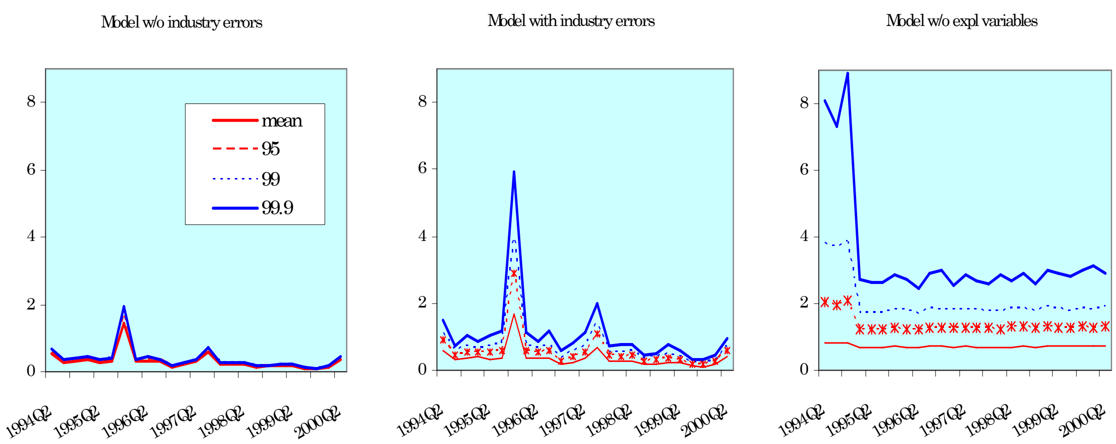
Figure 4 displays, for each of the 25 quarters in the sample period, the mean and three commonly used percentiles from the default rate distribution generated by three different models of firm default risk. Table 4 provides some further numerical details on the default rate distributions.[19]

The first box contains the distribution that results when model (4) is used, the middle box the distribution for our model with cluster errors, and the percentiles in the third box are derived using a cluster error model without any explanatory variables. A first glance at the three boxes reveals both similarities and important differences. Both the first and second model appear

[19] It should be noted that the calculation of VaR presumes that $\sigma_{\varepsilon_k} = \widehat{\sigma}_{\varepsilon_k}$, which of course is susceptible to uncertainty. If $\sigma_{\varepsilon_k} > \widehat{\sigma}_{\varepsilon_k}$, then VaR would be greater than reported.

Figure 5: Percentiles of the loss rate distribution implied by three models.

The boxes display the mean and three percentiles from the credit loss distributions that have been generated using three different models of default risk. In Box 1, the loss distribution has been generated with a conventional duration model. A cluster specific error structure has been added to this model in Box 2. In Box 3 all explanatory variables have been deleted from the cluster error model.



to capture the broad movements in default rates over the sample period. The peaks in 1995 and 1997 are clearly captured and even the general downward trend up to the end of 1999 is identified and the mean default rate is more or less identical for models 1 and 2. The model without industry specific errors has default rate distributions, for each point in time, that are highly concentrated though. The lower and upper percentiles almost coincide, while for the model with industry specific errors, default rate percentiles are at least 50 percent and up to 100 percent larger.

To be able to single out the contribution of the cluster error structure to the shape of the loss and default rate distributions, it is useful to consider how other factors will affect these. Firstly, there's the size of the portfolio: larger portfolios, with more borrowers, will produce loss distributions that tend to converge to the mean and have smaller tail losses. Secondly, the *skewness* of the distribution of loan sizes will influence the location of the tails: portfolios in which the biggest loans are exceptionally large in relation to the remainder, will tend to generate bigger tail losses. Finally, there's the cluster errors, that capture the interdependency between firms within industries. A closer look at the boxes in Figures 4 and 5 makes what their impact is and that they, in this particular case, are the single driving force behind the larger default rates in the tail percentiles of the cluster model (Box 2). To see this, consider that in Figure 4 the single factor causing differences between Boxes 1 and 2 is the addition of the cluster errors. Hence, it is the existence of an interdependence between companies within industries that causes the higher percentiles of the default rate distribution to exceed the mean default rate by up to one and a half percent points. When we instead look at the loss distributions, in Figure 5,

Table 5: The effect of including industry specific errors on the estimation of Value-at-Risk.
Table shows the mean and various percentiles of the loss distribution that is generated by (i) the model with industry specific common errors, and (ii) the model without common industry specific errors. Normally styled figures refer to the model with industry specific errors, italics to the model witout industry errors.

		Simulated portfolio loss rates at loss distribution percentiles						
Quarter	mean	*mean*	95	*95*	97.5	*97.5*	99.9	*99.9*
1994Q2	0,59	*0,54*	0,90	*0,60*	1,00	*0,62*	1,51	*0,67*
1994Q3	0,30	*0,27*	0,47	*0,32*	0,52	*0,32*	0,73	*0,36*
1994Q4	0,34	*0,30*	0,55	*0,35*	0,62	*0,36*	1,04	*0,40*
1995Q1	0,39	*0,35*	0,56	*0,38*	0,60	*0,39*	0,85	*0,43*
1995Q2	0,33	*0,29*	0,54	*0,33*	0,61	*0,33*	1,05	*0,36*
1995Q3	0,38	*0,33*	0,61	*0,37*	0,70	*0,38*	1,18	*0,41*
1995Q4	1,66	*1,44*	2,88	*1,65*	3,32	*1,70*	5,91	*1,93*
1996Q1	0,36	*0,31*	0,57	*0,34*	0,65	*0,35*	1,11	*0,38*
1996Q2	0,37	*0,32*	0,56	*0,37*	0,62	*0,39*	0,87	*0,44*
1996Q3	0,36	*0,31*	0,57	*0,34*	0,65	*0,35*	1,16	*0,38*
1996Q4	0,17	*0,15*	0,27	*0,16*	0,31	*0,17*	0,57	*0,18*
1997Q1	0,25	*0,21*	0,42	*0,23*	0,48	*0,24*	0,81	*0,25*
1997Q2	0,35	*0,30*	0,55	*0,32*	0,63	*0,33*	1,11	*0,35*
1997Q3	0,68	*0,59*	1,10	*0,64*	1,24	*0,66*	2,00	*0,71*
1997Q4	0,28	*0,24*	0,43	*0,26*	0,48	*0,27*	0,73	*0,29*
1998Q1	0,25	*0,22*	0,40	*0,24*	0,45	*0,24*	0,79	*0,26*
1998Q2	0,26	*0,22*	0,44	*0,24*	0,50	*0,25*	0,78	*0,27*
1998Q3	0,16	*0,14*	0,26	*0,15*	0,30	*0,16*	0,47	*0,17*
1998Q4	0,19	*0,16*	0,30	*0,18*	0,33	*0,18*	0,52	*0,20*
1999Q1	0,21	*0,18*	0,37	*0,19*	0,42	*0,20*	0,77	*0,21*
1999Q2	0,22	*0,19*	0,32	*0,20*	0,36	*0,20*	0,57	*0,22*
1999Q3	0,12	*0,10*	0,17	*0,11*	0,19	*0,11*	0,30	*0,12*
1999Q4	0,10	*0,09*	0,16	*0,09*	0,18	*0,09*	0,31	*0,10*
2000Q1	0,17	*0,14*	0,25	*0,16*	0,28	*0,16*	0,47	*0,18*
2000Q2	0,39	*0,34*	0,58	*0,39*	0,65	*0,40*	0,97	*0,45*

taking account of the cluster error structure appears to lead to increases in the VaR percentiles during most quarters that are comparable to those in the default rate distribution in Figure 4 and Table 4. Depending on the percentile and the quarter one considers, the model without cluster errors underestimates tail losses by 50-200 percent. Clearly, the exceptional peak in the loss distribution in the beginning of 1996, that exceeds the jump in the default rate distribution, has to be attributed to fact that, coincidentally, large exposures were being lost at a time when the average rate of default was also high. Table 5 provides numerical details for the boxes in Figure 5.

When we compare Box 2 in Figures 4 and 5 to the model with cluster errors but without explanatory variables in Box 3, one observes clearly that, through time, the difference between expected risk and tail risk in the outer tails increases when the mean default rate rises. Aggregate or "macro" risk, beyond affecting expected default rates, thus appears to be important for portfolio credit risk in the sense that periods with high rates of default among businesses are associated with more than proportional increases in unexpected defaults. The third model is, however, in both its mean and the lower percentiles, much less successful at capturing the trend in the average default rate. Only the 99.9th and, to a lesser extent, the 99th percentile, manage to produce a peak at the start and a slight downward movement over the remainder of the sample period. But the peak at the start is disproportionately high, while other increases in the default rate in the middle of the sample period, especially the one in 1997 are more or less fail to be captured. So although a model without a systemic part can generate peaks and troughs in the relevant range of the VaR percentiles, it is quite uninformative about when a bank should hold a large economic capital stock. In other words: a model with systematic factors but without a cluster error structure is well able to fit the broad trends in mean portfolio default rates and credit losses, but fails to appropriately capture the credit losses in bad times. Typically these are the times when banks are hit by large "unexpected" credit losses.[20] This model thus performs best when it is least needed. A model that accounts for the dependencies within each industry but lacks a systematic part produces quite some variation in the tails, but the shifts in the estimated loss distribution do not match those in the bank's actual loss experience. It would thus force the bank to hold an unnecessarily large economic capital stock for most of the time. Only the model that includes both a systematic part and cluster specific variances manages to follow both the trend in credit losses and produce industry driven fluctuation in losses around that trend. Consequently, the economic capital requirements that are derived from it are larger for periods when times of large "aggregate" disturbances occur and smaller when the economy is doing well. To what extent actual capital buffers will follow this economic requirement will depend on individual banks' "buffer smoothing" policy. Most likely, many banks will avoid too large fluctuations in their capital stock over time.

5 Discussion

Currently available portfolio credit risk models fail to allow for default risk dependency across loans other than through common risk factors. As a consequence they ignore the close ties that usually exist between companies, due to legal, financial and business relations, and tend to lead to clustering of bankruptcies. As a result, many models commonly used by regulators and financial institutions are likely to underestimate credit risk peaks in periods when industry specific shocks, as opposed to firm specific (idiosyncratic) and country wide factors, manifest themselves.

To address this shortcoming, we have attempted to integrate the insights from theoretical

[20] When studying credit losses, authors generally refers to the difference between the credit loss at a specific percentile of the loss distribution minus the mean losses as "unexpected" credit losses.

models of default correlation, like Giesecke [22] and Zhou [55], into a conventional duration model of default risk and the implied portfolio credit risk estimates by explicitly introducing a relation between firms' default risk through industry specific errors. In this paper, we also illustrate how this model of default risk can be operationalized by practitioners by applying it to a pooled data set from two Swedish banks' business loan portfolios over the period 1994-2000.

The results from the application are encouraging. Our estimates reveal substantial differences in the intra-industry inter-firm correlation of defaults. The retail industry and the "public and subsidized" sector display substantially higher cluster variation than most other industries, while the "real estate and finance" business demonstrates by far the largest correlation.While accounting for dependencies within industries does not affect estimates of individual default risk to any large extent, intra-industry dependencies are shown to be quantitatively important for credit risk. Estimates of VaR can increase by 50-200 percent, depending on the point in time and the aggregate economic conditions. Most commonly used credit risk models, that ignore these intra-industry effects, will thus substantially underestimate actual portfolio risk - by a degree that varies over the business cycle. "Conventional" models that include only systematic factors, although able to fit the broad trends in credit losses, fail to capture the fluctuations in unexpected credit losses, particularly in bad times, when banks typically are confronted with larger "unexpected" credit losses. The model we propose does manages to follow both the trend in credit losses and produces industry driven, time-varying fluctuations in losses around that trend. Our model will thus be a better aid for banks in determining the appropriate size of economic capital buffers.

What do these findings imply for financial institutions, financial supervisors and financial regulators? As we noted, the economic capital requirements that result from our model will be larger compared with the those implied by "conventional" models, especially in periods with large "aggregate" disturbances. All other equal, this should imply costlier lending for banks as they will be in need of larger capital buffers than one has thought until now. Another way to look at these results is that the trade-off banks face when determining how much to diversify is different. Manove, Padilla and Pagano ([38]) and Acharya, Hasan and Saunders ([3]) have, for example, discussed the costs that diversification bring about for banks and the trade-off they may face when balancing these against the gain from reducing tail credit losses. In this contaxt, our results seem to imply that the optimal degree of diversification is higher for banks than they inferred until now from less accurate models.

From a regulatory perspective, our findings will not have any immediate consequences, since regulatory capital requirements in the Basel II framework only depend on *average* probabilities of default. This paper is one in a larger literature that shows that the Basel risk weight mappings may not capture the actual risk properties of credit portfolios and that the success of implementing the Basel II framework will hinge on the specific characteristics (parameters) of banks' portfolios. For supervisors, it does mean, however, that the fragility of banks, and thus most likely the banking system, has been underestimated. In the long run, our findings ought to be a reason to make capital requirements contingent on portfolios' industrial composition, i.e. degree of diversifiaction.

As any model, ours assumes a specific structure. Obviously, there may be other ways to model intra-industry comovements between firms. We imposed that shocks to default risk are i.i.d. and firm default risk within an industry was positively correlated. One could think of some firms (within one or several industries) with a horizontal relation to display negative comovements, for example because a one firm's default creates new business opportunities for another. Another avenue may be to impose more structure on the cluster errors by using "input-output" data or other information documenting the intensity of contacts between firms and/or clusters. We believe that future research could explore these channels of inter-firm interaction.

References

[1] Allen, Linda and Anthony Saunders, (2004), Incorporating systemic influences into risk measurements: A survey of the literature, *Journal of Financial Services Research.*26 (2), pp. 161-191.

[2] Acharya, Viral, Sreedhar Bharat and Anand Srinivasan, (2004), Understanding the recovery rates on defaulted securities, Forthcoming in :*Journal of Finance.*

[3] Acharya, Viral, Iftekhar Hasan and Anthony Saunders, (2006), Should Banks Be Diversified? Evidence from Individual Bank Loan Portfolios, *Journal of Business.*79 (3), pp. 1355-1412

[4] Altman, Edward I., (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance* 23 (4), pp. 589-611.

[5] Altman, Edward I., (1971), Railroad bankruptcy propensity, *Journal of Finance* 26 (2), pp. 333-345.

[6] Altman, Edward I., (1973), Predicting railroad bankruptcies in America, *Bell Journal of Economics* 4 (1), pp. 184-211.

[7] Altman, Edward I., (1984), The success of business failure prediction models, *Journal of Banking and Finance* 4 (X), pp. 171-198.

[8] Arminger, Gerhard, Daniel Enache and Thorsten Bonne, (1997), Analyzing credit risk data: a comparison of logistic discrimination, classification tree analysis and feedforward networks, *Computational Statistics* 12, 293-310.

[9] Ashby, M., Neuhaus, J. M., Hauck, W. W., Bacchetti, P., Heilbron, D. C., Jewell, N. P., Segall, M. R. and R. E. Fusaro, (1992), An annotated bibliography of methods for analysing correlated categorical data, *Statistics in Medicine* 11:67-99.

[10] Bangia, Anil, Francis X. Diebold, André Kronimus, Christian Schagen and Til Schuermann, (2002), Ratings migration and the business cycle, with application to credit portfolio stress testing, *Journal of Banking and Finance* 26, pp. 445–474.

[11] Basel Committee, (2003), *The New Basel Capital Accord, Third Consultative paper*, www.bis.org/publ/bcbsca.htm.

[12] Breslow, N. E. and D. G. Clayton (1993), Approximate inference in generalized linear mixed models, *Journal of American Statistical Association*, 88:421, pp. 9-25.

[13] Butler, J. S. and Moffitt, R., (1982), A computational efficient quadrature procedure for the one-factor multinomial probit model, *Econometrica* 50(3):761-764.

[14] Calem, Paul and Michael LaCour-Little, (2004), Risk-based capital requirements for mortgage loans, *Journal of Banking and Finance* 28 (3), pp. 647-672

[15] Campbell, John Y., Jens Hilscher, and Jan Szilagyi, (2006), *In Search of Distress Risk*, NBER Working Paper 12362

[16] Carey, Mark, (1998), Credit risk in private debt portfolios, *Journal of Finance*, 52 (4), pp. 1363-1387.

[17] Carling, Kenneth, Tor Jacobson, Jesper Lindé and Kasper Roszbach (2006), *Corporate credit risk modeling and the macroeconomy*, Forthcoming in: *Journal of Banking and Finance.*

[18] Carey, Mark and Mark Hrycay, (2001), Parameterizing credit risk models with rating data, *Journal of Banking and Finance* 25, pp. 197-270.

[19] Das, Sanjiv R., Darrell Duffie, Nikunj Kapadia, and Leandro Saita, (2007), Common Failings: How Corporate Defaults are Correlated, Forthcoming in: *Journal of Finance*

[20] Duffie, Darrell, Andreas Eckner, Guillaume Horel and Leandro Saita, (2006), Frailty Correlated Default, manuscript Stanford University.

[21] Frydman, Halina, Edward I. Altman, and Duen-Li Kao, (1985), Introducing recursive partitioning for financial classification: the case of financial distress, *Journal of Finance*, Vol. XL No. 1, pp. 269-291.

[22] Giesecke, Kay, (2004), Correlated default with incomplete information, Forthcoming in: *Journal of Banking and Finance.*

[23] Giesecke, Kay, and Stefan Weber, (2004), Cyclical correlations, credit contagion, and portfolio losses, Forthcoming in: *Journal of Banking and Finance.*

[24] Gordy Michael B., (2000), A comparative anatomy of credit risk models, *Journal of Banking and Finance*, 24 (1-2) , pp. 119-149.

[25] Gordy, Michael B., (2003), A risk-factor model foundation for ratings-based bank capital rules, *Journal of Financial Intermediation* 12, pp. 199-232.

[26] Greene, W. H., (1991), Limdep version 6, Econometric Software, New York.

[27] Guilkey, D. K. and Murphy, J. L., (1993), Estimation and testing in the random effects probit model, *Journal of Econometrics* 59, pp.301-317.

[28] Hackbarth, D., J. Miao, and E. Morellec, (2006), Capital structure credit risk and macroeconomic conditions, Forthcoming in: *Journal of Financial Economics.*

[29] Heckman, J. J. and Willis, R., (1976), Estimation of a stochastic model of reproduction: An econometric approach, in: N. Terleckyj, ed. Houshold production and consumption, National Bureau of Economic Research, New York.

[30] Honjo, Yuji, (2000), Business failure of new firms: an empirical analysis using a multiplicative hazards model, *International Journal of Industrial Organization 18*, pp. 557-574.

[31] Jacobson, Tor, Jesper Lindé and Kasper Roszbach, (2006), Internal ratings systems, implied credit risk and the consistency of banks' risk classification policies, *Journal of Banking and Finance* 30 (7), pp.1899-1926

[32] Jacobson, Tor, Jesper Lindé and Kasper Roszbach (2005), Exploring interactions between real activity and the financial stance, *Journal of Financial Stability* 1 (3), pp.308-342.

[33] Jorion, Philippe, and Gaiyan Zhang, (2006), *Intra-industry Credit Contagion: Evidence from the Credit Default Swap and Stock Markets*, working paper University of California at Irvine

[34] Koopman, Siem Jan, Roman Kräussl, André Lucas, and André Monteiro, (2005), *Credit Cycles and Macro Fundamentals*, working paper Tinbergen Institute.

[35] Koyluoglu, H. Ugur, and Andrew Hickman, (1998), *A generalized framework for credit risk portfolio models*, Working paper, Oliver, Wyman & Company and Credit Suisse Financial

Products. Published in abridged version as: "Reconcilable Differences", *Risk* (October 1998), pp. 56-62.

[36] Li, Kai, (1999), Bayesian analysis of duration models: an application to Chapter 11 bankruptcy, *Economics Letters* 63, pp. 305-312.

[37] Littell, R. C., Milliken, G. A., Stroup, W. W. and R. D. Wolfinger, (1996), *The SAS system for mixed models*, SAS Inst. Inc., Cary, North Carolina.

[38] Manove, Michael, Jorge Padilla and Marco Pagano, (2001), Collateral vs. Project Screening: A Model of Lazy Banks, *RAND Journal of Economics*, 32 (4).

[39] McCullagh, P. and Nelder, J. A., (1989), *Generalized linear models*, London: Chapman and Hall.

[40] Merton, Robert C., (1974), On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", *Journal of Finance* 29, pp. 449-470

[41] Mingo, John J., (2000), Policy implications of the Federal Reserve study of credit risk models at major U.S. banking institutions, *Journal of Banking and Finance* 24 (1-2) , pp. 15-33.

[42] Narendranathan, W. and Stewart, M., (1993), Modelling the probability of leaving unemployment: Competing risks models with flexible baseline hazard, *Applied Statistics* 42, pp. 63-83.

[43] Olsson, U, (2002), "Generalized Linear Models: an applied approach", Studentlitteratur, Lund.

[44] Nickell, Pamela, William Perraudin and Simone Varotto, (2000), Stability of rating transitions, *Journal of Banking & Finance* 24, pp.203-227.

[45] Pesaran, Hashem, Til Schuermann, Björn-Jakob Treutler, and Scott Weiner, (2006), Macroeconomic Dynamics and Credit Risk: A Global Perspective, *Journal of Money Credit and Banking* 38 (5), pp. 1211-1262.

[46] Platt, Harlan D., and Marjorie B. Platt, (1991), A note on the use of industry-relative ratios in bankruptcy prediction , *Journal of Banking and Finance* 15, pp. 1183-1194.

[47] Statistics Sweden, (2002), Industry code - SNI2002, Stockholm, Sweden. www.scb.se/templates/Listning2____35025.asp

[48] Shumway, Tyler, (2001), Forecasting bankruptcy more accurately: a simple hazard model, *Journal of Business* 74 (1), pp. 101-124.

[49] Song, J. X. and C. W. Ahn, (2003), An evaluation of methods for the stratified analysis of clustered binary data in community intervention trials, *Statistics in Medicine* 22:2205-226.

[50] Szegö, Giorgio, (2002), Measures of risk, *Journal of Banking and Finance* 26, pp. 1253-1272.

[51] Treacy, William and Mark Carey, (2000), Credit risk rating systems at large US banks, *Journal of Banking and Finance* 24 (1-2), pp. 167-201.

[52] Wilson, Thomas, (1997), Portfolio credit risk (I), *Risk* 10 (9), pp. 111-116

[53] Wolfinger, R. and M. O'Connell, (1993), Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation Simulation* 48, pp. 233-243.

[54] Zhou, X.-H., Perkins, A. J. and Hui, S. L., (1999), Comparisons of software packages for generalized linear multilevel models, *The American Statistician* 53 (3), pp. 282-290.

[55] Zhou, Chunsheng, (2001), An analysis of default correlations and multiple defaults, *Review of Financial Studies* 14 (2), pp. 555-576.

A Appendix

Table 6. Definition of clusters and their distribution in the data.

Variable	Included SNI-codes					% share
Public or subsidized sector	0-9999	22100-22330	40000-41000	75000-91330	99000-99999	13.3
Wholesale	51000-51700					12.7
Retail	50000-50500	52000-55529	71000-71100	71400-71402	92000-98999	19.4
Transport and communication	60000-64203	71200-71340	72000-74849			23.5
Manufacturing	27000-35999					7.8
Construction, forest and other industries	10000-22000	23000-26829	36000-39999	45000-45500		14.9
Estate and finance	65000-70329					8.4