Estimation of rating class transition probabilities with incomplete data

Thomas Mählmann

Chair of Banking, University of Cologne

Albertus-Magnus-Platz

50923 Koeln/Germany

Tel: (0049)+221-4702628

Fax: (0049)+221-4702305

maehlmann@wiso.uni-koeln.de

Estimation of rating class transition probabilities with incomplete data

Abstract

This paper shows that the well known "duration" and "cohort" methods for estimating transition probabilities of external bond ratings are not suitable for internal rating data. More precisely, the duration method *cannot* and the cohort method *should not* be used in connection with bank rating data. Structural differences within the borrower monitoring process of banks and rating agencies are responsible for this result. A Maximum Likelihood (ML) estimation procedure, which accounts for the peculiarities of internal bank ratings, is introduced and applied to data from a German bank. The empirical results indicate that the differences between cohort and ML transition matrices are both, statistically and economically significant. Furthermore, evidence of rating reversals, business cycle dependent transition probabilities and on the factors which determine the borrower monitoring intensity of banks is provided.

Key words: internal ratings, transition probabilities, Markov process, borrower monitoring

JEL classification codes: C24, C41, G21, G28

1 Introduction

Over the past few years, banks have attempted to mirror the rating behavior of external rating agencies. But, are internal and external rating procedures really very similar? Or, more specifically, can methods developed for external ratings (e.g. migration probability estimation or validation techniques) be applied to internal bank ratings? This paper tries to provide answers to these questions in the context of migration risk, i.e. the probability of moving from one rating class to another within a given amount of time.

At first sight, the estimation of the migration or transition probability matrix for internal ratings seems to be an easy task in face of the methodology already available for estimating transition probabilities of external bond ratings. However, as will be seen, there is at least one fundamental difference between external and internal rating data which prevents the transfer of the methods developed for agency ratings. In particular, we argue that internal ratings are not continuously monitored and so internal rating data is discrete, i.e. observations are taken at specified time points. In the absence of continuous monitoring, a borrower could change credit grade without anyone noticing. Moreover, banks frequently base the decision of when to monitor a borrower on an estimate of the borrowers expected loss (see Basel Committee on Banking Supervision, 2000, or Treacy and Carey, 2000). For example, borrowers with cash-collateralized loans or borrowers in higher-quality grades may receive a refreshed rating only once every two years whereas low-quality borrowers may be reviewed several times within one year. In sum, internal rating data can be characterized by two important features: (i) Exact times of movement between rating classes and the class occupancy in between the observation times are unknown and (ii) observation and interexamination times vary for different borrowers.

The traditional discrete "cohort" or "multinomial" type estimator as well as the continuous "duration" estimator (see Jafry and Schuermann, 2004, or Lando and Skødeberg, 2002) are not capable to deal appropriately with such data. In concrete terms, the first data feature prevents the application of the duration approach and the second feature brings about an efficiency loss for the cohort estimator. Therefore, an alternative estimation method that is suitable for incompletely observed rating data is proposed and applied to data from a German bank. This method derives ML estimates of transition probabilities under the assumption that rating dynamics follow a time-homogeneous first-order Markov process.

The empirical part of this paper studies the following three questions: (i) Which factors influence the monitoring intensity of bank borrowers? (ii) Are cohort and ML estimates of transition probabilities statistically and economically distinguishable? (iii) Can the process underlying internal rating changes be described as a time-homogeneous first-order Markov process?

Concerning the first question, our evidence suggests that the more favorable the current rating the greater, *ceteris paribus*, the probability for a review within less than a one-year period. This somewhat surprising finding might be explained by agency issues between the lending officer in the bank and senior management. Furthermore, a previous upgrade pattern and a longer duration of the bank-borrower relationship reduce the review frequency, and the effect of borrower size is U-shaped.

With respect to one-year transition probability matrices, it seems to be that two countervailing effects drive the differences between the ML and the cohort estimates: A duration dependence effect states an inversely U-shaped relation among the time between two subsequent monitoring steps and the probability of a rating change, and a downgrade effect reflects the empirical observation (see Lando and Skødeberg, 2002) that defaulted borrowers are frequently downgraded to the highest risk class shortly before they actually default. Taking both effects together results in a small, but statistically and economically significant value for the distance metric proposed by Jafry and Schuermann (2004) for comparing migration matrices.

Finally, by using a form of a proportional hazard model we find evidence for business cycle influence on rating intensities and for rating reversals, i.e. an upgrade followed by a downgrade, or a downgrade followed by an upgrade. Such rating bounces are consistent with the assumption of Gaussian distributed credit quality changes in structural models of rating transitions (Gordy and Heitfield, 2001, or Löffler, 2005). However, business cycle dependence and rating reversals undermine the assumptions of any of the estimators discussed above and thus leave open some important questions for future research on the analysis of internal ratings.

The paper proceeds as follows. Section 2 points out key differences between what we are used to seeing, i.e. histories of agency ratings, and bank data. In section 3, a regression analysis is performed to characterize the specific nature of internal rating data from a German bank. The new generator-based ML transition probability estimator is described in section 4 and compared to the widely applied cohort estimator in section 5. The paper concludes with section 6.

2 Structure of internal rating data

Obviously, before estimating transition probabilities one should ask whether all transitions made by a single borrower over a specified observation period and the corresponding exact transition times are known. Or, in short, whether the rating data is continuous or discrete. To answer this question for the kind of data that can be obtained from agencies and banks, the monitoring intensity of both has to be analyzed.

Regarding rating agencies, it seems reasonable to expect that their data is on a approximately continuous scale, i.e. roughly exact transition times are available.[1] Agencies are able to perform almost a continuous monitoring mainly for two reasons. First, they typically rate only large public firms which do regularly issue fixed income securities. These firms are forced to disclose a considerable amount of information, which is digested and filtered by many market participants. Moreover, since agencies are directly paid by the issuers of debt, they can spend substantial resources on discovering and updating private information.

In contrast, commercial banks are faced with a different situation. They have to rate a bulk of small and middle-market loans and the costs of producing ratings must be covered by revenues on credit products. The expenditure of resources at a rate similar to that of the rating agencies would make the business with small and middle-market loans unprofitable. In addition, monitoring private firms with poor and infrequent information is much more costly. Therefore, to reduce costs banks are forced to perform a formal review of each rating on an annual basis only. However, Machauer and Weber (1998) report that the time between two consecutive monitoring processes does not always equal one year in their sample. Indeed, it seems to be common practice to base the decision about the frequency of reviews on the current rating or the collateral (see Basel Committee

[1]Note that even rating agencies do not perform a truly continuous monitoring. This can be seen from the existence of watchlists, which are of no use if all issuers were under a continuous review anyway. In addition, rating agencies are often accused of being too slow to adjust their ratings. For example, empirical evidence by Delianedis and Geske (1999) shows that rating changes lag changes in default probabilities. A possible reason for not reporting exact transition times is the agency's "rating reversal aversion" (see Löffler, 2005). A continuous monitoring, however, does not imply that ratings are changed every time new information is obtained. According to their "through-the-cycle" point of view external ratings are stable because they are intended to measure default risk over long investment horizons, and thus they are changed only when agencies are confident that observed changes in a firm's risk profile are likely to be permanent.

on Banking Supervision, 2000, p. 37, Treacy and Carey, 2000, p. 182). In addition, the theoretical work of Diamond (1989, 1991) suggests that holding credit risk constant, monitoring and borrower reputation are close substitutes. Then, as firms successfully complete loan transactions with banks and thus acquire reputation, banks monitor them less frequently and, ultimately, charge them lower interest rates. Blackwell and Winters (1997) provide empirical evidence that banks less frequently monitor firms with whom they have closer relationships.

From the preceding arguments we might conclude that in almost all cases internal rating data will be discrete and the choice of the observation times depends on: (i) The current rating (reflects the probability of default, PD), (ii) the terms of the loan (reflect the loss given default, LGD), (iii) the loan size (reflects the exposure at default, EAD) and (iv) the closeness of the bank-borrower relationship (reflects the moral hazard component of credit risk). Thus the structure of the monitoring times will vary for each borrower, depending on his level of expected loss (EL=PD·LGD·EAD).[2]

3 Empirical analysis of monitoring intensities

3.1 The data

Our data comes from the internal rating system of a medium-sized German bank. This rating system is composed of two parts: A quantitative subrating, called financial rating, and a qualitative subrating, called corporate situation rating. As opposed to the corporate situation rating, which relies on a subjective assessment of qualitative criteria (e.g. market position, management quality etc.), the financial rating is based on a statistical model, which essentially combines ratio analysis and industry sector information. The overall rating results by applying the subjectively determined, fixed weights 0.7 and 0.3 to the financial and corporate situation subratings, respectively. The bank provided us with the rating history of all borrowers in its loan portfolio as at January 31, 2002, that did not default by the end of that year. In addition, rating information for all borrowers which defaulted within the period 1996-2002 is obtained. Here legal bankruptcy proceedings and loan loss

[2]It should be noted that the Basel Committee on Banking Supervision, when formulating the minimum requirements for the application of the IRB approach, acknowledges the impossibility of a continuous monitoring of small and medium-sized borrower ratings. Banks are only required to refresh their borrower and facility ratings "at least on an annual basis", whereas the frequency of reviews must be risk-sensitive (see Basel Committee on Banking Supervision, 2004, § 387).

provisions are used as proxies for default. Because information about borrowers which defaulted before 1996 is not stored by the bank, extending the observation period behind 1996 would result in an oversampling of non-defaulted borrowers. To prevent this kind of survivorship bias, we restrict the observation period to 1996-2002. Moreover, we do not consider very small firms with an annual turnover of less than €2.5 Mio. Besides rating and default data we collected some additional information (e.g. annual turnover, industry sector, legal form etc.) for each borrower. The final database contains 12,252 observations from 2,247 non-defaulted and 110 defaulted borrowers.

While the original rating system has 24 grades for non-defaulted borrowers, we built up a simpler six-position rating system, i.e. $l_D - 1 = 6$.[3] Rating grade 1 corresponds to the lowest and grade 6 to the highest degree of credit risk. Table 1 presents the distribution of the 12,142 rating and the 110 default observations over five different size categories. Here size is measured by the annual turnover given in the last available financial statement. As can be seen, large firms have a higher (lower) frequency of high (low) quality grades than small firms. Furthermore, the default proportion is monotonically decreasing with firm size. These findings also appear in Dietsch and Petey (2004, Table 2). As expected for a lending portfolio of a large bank, the majority (more than 85% in this case) of borrowers are SMEs (annual turnover $\leq$ €50 Mio.).

The temporal distribution of the observations, given in Table 2, demonstrates two interesting points. First, at least the high- and low-quality ratings show a dependency on time. For example, the proportion of class 1 ratings reaches its maximum in 1999 and decreases monotonically afterwards. Second, the number of defaults increases significantly for the years 2000, 2001 and 2002, i.e. the overall weighted average annual default rate equals 0.90% for 1996-2002 and just 0.36% for 1996-2000. This reflects the influence of the business cycle and shows that defaults move counter-cyclically. Note that the annual German GDP growth rates from 1996 to 2002 equal 1.0%, 1.8%, 2.0%, 2.0%, 3.2%, 1.2% and 0.2%.[4]

[3] In fact, the bank's system has eight main grades, each of which can be supplemented with a plus or a minus to provide a finer indication of risk. To construct our six grade scale, we consider only the main grades and translate the first four into two new grades.

[4] The Economic Cycle Research Institute (see http://www.businesscycle.com) provides a business cycle chronology for the main European countries based on the NBER methodology. The corresponding quarterly recession dates for Germany after reunification are: 1991:Q1−1994:Q1 and 2001:Q1−2002:Q2.

3.2 Results

It was argued that for internal ratings the interexamination times (times between successive ratings) cannot expected to be equally spaced in general. In our data set 92.77% of the 9,895 recorded interexamination times display an equal length of 12 months, whereas 1.99% are smaller and 5.24% are larger than 12 months. Actual observation times, i.e. the dates the credit department signs off on the credit review, are centered around the months March and September of each year, reflecting the time lags with which accounting information becomes available. To obtain more insight into the determinants of interexamination times a regression analysis is performed. The variables used for this analysis are shown in Table 3.

The dependent variable MONTHS is defined as the number of months between two consecutive rating assignments for borrower i at times t_{ij} and t_{ij+1}, respectively, where $j = 0, 1, \ldots, j_i - 1$ indexes the monitoring step. Variables on *risk* characterize borrower risk, i.e. the default probability of borrowers, without taking loans terms into account. Mester et al. (2002) provide empirical evidence that banks intensify monitoring activities as borrower quality deteriorates - rating reviews become lengthier and are more frequent. Besides controlling for the present (at t_{ij}) risk level through the rating category dummies R1-R6, we are also interested in examining the influence of the last observed rating change on MONTHS. If t_{ij-1} denotes the monitoring time preceding t_{ij}, the dummies UPGRADE and DOWNGRADE are calculated by comparing the rating at t_{ij-1} with that at t_{ij} and grouping together all kinds of downgrades or upgrades. One might suppose that a favorable past rating pattern, i.e. UPGRADE=1, is connected with an increase in MONTHS, given the rating at t_{ij}. In contrast, a recently upgraded firm might be subject to a faster review because an unwarranted upgrade decision will be more harmful to the career of a risk averse credit officer than a wrongly downgrade.

LOGSIZE measures the natural logarithm of a firm's sales per year as reported in the last financial statement before t_{ij}. It serves as a proxy for firm size. Blackwell and Winters (1997) find that banks monitor larger firms less frequently, maybe because larger firms have more valuable assets to pledge as collateral. However, they do not test for a nonlinear relationship. Since information is more difficult and costly to obtain for very small firms, the overall size effect might be U-shaped.

To measure the likely exposure of each borrower in the case of default, we simply assume that this exposure is the balance sheet amount (at t_{ij}) of debt supplied by the bank. Because a higher exposure at default is connected with a higher expected loss, *ceteris paribus*, the variable EXPOSURE is expected to be negatively related to MONTHS. Evidence for this kind of relationship is provided by Udell (1989), who reports that the vast majority of surveyed banks used loan size as a selection criteria for determining the number of reviews performed by an independent loan review department.

Relationship lending is captured by the variable HOUSEBANK, i.e. the proportion (at t_{ij}) of total bank debt provided by our cooperating bank. When HOUSEBANK tends to 1, the firm's present and potential borrowing concentration with the bank is greater and the banking relationship is stronger. In addition, the variable DURATION measures the strength of the relationship in terms of its temporal length - the amount of time the bank has provided loan, deposit, or other services to the firm. As already mentioned, theoretical arguments as well as empirical evidence suggests that banks less frequently monitor firms with whom they have closer relationships.

The governance-related influences are summarized by the dummy LIMLIAB, which takes on a value of 1 if the firm is incorporated and 0 otherwise. As limited liability restricts the bank's access to private assets of the owners in distress situations, we expect to find a negative effect. Finally, Y96-Y01 are control variables for the years 1996-2001 and reflect possible macroeconomic effects on MONTHS. Since no completed interexamination time is starting in 2002, a dummy for this year is not necessary.

Unfortunately, complete data on all covariates is only available for 5,640 (57%) of the 9,895 rating pairs. Furthermore, we have to face two additional problems. First, the dependent variable MONTHS is clustered around 11 distinct realizations and is clearly not normally distributed. So OLS regression is inappropriate. We decided to transform MONTHS into the ordinal variable MONTHSCAT which equals 0, 1, and 2 if MONTHS is below, equal or above 12 months, respectively. Obviously, such a categorization leads to a loss of information in the sense that merely the direction but not the economic effect (in months) of a unit increase in each covariate can be estimated.

Second, since we have no information about terms of lending (e.g. collateral, covenants, loan commitments) there is likely to be unobserved heterogeneity between borrowers. Even if we deal here with updates of borrower ratings, not facility ratings, it seems reasonable to suppose that some loan characteristics also influence the frequency of borrower rating reviews. For example, a fully collateralized lender is immunized from borrower performance if the value of the collateral is unaffected by any actions the borrower undertakes once he or she has been funded. Then the provision of security completely obviates the bank's need for investigating the borrower. Manove et al. (2001) provide a model which shows that the use of collateral in debt contracts may reduce the screening and monitoring effort of banks below its socially efficient level and lead them to fund too many worthless investment projects. Moreover, since the benefits from monitoring and control functions of banks are subject to free-riding by other stakeholders, such as trade creditors, employees, and the government, the bank's incentives to acquire costly private information are reduced in general. Rajan and Winton (1995) show that when monitoring is socially beneficial long-term debt with covenants may be preferable to covenant-free short-term debt because effective use of covenants forces the lender to do some monitoring. We control for the combined effect of all unobserved borrower-specific covariates by including a Gaussian random effect into the regression model.

The results from a random-effects ordered probit model are given in Table 4. Most surprisingly, all rating dummies, which are significant at better than the 5% level, show a negative coefficient.[5] This implies that a rating better than class 6 increases the probability for a short ($<$12 months) interexamination time. Furthermore, with the exception of class 1, the more favorable the rating the greater, *ceteris paribus*, the probability for a review within less than a one-year period. Possibly, this phenomenon can be explained by two arguments. First, it seems reasonable to expect that for high-quality borrowers the completion of the rating review should take less time than for borrowers that have deteriorated in quality, since for healthy borrowers the loan officer is less likely to find troublesome information that takes longer to evaluate. Moreover, in the case of bad news, the rating review is typically prolonged by the bank's requests for additional information, such as more complete financial statements, and higher ranking credit officers may become involved in the review. In addition, because of reluctance to provide the bank with unfavorable financial results or management's inability to generate and package a set of financial reports, a failure to submit the requested financial information in time is more likely for low-quality borrowers.[6] This reasoning is

[5] We performed several robustness checks on the model specification. For example, we excluded the variables UPGRADE, DOWNGRADE, HOUSEBANK and EXPOSURE with missing data and repeated the analysis using all 9,895 interexamination times. The results for the remaining variables are qualitatively unchanged. Especially, the rating dummies again have significant negative coefficients.

[6] Empirical studies by Lawrence (1983) and Whittred and Zimmer (1984) into the timeliness of financial reporting and financial distress in large firms generally support the proposition that failing firms tend to take longer to produce

in line with empirical evidence on the duration of rating reviews provided by Mester et al. (2002).

Second, an internal incentive conflict between a bank and its loan officer can result in the loan officer systematically delaying or even concealing rating updates when disclosure of such updates reduces his private benefits (Udell, 1989). For example, loan officers might fear that the discovery of borrower quality deterioration will reflect on their original credit judgement and thus harm their future career. This may encourage the loan officer not to reveal any new bad information about the borrower in the hope that subsequent events will produce a restoration of credit quality.

As expected, the effect of borrower size is U-shaped, with inflection point at €60.616 Mio. annual sales, and limited liability increases the probability of a short ($<$12 months) interexamination time, but this effect is only weakly significant.[7] Furthermore, while a previous upgrade pattern reduces the review frequency, a recently downgraded firm is not exposed to an increased monitoring intensity. The results for the two variables measuring the strength of the bank-borrower relationship are contradictory. First, the probability of a long ($>$ 12 months) interexamination time increases with the duration of the bank-borrower relationship. This is in line with the empirical results of Blackwell and Winters (1997) and the reputation-based arguments of Diamond (1989, 1991), suggesting a reduction of information asymmetry between lender and borrower over time as firms successfully complete loan transactions with banks. In contrast, the variable HOUSEBANK is not significant. This implies that the bank does not monitor borrowers more or less frequently for which it serves as the dominant provider of bank debt. We tested a model without DURATION but found HOUSEBANK to be still not significant.

Finally, since most of the year dummies are insignificant the monitoring intensity does not vary systematically across years. The significant positive coefficient for Y00 represents an exception, indicating that, in comparison to 2000, the bank increased its overall monitoring activity for the year 2001, the beginning of an economic downturn in Germany. However, because we do not have rating information for the years after 2002, this result might in part be due to an oversampling of short interexamination times starting in 2001.

The bottom part of Table 5 shows that the intraclass correlation ρ, i.e. the proportion of the total variance contributed by the panel (borrower)-level variance component, is statistically signif-

their annual financial statements than healthy firms.

[7]Because quadratic $ax^2 + bx + c$ turns over at $x = -b/2a$, which for our LOGSIZE and LOGSIZE^2 coefficients is $4.5591/(2 \times 0.2070) \approx 11.01$, the inflection point is $exp(11.01) =$ €60.616 Mio.

icant (p-value<0.0001). Therefore, the panel-level variance component is important and the panel estimator is different from the pooled estimator. Having understood the data, we can now turn to our main object of interest, namely to estimate transition probabilities with the different kinds of data.

4 Methods for estimating transition probabilities

We assume that credit quality dynamics can be described by a first-order continuous-time time-homogeneous Markov chain $X(t)$ denoting the rating class occupied at time t ($t \geq 0$). Furthermore, we assume that the chain has a finite number of states, $1, 2, \ldots, l_D$. Here $1, \ldots, l_D - 1$ are defined by decreasing levels of credit quality and l_D is the default state, which is absorbing. Given that class l is entered at calendar time t and is still occupied at $t + s$, the transition out of l is determined by the set of $l_D - 1$ transition intensities $q_{ll'}(t, s) = q_{ll'}$, with $q_{ll'} \geq 0$, $q_{ll} = -\sum_{l' \neq l} q_{ll'}$. Let Q denote the $l_D \times l_D$ intensity matrix or generator, and $P(s)$ the $l_D \times l_D$ transition probability matrix whose (ll')'th element $p_{ll'}(s)$ is the probability of migrating from state l to state l' within a time interval of length s. It is well known (see, for example, Cox and Miller, 1965, p. 182) that for a time-homogeneous Markov chain the relationship between generator Q and transition matrix $P(s)$ is given by:

$$P(s) = \exp(Qs) = \sum_{r=0}^{\infty} Q^r \frac{s^r}{r!}. \tag{1}$$

If the credit quality process can be completely observed for each individual, i.e. if the exact transition times are known, there are no serious difficulties for analyzing such data either in a parametric (see Lando and Skødeberg, 2002) or in a nonparametric fashion (see Fledelius et al., 2004). Under the assumption of time-homogeneity[8], the ML estimator of the $l \to l'$ transition intensity is given by

$$\widehat{q}^{D}_{ll'} = \frac{\sum_{i=1}^{n} N^i_{ll'}}{\sum_{i=1}^{n} T^i_l}, \quad l' \neq l, \tag{2}$$

where $N^i_{ll'}$ denotes the total number of $l \to l'$ transitions made by borrower i and T^i_l is the

[8] A modified version of the duration estimator, the Aalen-Johansen estimator, is available for the time-inhomogeneity case (see Lando and Skødeberg, 2002).

total time spent by i in class l. This duration estimator counts all rating changes and divides by the total time spent in each rating. However, if the nature of the data is discrete in the sense that the observations consist of the classes occupied by the borrowers at a sequence of discrete time points $t_{i0} < t_{i1} < t_{ij} < \ldots < t_{ij_i}$, $\widehat{q}^D_{ll'}$ cannot be used to estimate transition intensities. With no information available about the timing of events between observation times or about the exact transition time, neither the numerator nor the denominator in (2) can be calculated.

Assuming that the data is discrete and that the observation times of the Markov chain are identically for each borrower, i.e. $t_0 < t_1 < t_j < \ldots < t_{j_n}$, and equally spaced such that $s_j = t_j - t_{j-1} = s$ for all $j = 1, \ldots, j_n$ monitoring steps, then the ML estimators of the stationary transition probabilities

$$\widehat{p}^C_{ll'}(s) = \frac{N_{ll'}}{\sum_{l'=1}^{l_D} N_{ll'}}, \quad l,\ l' = 1, \ldots,\ l_D. \tag{3}$$

are the common cohort estimators. Here, $N_{ll'} = \sum_{j=1}^{j_n} N_{ll'j}$ is the total number of recorded transitions from l to l'. Clearly, if observation times are not equally spaced and identical for all borrowers, then $\widehat{p}^C_{ll'}(s)$ is not a ML estimator for $p_{ll'}(s)$. For example, to estimate the one-year transition matrix $P(1)$ the cohort method uses only the ratings at the beginning and end of a calendar year, subsequent ratings more than or less than one year away from the initial rating are ignored. The inefficiency of the cohort method, documented by Lando and Skødeberg (2002) or Jafry and Schuermann (2004) for the continuous-time case, can thus be extended to discrete data with a rather irregular pattern of times at which observations are taken that might also vary from borrower to borrower.

The evidence presented in section 3.2 indicates that the time between consecutive rating reviews depends on borrower-specific characteristics and is therefore unlikely to be equal for all borrowers at all points in time. Then, irrespective of what time period is used, the cohort method yields inefficient estimates when applied to such discrete data. In the following, we outline a method, based on original work of Kay (1986), which is capable of providing ML estimates of the generator Q when the exact times of movement between rating classes and the class occupancy in between the observation times are unknown. Furthermore, the observation times are assumed to be arbitrary.[9]

[9]Some related work is provided by Kalbfleisch et al. (1983) and Kalbfleisch and Lawless (1985). However, they

The key point in deriving ML estimates of the intensities is to formulate the transition probabilities $p_{ll'}$ in terms of the intensities. Let θ denote the vector of intensities $q_{ll'}$ $(l \neq l')$ which are to be estimated. Then, using a canonical decomposition and assuming $Q(\theta)$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_{l_D}$, from (1) we have (see Cox and Miller, 1965, p. 184):

$$P(s) = A_0 \, (e^{\lambda_1 s}, \ldots, \; e^{\lambda_{l_D} s}) A_0^{-1} \, , \tag{4}$$

where A_0 is a $l_D \times l_D$ matrix whose lth column is the right eigenvector for λ_l. In general, $P(s)$ is a complicated function of θ but it is possible numerically to obtain the eigenvalues $\lambda_1, \ldots, \lambda_{l_D}$, the eigenvectors making up A_0 and then $P(s)$. Let $t_{i0} = 0, t_{i1}, \ldots, t_{ij_i}$ represent the times at which the rating of borrower i is recorded as $x_{i0}, x_{i1}, \ldots, x_{ij_i}$, respectively, with $x \in \{1, \ldots, l_D\}$. If borrower i does not default during the period studied, he is censored at the end of the observation period, i.e. these borrowers are known to be alive but with an unknown rating at t_{ij_i+1}.[10] Accounting for such censored observations x_{ij_i+1}, known only to be a state in the set $R = \{1, \ldots, l_D - 1\}$, the contribution of i to the likelihood is

$$L_i(\theta) = \prod_{j=0}^{j_i - 1} p_{x_{ij} x_{ij+1}}(t_{ij+1} - t_{ij}) p^{\#}_{x_{ij_i}}(t_{ij_i+1} - t_{ij_i}), \tag{5}$$

where for $l = 1, \ldots, l_D - 1$,

$$p^{\#}_l(t_{ij_i+1} - t_{ij_i}) = \sum_{r \in R} p_{lr}(t_{ij_i+1} - t_{ij_i}). \tag{6}$$

Correspondingly, if borrower i defaults at time t_{ij_i} the likelihood contribution is

$$L_i(\theta) = \prod_{j=0}^{j_i - 2} p_{x_{ij} x_{ij+1}}(t_{ij+1} - t_{ij}) p^{*}_{x_{ij_i-1} l_D}(t_{ij_i} - t_{ij_i-1}), \tag{7}$$

where for $l = 1, \ldots, l_D - 1$,

---

both consider a situation in which the observation times are equal for all individuals. Furthermore, Kalbfleisch et al. (1983) suppose that not the transition counts $N_{ll'j}$ but only aggregate data, i.e. the number of individuals in each state at any specified observation time, are available. They show how to obtain approximate ML estimates in this case.

[10] In the empirical part of this paper, the observation period ends at December 31, 2002. No information about the default status of the remaining borrowers is available afterwards. The last date at which a rating is observed is June 16, 2002.

$$p^*_{ll_D}(t) = \sum_{l^{\prime}=1}^{l_D-1} p_{ll^{\prime}}(t-1) q_{l^{\prime} l_D} \ . \tag{8}$$

Here it is assumed that the exact time of default is known, but the rating class on the previous instant before default is unknown. If, for example, time is measured in days, then the contribution to the likelihood is summed over the unknown class $l^{\prime} \neq l_D$ on the day before default. The full likelihood function, conditional on the distribution of borrowers among states at t_{i0}, is simply the product of the likelihood contributions over all n borrowers, i.e.

$$L(\theta) = \prod_{i=1}^{n} L_i(\theta). \tag{9}$$

For general $l_D \geq 2$, a Quasi-Newton algorithm can be employed to find the ML estimates of this function, using finite differences to obtain numerical approximations of the derivatives.[11] Starting values for the iterative procedure can be calculated with the duration estimator in (2) by assuming that the observation times represent the exact transition times between grades.[12] In the following, using the data described in section 3.1, one-year transition probabilities are estimated and possible deviations from the time-homogeneity and Markov assumptions are analyzed.

5 Empirical application

5.1 Estimates of transition probabilities

By maximizing the likelihood function (9) using all available 12,252 rating observations (9,895 rating pairs) and taking the matrix exponential of the estimated generator we get the one-year ($s =$ 12 months) transition probability matrix $\widehat{P}_{ML}$ given in Panel A of Table 5.[13] These probabilities can be compared with the cohort estimates reported in Panel B, which are based only on the

[11]Kalbfleisch et al. (1983) describe a way of computing the first derivatives of the entries of $P(s)$ with respect to θ. Given the high costs of the numerical evaluation of the likelihood function, the algorithm can be significantly accelerated by using their analytic expression for the first derivatives.

[12]Note that for the function in (9) to be the true (conditional) likelihood function we have to assume that there is no migration correlation between different borrowers. Moreover, observation times are not allowed to depend on unobserved borrower characteristics.

[13]Below, we restrict our discussion to one-year matrices, so that the conditioning on s can be ignored for notational convenience.

subset of the rating pairs with a one-year interexamination time. As can be seen, one important difference between $\widehat{P}_C$ and $\widehat{P}_{ML}$ is the fact that, referring to the ML approach, there is a measurable strictly positive probability for the $6 \rightarrow 1$ transition. In contrast, the cohort method estimates this probability to zero. Because there are observed transitions from class 6 to class 5 and from class 5 to class 1 (but not from class 6 to class 1 directly), the estimate for transitions from class 6 to class 1 should be non-zero. The ML estimator captures this whereas the cohort method does not.

Presumably more importantly, both methods generate different default probability estimates. Except for the highest- and lowest-quality grades, the less efficient cohort method overestimates default risk. This results from the fact that much of the observations for grades 2-5 not used by the cohort estimator are non-default transitions. Moreover, there is a sizeable difference of more than 10% in the one-year default probability estimate of a firm in class 6. It is important to note that this structure of differences between both methods in terms of default probability estimates is similar to the differences reported by Lando and Skødeberg (2002) and Jafry and Schuermann (2004) when comparing the duration and cohort method using rating agency data.

Since the regression analysis in section 3.2 suggests that, *ceteris paribus*, a high- (low-) quality credit rating increases the chance for a short (long) time to the next review, one would expect the differences between the two estimation methods to be more pronounced at both, the lower and upper ends of the rating scale. However, because of a mixing of several effects, this is not the case, at least for the high-quality grades. To provide more evidence in this respect we estimated an ordered probit rating prediction model using the following variables: four financial ratios[14], LOGSIZE, LOGSIZE^2, DURATION, and sector and time dummies. The signs of all estimated parameters[15] match expectations, in particular, the effect of size is U-shaped and the probability of obtaining high-quality ratings increases with the length of the bank-borrower relationship. This implies an indirect effect of DURATION on MONTHSCAT which is contrary to its direct effect reported in Table 4. More precisely, controlling for the rating, a longer bank-borrower relationship

[14]The financial ratios are: (net working capital)/(total assets), (retained earnings)/(total assets), (earnings before interest and taxes)/(total assets) and (book value of equity)/(book value of debt). With the exception of replacing book value of equity by its market value these ratios are also employed by Altman and Rijken (2004) in the context of predicting agency ratings. In addition, the same log-transformations of the ratios as given in Altman and Rijken (2004) are applied. Because of missing financial statements, complete data on all covariates is only available for 7,538 of the 12,252 rating observations.

[15]Not reported here due to size limitations, but available from the author upon request.

increases the chance for a long interexamination time, but is also connected with a better rating, which itself predicts a shorter time to the next review.[16]

To answer the question whether both methods yield matrices that are statistically distinguishable, we need a metric which measures the scalar difference between these matrices. Jafry and Schuermann (2004) consider a set of criteria by which the performance of a proposed metric should be judged. One of their main requirements is *distribution discriminatory*, i.e. the metric should discriminate between matrices having the same diagonal probabilities but different off-diagonal distributions. Such a distinction between matrices with the same amount of mobility is important in the context of credit risk since far migrations have different economic and financial consequences than near migrations. The metric proposed by Jafry and Schuermann (2004) for a transition probability matrix P of dimension $l_D \times l_D$ is

$$M_{SVD}(P) = \frac{\sum_{l=1}^{l_D} \sqrt{\lambda_l \left(\widetilde{P}' \widetilde{P} \right)}}{l_D}, \tag{10}$$

where $\widetilde{P} = P - I$ and $\lambda_l(E)$ denotes the lth eigenvalue (arranged in the sequence from the largest to smallest absolute value) of a $l_D \times l_D$ matrix E. Note that by subtracting the identity matrix I from the migration matrix only the dynamic part of the original matrix is left, which reflects the "magnitude" of P in terms of the implied mobility. Jafry and Schuermann (2004) show that the value of $M_{SVD}(P)$ indicates something like the "average amount of migration" contained in P.

The observed value for the distance metric $\Delta M_{SVD}(\widehat{P}_{ML}, \widehat{P}_C,) = M_{SVD}(\widehat{P}_{ML}) - M_{SVD}(\widehat{P}_C)$ between the migration matrices estimated using the ML and the cohort approach for the period 1996-2002 equals 0.00047. This somewhat low[17] number results from the impact of two countervailing effects: A *duration dependence* and a *downgrade* effect. Concerning the first, much of the interexamination times shorter or longer than 12 months yield no rating change, whereas more ratings are altered after 12 months. That is, the data suggests that a short or a long interexamina-

[16]Moreover, because a large fraction of the high-quality ratings are previously upgraded, there is a mixing of the rating effect, suggesting an increase in the review frequency, and the UPGRADE effect, predicting a longer waiting period until the next review.

[17]As a comparative value, Jafry and Schuermann (2004) report that transition matrices estimated with the parametric (time homogeneous) duration method for expansion and recession periods have a difference in M_{SVD} of 0.0434.

tion time increases the probability for a firm not to change its rating.[18] Note that this definition of duration dependence is different from the duration dependence effect discovered by Lando and Skødeberg (2002) for external bond ratings. They estimate that the probability of changing its rating decreases with the time a firm spends in its current rating. In our case, duration dependence is related to the time since the last rating review, and not the last rating change, which is possibly not observed. Since only the ML approach includes all interexamination times, the ML estimates for the diagonal elements of the migration matrix should be higher compared to the cohort estimates. This is indeed the case, except for the lowest and highest risk grade.

Concerning the downgrade effect, most of the defaulted firms are downgraded to grade 6 shortly before they actually default. However, these downgrades are often not observed by the cohort method, resulting in an default probability estimate (3.79%) which is almost four times lower than the ML estimate (14.59%). In sum, whereas the downgrade effect increases the distance metric $\Delta M_{SVD}(\widehat{P}_{ML}, \widehat{P}_C)$, the duration dependence effect reduces it.

To test whether the two matrices are statistically different, the distributional properties of $\Delta M_{SVD}(\widehat{P}_{ML}, \widehat{P}_C)$ are obtained through a nonparametric bootstrap experiment along the lines of Hanson and Schuermann (2005), i.e. we pick realized firm rating-histories randomly and replace them in the pool until the number of firm-years equals the amount of total observed firm-years in the real data set, perform ML and cohort estimation, and repeat $B = 1,000$ times.[19] Table 6 shows the bootstrap results. Because the 95% confidence interval from the 2.5% to the 97.5% percentile does not include zero, the difference of 0.00047 is statistically significant at better than the 5% level. To control for the effect of temporal heterogeneity, we replicated all of our computations using only data from the period 1996-2000, excluding the recession years 2001 and 2002. Since this shorter period includes just 29 default events (see Table 2), the default probability estimates are much lower and the downgrade effect is not of particular importance, i.e. the default probability estimates for grade 6 are 0.78% (ML) and 0.68% (cohort). Therefore, because of the low

[18]A logit analysis performed on all 9,895 rating pairs confirmed this conjecture. In particular, the binary dependent variable CHANGE takes the value 1 if a rating change occurred between two subsequent monitoring steps. The dummies SHORT and LONG indicate whether the interexamination time is short (<12 months) or long (>12 months), respectively. The estimated coefficients are (p-values in parentheses): -0.6644 (0.001) for SHORT and -0.8897 (0.000) for LONG.

[19]In contrast to Hanson and Schuermann (2005), since we resample the same number of rating-histories from each stratum (defined as histories of equal length), the total number of firm-years does not vary across the bootstrap samples.

downgrade effect, we should expect a negative value for the distance metric. This conjecture is confirmed by the results given in Table 6.

As an illustration to assess the economic relevance of the difference in migration matrices, we look at credit risk capital levels implied by a one-factor CreditMetrics®-type portfolio model (see, e.g., Gordy, 2000) applied to stylized portfolios in which each firm is associated with a fictitious three-year, $100, coupon paying loan and the year-end portfolio values are computed using the two different migration matrices. Important characteristics of the sample portfolios are displayed in Table 7. Taking into account that the bank does not consider AAA equivalent ratings for its SME portfolio, internal ratings are mapped to the S&P scale such that corresponding default probabilities approximately agree. Credit spreads and the one-year forward zero-coupon U.S. Treasury yield curve that prevailed on July 19, 2005 are taken and a 40% recovery rate is assumed. Coupons are assigned such that pricing differs not to significantly from par = $100.

Exploiting the fact that for equal-sized loans the credit quality of a portfolio is governed by the proportion of firms in each rating grade, we construct three different portfolios: a "high quality" portfolio with the rating distribution taken from the year 1999, an "average quality" portfolio where the rating distribution equals the weighted average number of firms in each rating grade over the period 1996-2002, and a "low quality" portfolio using the distribution from 2002. The remaining columns in Table 7 show the weights of the single systematic risk factor, which equal the square root of the asset return correlation of two firms with the same rating. These weights or sensitivities can be derived separately for the ML and cohort approach by calculating a transition probability matrix for each of the six one-year periods from 1996-2002, and then estimating means and variances of rating class default probabilities from these period-specific (i.e. conditional) matrices. The last column shows the weights reported in Table 2 of Gordy (2000), which are for the majority of grades greater than the corresponding ML or cohort weights.

We summarize our findings in Table 8 which displays for the varying credit quality distributions the corresponding mean and standard deviation of portfolio horizon value and VaR (value-at-risk) figures at 99% and 99.9%, calculated separately for each of the three factor sensitivity sets. As can be seen, the largest deviations (up to 13%) from 100% of the VaR ratios (cohort to ML) occur when using Gordy's factor loadings, resulting from the fact that they imply higher default correlations under the assumptions of the one-factor model, especially for the grades 2, 4, 5, and 6. Because the economic impact of default is serve, much more so than a downgrade to some other rating, and the cohort method actually tends to overestimate default probabilities, except for the riskiest grade 6, the ratio of capital implied by the cohort and ML method will increase for a given portfolio quality distribution when the default correlations for grades 2-5 are rising. In contrast, a higher default correlation for firms in grade 6 decreases, *ceteris paribus,* the capital ratio.[20] However, application of the ML and cohort factor weights also leads to differences in risk capital which are economically meaningful, ranging in most cases between 2% and 7%. Finally note that for all credit quality distributions and sets of factor weights, the differences in portfolio mean horizon value change very little, whereas the differences in risk capital are often substantial.

5.2 Business cycle effects and non-Markov behavior

The model considered so far for the rating migration process makes the following quite specific assumptions: (i) cross-sectional and temporal homogeneity of transition intensities, and (ii) first-order Markov property, i.e. transitions from each rating class are independent of the process history. Assumption (i) may be checked by including covariates and time dummies in the modeling process. Similar to Lando and Skødeberg (2002) we use a form of a proportional hazards model in which the transition intensity matrix elements $q_{ll'}$ are parameterized with

$$q_{ll'}\left(Z_{ij}\right) = q_{ll'}^{(0)} \exp\left(\beta_{ll'}' Z_{ij}\right),\ l \neq l' \qquad (11)$$

where Z_{ij} denotes a vector of, possible time dependent, covariates. If the product of covariates and regression parameters is non-zero, then the intensities deviate from the baseline $q_{ll'}^{(0)}$. Thus, if the rating migration process exhibits non-homogeneity the regression coefficients are significantly different from zero. In contrast to Lando and Skødeberg (2002), where the baseline intensities are unspecified functions of time, we have to assume that the $q_{ll'}^{(0)}$ are constant in order to reduce the number of parameters to estimate. We will test for time-homogeneity by including the dummy RECESSION into model (11). RECESSION equals 1 if the rating observation falls into the recession years 2001 or 2002 and 0 otherwise.

[20]Recall that in the context of the one-factor model, the joint default probability of two borrowers having the same rating is a monotonically increasing and convex function with respect to the rating class default probability and the asset return correlation (i.e. the square of the factor weight). Looking upon portfolio quality, the (cohort to ML) VaR ratio is also decreasing, *ceteris paribus*, if the percentage of low quality (i.e. grade 6) firms is increasing.

Assessing assumption (ii) by accounting for the process history is difficult with merely discrete data available as the process is only observed through a series of snapshots. For example, we cannot calculate the time spend in the current rating in the absence of data on exact transition times. Furthermore, we do not know whether the borrower was upgraded or downgraded into the present rating class because nothing is known about the migration behavior between observation times. The most we can do is to condition the model on the last observed rating change using the variables UPGRADE and DOWNGRADE, defined in Table 3. In this way, it is examined if intensities are different for borrowers showing a positive, a negative or a stable rating tendency since the last observed rating.

The results for a model with the three dummy variables RECESSION, UPGRADE and DOWNGRADE are presented in Table 9. The parameter of interest is the regression coefficient $\beta_{ll'}$ which gives the linear effect on the log transition intensity $\ln(q_{ll'})$ of the respective covariate. As in Lando and Skødeberg (2002), we also consider only transitions to a neighboring rating class to get enough observations for meaningful inference. This comprises 78.17% of all recorded transitions. The covariate effects on non-neighboring transitions are constrained to zero.

Table 9 shows that downgrade intensities are significantly increased and upgrade intensities are significantly reduced when entering the recession period 2001-2002. This is consistent with an objective of maintaining a certain one-year PD for each rating grade through time. Interestingly, downgrade intensities are higher for firms with a favorable past rating pattern and upgrade intensities are lower. In addition, the results for DOWNGRADE are almost mirror-inverted, i.e. downgrade intensities are decreased and upgrade intensities are increased when a firm has experienced a recent downward rating pattern. In sum, Table 9 presents strong evidence for rating reversal activity - an upgrade followed by a downgrade, or a downgrade followed by an upgrade. This is in contrast to the "rating reversal aversion" of rating agencies: for example, Moody's takes a rating action only "when it is unlikely to be reversed within a relatively short period of time" (see Cantor, 2001, p. 175).[21] However, rating reversal activity is in line with structural rating models

[21] Fledelius et al. (2004) give empirical evidence consistent with this stated objective of rating agencies. Simulations conducted by Löffler (2005) show that the wish to avoid frequent reversals of credit ratings could account for some of the stylized facts of agency ratings such as their relative stability or the serial correlation and predictability of rating changes. In addition, Altman and Rijken (2004) confirm that the agency-rating migration policy is an even more important factor underlying agency-rating stability than their focus on long-term investment horizons.

of the type presented in Gordy and Heitfield (2001) or Löffler (2005). These authors model ratings as a mapping of a continuous credit quality variable (e.g., the borrower's distance to default) into discrete categories. A bias towards rating reversals can then be explained by changes in the credit quality variable following a bell-shaped distribution, i.e. a distribution whose density is declining monotonically towards the tails.

Since observing rating volatility, business cycle dependence and relatively frequent rating reversals is evidence of a point-in-time rating system, this paper finds that internal ratings are more point-in-time than external ratings.

6 Concluding remarks

Rating migration matrices, which measure the expected changes in credit quality of borrowers, are cardinal inputs to many applications, including credit portfolio monitoring and management, capital allocation and pricing of credit derivatives. Moreover, since the new Basel capital accord will allow certain financial institutions to use an Internal Ratings-Based (IRB) approach to determine the capital requirements for a given exposure, bank management and regulators are keenly interested in accurately estimated transition probabilities associated with internal rating grades. Against this background, the findings of this paper have some important implications for both, risk management and banking supervision.

First, since banks base, for efficiency reasons, the decision about rating review frequencies on borrower characteristics, the use of the cohort method gives misleading estimates, especially for the diagonal elements and the last column of the migration matrix. This will translate into inaccurate credit risk capital levels estimated by credit portfolio models. Second, with respect to the possible acceptance of internal models for the determination of regulatory capital requirements in the future, regulators as well should be aware of the shortcomings of the cohort method. However, the current version of the Basel capital accord promotes the cohort approach. In particular, IRB banks are required to use the "long-run average of one-year realized default rates" as estimator for grade-specific default probabilities (Basel Committee on Banking Supervision, 2004, § 447).

Besides estimating transition probabilities, this paper also contributes to the debate about internal ratings and their properties. We find that internal ratings are more point-in-time than external ratings, which implies that up- and downgrade intensities vary with the business cycle, consistent with an objective of maintaining a certain one-year PD for each rating grade through time. In addition, this effect also induces a non-Markov tendency for rating reversals (contrary to the rating reversal aversion seen in external rating data). Both effects undermine the assumptions of any of the estimators discussed above and thus leave open some important questions for future research on the analysis of internal ratings.

7 References

Altman, E.I., Rijken, H.A., 2004. How rating agencies achieve rating stability. Journal of Banking and Finance 28, 2679-2714.

Basel Committee on Banking Supervision, 2000. Range of practice in banks' internal rating systems. Discussion paper, January 2000, Basel.

Basel Committee on Banking Supervision, 2004. International convergence of capital measurement and capital standards. June 2004, Basel.

Blackwell, D.W., Winters, D.B., 1997. Banking relationships and the effect of monitoring on loan pricing. Journal of Financial Research 20, 275-289.

Cantor, R., 2001. Moody's Investors Service response to the consultative paper issued by the Basel Committee on Banking Supervision and its implications for the rating agency industry. Journal of Banking and Finance 25, 171-185.

Cox, D.R., Miller, H.D., 1965. The theory of stochastic processes. Chapman & Hall, London.

Delianedis, G., Geske, R., 1999. Credit risk and risk neutral default probabilities: Information about rating migrations and defaults. Working paper, University of California, Los Angeles.

Diamond, D.W., 1989. Reputation acquisition in debt markets. Journal of Political Economy 97, 828-862.

Diamond, D.W., 1991. Monitoring and reputation: The choice between bank loans and directly placed debt. Journal of Political Economy 99, 689-721.

Dietsch, M., Petey, J., 2004. Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs. Journal of Banking and Finance 28, 773-788.

Fledelius, P., Lando, D., Nielsen, J.P., 2004. Non-parametric analysis of rating transition and default data. Journal of Investment Management 2, 71-85.

Gordy, M., 2000. A comparative anatomy of credit risk models. Journal of Banking and Finance 24, 119-149.

Gordy, M., Heitfield, E., 2001. Of Moody's and Merton: A structural model of bond rating transitions. Working paper, Board of Governors of the US Federal Reserve System.

Hanson, S., Schuermann, T., 2005. Confidence intervals for probabilities of default. Working paper, Federal Reserve Bank of New York.

Jafry, Y., Schuermann, T., 2004. Measurement, estimation and comparison of credit migration matrices. Journal of Banking and Finance 28, 2603-2639.

Kalbfleisch, J.D., Lawless, J.F., 1985. The analysis of panel data under a markov assumption. Journal of the American Statistical Association 80, 863-871.

Kalbfleisch, J.D., Lawless, J.F., Vollmer, W.M., 1983. Estimation in markov models from aggregate data. Biometrics 39, 907-919.

Kay, R., 1986. A markov model for analyzing cancer markers and disease states in survival studies. Biometrics 42, 855-865.

Lando, D., Skødeberg, T.M., 2002. Analyzing rating transitions and rating drift with continuous observations. Journal of Banking and Finance 26, 423-444.

Lawrence, E.C., 1983. Reporting delays for failed firms. Journal of Accounting Research 21, 606-610.

Löffler, G., 2005. Avoiding the rating bounce: Why rating agencies are slow to react to new information. Journal of Economic Behavior and Organization 56, 365-381.

Machauer, A., Weber, M., 1998. Bank behavior based on internal credit ratings of borrowers. Journal of Banking and Finance 22, 1355-1383.

Manove, M., Padilla, J.A., Pagano, M., 2001. Collateral versus project screening: A model of lazy banks. RAND Journal of Economics 32, 726-744.

Mester, L.J., Nakamura, L.I., Renault, M., 2002. Checking accounts and bank monitoring. Working paper, Federal Reserve Bank of Philadelphia.

Rajan, R., Winton, A., 1995. Covenants and collateral as incentives to monitor. Journal of Finance 50, 1113-1146.

Treacy, W.F., Carey, M., 2000. Credit risk rating systems at large US banks. Journal of Banking and Finance 24, 167-201.

Udell, G. F., 1989. Loan quality, commercial loan review and loan officer contracting. Journal of Banking and Finance 13, 367-382.

Whittred, G., Zimmer, I., 1984. Timeliness of financial reporting and financial distress. The Accounting Review 59, 287-295.

Table 1: Risk/size distribution (1996-2002)

Rating class	Size class (turnover in € Mio.)											
	[2.5-5]		(5-10]		(10-25]		(25-50]		>50		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
1	8	0.21	39	1.26	348	14.07	198	15.84	255	15.69	848	6.92
2	1,051	27.53	925	29.97	793	32.07	395	31.60	579	35.63	3,743	30.55
3	1,013	26.53	852	27.61	487	19.69	243	19.44	331	20.37	2,926	23.88
4	1,082	28.34	820	26.57	386	15.61	219	17.52	282	17.35	2,789	22.76
5	496	12.99	357	11.57	255	10.31	127	10.16	110	6.77	1,345	10.98
6	114	2.99	69	2.24	186	7.52	61	4.88	61	3.75	491	4.01
Default	54	1.41	24	0.78	18	0.73	7	0.56	7	0.43	110	0.90
Total	3,818	100.00	3,086	100.00	2,473	100.00	1,250	100.00	1,625	100.00	12,252	100.00

Table 2: Rating distribution over time. The observation time for each rating is based on the date the credit department signs off on the rating review. Note that the default count for the year 2002 also includes seven defaults from the first month of 2003.

Year	Rating class (from low risk - class 1 - to high risk - class 6)															
	1		2		3		4		5		6		Default		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
1996	83	6.61	406	32.35	277	22.07	277	22.07	159	12.67	49	3.90	4	0.32	1,255	100
1997	96	6.84	414	29.49	342	24.36	330	23.50	151	10.75	68	4.84	3	0.21	1,404	100
1998	109	6.87	467	29.43	385	24.26	406	25.58	160	10.08	56	3.53	4	0.25	1,587	100
1999	135	7.55	539	30.15	448	25.06	418	23.38	184	10.29	63	3.52	1	0.06	1,788	100
2000	145	7.21	624	31.04	489	24.33	438	21.79	225	11.19	72	3.58	17	0.85	2,010	100
2001	141	6.59	644	30.09	522	24.39	466	21.78	246	11.50	93	4.35	28	1.31	2,140	100
2002	119	5.75	619	29.93	463	22.39	454	21.95	256	12.38	104	5.03	53	2.56	2,068	100
Total	848	6.92	3,743	30.55	2,926	23.88	2,789	22.76	1,345	10.98	491	4.01	110	0.90	12,252	100

Table 3: Definitions of variables and hypotheses

Variable	Definition	Hypothesis
Rating process		
MONTHS	Number of months between t_{ij} and t_{ij+1}	
MONTHSCAT	Ordinal variable: 0 if MONTHS$<$12, 1 if MONTHS=12, 2 if MONTHS$>$12	
Risk		
R1 to R5	Rating-Dummies, 1 is best and 6 is worst	Positive
UPGRADE	Dummy, 1 if the rating class difference between t_{ij} and t_{ij+1} is $>$0	Ambiguous
DOWNGRADE	Dummy, 1 if the rating class difference between t_{ij} and t_{ij+1} is $<$0	Negative
Size		
LOGSIZE	Natural logarithm of the firm's annual sales	U-shaped
EXPOSURE	Natural logarithm of the balance sheet amount of debt supplied by the bank	Negative
Relationship		
HOUSEBANK	Debt supplied by the bank divided by total bank debt	Positive
DURATION	Temporal length (in years) of bank-borrower relationship	Positive
Governance		
LIMLIAB	Dummy, 1 if the firm is incorporated	Negative
Years		
Y96, Y97, Y98, Y99, Y00, Y01	Dummies, 1 if t_{ij} is in year 1996, 1997, 1998, 1999, 2000 or 2001	Ambiguous

Table 4: Determinants of interexamination time. Random-effects ordered probit regression based on 5,640 rating pairs for the period 1996-2002, dependent variable: MONTHSCAT. For definitions of variables see Table 3. The intraclass correlation ρ is given by $\sigma_v^2/(\sigma_v^2+1)$, where σ_v^2 denotes the panel-level variance component.

Variable	Coefficient	p-value
R1	-0.9905	0.0477
R2	-1.4196	< 0.0001
R3	-1.1395	0.0002
R4	-1.0323	0.0002
R5	-0.5600	0.0404
R6	Reference	Reference
UPGRADE	0.4082	0.0051
DOWNGRADE	0.2230	0.1416
LOGSIZE	-4.5591	< 0.0001
LOGSIZE^2	0.2070	< 0.0001
EXPOSURE	0.0166	0.7263
HOUSEBANK	-0.0964	0.6384
DURATION	0.1680	0.0002
LIMLIAB	-0.2961	0.0949
Y96	-0.0399	0.8493
Y97	-0.0226	0.9124
Y98	-0.0058	0.9761
Y99	0.0380	0.8729
Y00	0.8837	< 0.0001
Y01	Reference	Reference
ρ	0.5479	< 0.0001

Table 5: Transition probabilities: comparing estimation methods. Both transition matrices are estimated over the period 1996-2002. However, whereas the ML method uses all available rating observations, the cohort method is applied to only the 9,180 rating pairs (92.77% of 9,895) with an interexamination time of 12 months. This results in a loss of 761 single rating observations, i.e. 6.2% of 12,252. Standard errors, calculated using a nonparametric bootstrap approach with $B = 1,000$ replications, are given in parentheses. Diagonal entries are bolded for convenience.

From	To						Def.
	1	2	3	4	5	6	
Panel A. The one-year ML transition matrix							
1	**0.6614**	0.2547	0.0417	0.0309	0.0058	0.0036	0.0020
	(0.0207)	(0.0159)	(0.0033)	(0.0062)	(0.0015)	(0.0023)	(0.0027)
2	0.0638	**0.6802**	0.1798	0.0599	0.0124	0.0034	0.0005
	(0.0044)	(0.0101)	(0.0084)	(0.0049)	(0.0021)	(0.0009)	(0.0004)
3	0.0139	0.2363	**0.4737**	0.2246	0.0397	0.0106	0.0011
	(0.0021)	(0.0098)	(0.0101)	(0.0088)	(0.0036)	(0.0020)	(0.0003)
4	0.0068	0.0784	0.2341	**0.4960**	0.1466	0.0344	0.0037
	(0.0016)	(0.0058)	(0.0077)	(0.0101)	(0.0077)	(0.0046)	(0.0015)
5	0.0025	0.0343	0.0809	0.2870	**0.4432**	0.1331	0.0189
	(0.0008)	(0.0050)	(0.0046)	(0.0115)	(0.0171)	(0.0140)	(0.0096)
6	0.0007	0.0147	0.0460	0.0896	0.2406	**0.4624**	0.1459
	(0.0003)	(0.0033)	(0.0085)	(0.0114)	(0.0244)	(0.0265)	(0.0305)
Def	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**1.0000**
Panel B. The one-year cohort transition matrix							
1	**0.6643**	0.2654	0.0301	0.0316	0.0043	0.0029	0.0014
	(0.0208)	(0.0171)	(0.0059)	(0.0068)	(0.0024)	(0.0022)	(0.0016)
2	0.0646	**0.6786**	0.1823	0.0585	0.0118	0.0030	0.0010
	(0.0045)	(0.0104)	(0.0087)	(0.0049)	(0.0022)	(0.0010)	(0.0006)
3	0.0147	0.2394	**0.4727**	0.2247	0.0385	0.0082	0.0017
	(0.0026)	(0.0102)	(0.0104)	(0.0097)	(0.0046)	(0.0017)	(0.0009)
4	0.0077	0.0769	0.2437	**0.4919**	0.1433	0.0267	0.0099
	(0.0019)	(0.0057)	(0.0087)	(0.0108)	(0.0075)	(0.0036)	(0.0024)
5	0.0019	0.0382	0.0763	0.3111	**0.4408**	0.1059	0.0258
	(0.0015)	(0.0059)	(0.0085)	(0.0135)	(0.0177)	(0.0102)	(0.0049)
6	0.0000	0.0217	0.0596	0.1084	0.2954	**0.4770**	0.0379
	(0.0000)	(0.0071)	(0.0116)	(0.0179)	(0.0230)	(0.0296)	(0.0088)
Def	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**1.0000**

Table 6: Bootstrap results (mean, standard deviation, and several quantiles) for the distance metric $\Delta M_{SVD}(\widehat{P}_{ML}, \widehat{P}_C)$ between ML and cohort method. The number of replications for the nonparametric bootstrap experiment is $B = 1,000$.

	Period	
$\Delta M_{SVD}(\widehat{P}_{ML}, \widehat{P}_C)$	1996-2002	1996-2000
observed	0.00047	-0.00593
bootstrap distribution		
Mean	0.00051	-0.00545
Std. Dev.	0.00036	0.00167
Q_1	-0.00008	-0.00921
$Q_{2.5}$	0.00001	-0.00872
Q_5	0.00002	-0.00796
Q_{50}	0.00048	-0.00521
Q_{95}	0.00126	-0.00228
$Q_{97.5}$	0.00130	-0.00145
Q_{99}	0.00135	0.00011

Table 7: Characteristics of sample portfolios and calibration of portfolio model. All portfolios consist of fictitious three-year, \$100, coupon paying loans (one per firm) and the year-end values are computed using the ML and cohort migration matrices for the period 1996-2002. Internal ratings are mapped to external ones as shown in the second column. Coupons are assigned such that pricing differs not to significantly from par = \$100. Credit quality refers to the distribution of firms across rating grades for the years 1999 ("High"), 2002 ("Low") and the overall period 1996-2002 ("Average"). Factor weights are calculated in the standard way (Gordy, 2000, Dietsch and Petey, 2004), estimating (conditional) ML and cohort rating class default probabilities for each of the six one-year sample periods. The last column shows the weights reported by Gordy (2000).

Internal Rating	S&P scale	Coupon (annual)	Portfolio credit quality			Factor loadings		
			High (1999)	Average (1996-2002)	Low (2002)	ML	Cohort	Gordy (2000)
1	AA	3	135	122	119	0.316	0.289	0.285
2	A	3.5	539	547	619	0.202	0.167	0.279
3	BBB	4	448	432	463	0.080	0.156	0.121
4	BB	5	418	409	454	0.132	0.090	0.354
5	B	6.5	184	204	256	0.199	0.081	0.255
6	CCC	9	63	75	104	0.133	0.106	0.277
Total			1,787	1,789	2,015			

Table 8: Credit risk capital: comparing estimation methods. Credit risk capital for a one-year horizon as computed by a one-factor version of CreditMetrics®. Credit spreads and the one-year forward zero-coupon U.S. Treasury yield curve that prevailed on July 19, 2005 are taken and a 40% recovery rate is assumed. The sample portfolios and the three sets of factor weights are as described in Table 7. Monte Carlo simulations (100,000 trials) are used to obtain the portfolio value distributions one year hence.

	Factor loadings								
	ML			Cohort			Gordy (2000)		
	Cohort	ML	$\% \frac{\text{Cohort}}{\text{ML}}$	Cohort	ML	$\% \frac{\text{Cohort}}{\text{ML}}$	Cohort	ML	$\% \frac{\text{Cohort}}{\text{ML}}$
High quality portfolio (1999)									
Mean horizon value ($)	183,389	183,085	100.17%	183,388	183,084	100.17%	183,382	183,080	100.16%
Std. dev. of value ($)	634	642	98.75%	484	506	95.65%	1,119	1,100	101.73%
VaR (99%) ($)	1,834	1,808	101.44%	1,331	1,369	97.22%	3,662	3,393	107.93%
VaR (99.9%) ($)	2,711	2,654	102.15%	1,945	1,960	99.23%	6,215	5,489	113.23%
Average quality portfolio (1996-2002)									
Mean horizon value ($)	183,747	183,351	100.22%	183,747	183,351	100.22%	183,748	183,353	100.22%
Std. dev. of value ($)	666	689	96.66%	497	534	93.07%	1,153	1,169	98.63%
VaR (99%) ($)	1,929	1,925	100.21%	1,364	1,427	95.59%	3,785	3,599	105.17%
VaR (99.9%) ($)	2,906	2,835	102.50%	1,962	2,046	95.89%	6,236	5,798	107.55%
Low quality portfolio (2002)									
Mean horizon value ($)	207,204	206,608	100.29%	207,203	206,607	100.29%	207,212	206,617	100.29%
Std. dev. of value ($)	792	841	94.17%	577	639	90.30%	1,363	1,433	95.12%
VaR (99%) ($)	2,274	2,330	97.60%	1,573	1,696	92.75%	4,451	4,384	101.53%
VaR (99.9%) ($)	3,427	3,454	99.22%	2,284	2,383	95.85%	7,287	6,835	106.61%

Table 9: Results for the test of business cycle effects and non-Markov behavior of transition intensities to a neighboring rating class. The third, sixth and ninth column report the estimate of β for the RECESSION, UPGRADE and DOWNGRADE dummy, respectively. A positive (negative) β implies that the transition intensity is increased (decreased) by a factor of $\exp(\beta)$ when the corresponding dummy takes on the value 1.

Ratings		RECESSION			UPGRADE			DOWNGRADE		
From	To	$\widehat{\beta}$	Std. Err.	p-value	$\widehat{\beta}$	Std. Err.	p-value	$\widehat{\beta}$	Std. Err.	p-value
Panel A. Downgrade intensities										
1	2	0.2511	0.0421	<0.0001	0.8638	0.2117	<0.0001	–	–	–
2	3	0.2174	0.0309	<0.0001	0.8534	0.1985	<0.0001	-0.2654	0.0899	0.0032
3	4	0.3115	0.0498	<0.0001	0.5819	0.1472	<0.0001	-0.2303	0.0446	<0.0001
4	5	0.4766	0.0964	<0.0001	0.6980	0.1094	<0.0001	0.0740	0.0532	0.1642
5	6	0.7670	0.1436	<0.0001	0.3276	0.0884	0.0002	-0.1233	0.0362	0.0007
6	Def.	1.7665	0.3097	<0.0001	–	–	–	-0.1983	0.0631	0.0017
Panel B. Upgrade intensities										
2	1	0.2531	0.2271	0.2651	0.0416	0.0559	0.4568	1.1774	0.3462	0.0007
3	2	-0.1701	0.0496	0.0006	0.1424	0.2097	0.4971	0.6387	0.2097	0.0023
4	3	-0.4163	0.1004	<0.0001	-0.0747	0.0382	0.0505	0.6114	0.1968	0.0019
5	4	-0.5093	0.0896	<0.0001	-0.1081	0.0563	0.0548	0.7142	0.2437	0.0034
6	5	-0.6731	0.1552	<0.0001	–	–	–	0.4733	0.1506	0.0017