Bayesian Inference for Issuer Heterogeneity in Credit Ratings Migration

Ashay Kadam*

Cass Business School, City University, London.

Peter Lenk

Ross Business School, University of Michigan.

First Version: 17th October 2005

This Version: January 22, 2007

JEL Classification : C11,C13,C41,G12

Keywords: Credit risk, Risk Capital, Markov Chains, Bayesian Inference, Heterogeneity

*Corresponding author. Address: 106 Bunhill Row, London EC1Y 8TZ, United Kingdom. Email:A.Kadam@city.ac.uk Phone: +44-20-7040-8632. We are grateful to seminar participants at Quantitative Methods in Finance conference (University of Technology Sydney, Dec 2005) and Credit Risk Workshop (Cass Business School, London, May 2006) for providing valuable feedback. Helder Palaro provided excellent research assistance. The paper has benefitted immensely from valuable comments by two anonymous referees. We accept responsibility for all remaining errors.

Bayesian Inference for Issuer Heterogeneity in Credit Ratings Migration

Abstract

We explore sources of heterogeneity in rating migration behavior using a continuous time Markov chain. Working in continuous time circumvents the embedding problem, mitigates the censoring effect and facilitates term structure modelling with arbitrary prediction horizons. By adopting a Bayesian estimation procedure we are able to estimate for each issuer profile its own continuous time Markov chain generator. Using the Moodys corporate bond default database we identify significant country and industry effects on the determination of default intensity and conditional transition probabilities in general. We tabulate and compare these quantities for different issuer profiles to assess the heterogeneity in the sample. Using the Jafry-Schuermann mobility metric we show how distant the transition probability matrices are for different issuer profiles. Using the CreditRisk+ framework, and a sample credit portfolio, we find that ignoring heterogeneity can give erroneous estimates of VAR and a misleading picture of the risk capital.

1 Introduction

A time-homogenous, discrete-time Markov chain has been extensively used to model the ratings migration process for corporate bonds and bond issuers[1]. Such modelling has often further assumed that the rated entities are homogeneous with respect to their rating migration behavior. Deviation from this added assumption has been the subject of several studies that highlight sources of heterogeneity such as the issuer's age, country of domicile, stage in the business cycle etc[2]. Independent of this deviation, the ratings migration literature has witnessed departures from the discrete time framework as well[3]. In this paper we combine these two parallel developments in the rating migration literature, while using Bayesian techniques to estimate a flexible ratings based model of default risk.

We explore sources of heterogeneity in rating migration behavior using a continuous time Markov chain based model. In that sense our modelling framework is similar to Frydman and Kadam (2004) and Frydman and Schuermann (June 2005). Both of these apply Markov chain based mixture models to ratings data.[4] However, they use Maximum Likelihood Estimation to obtain the relevant Markov chain generators for sub-populations of rated entities. By adopting a Bayesian estimation procedure

[1]Having accepted this model, the actual reported transition probability matrices can vary a lot depending on the actual data and estimation methodology used, see Altman (1998) for a detailed discussion on the popular methods used in practice.

[2]See for instance Asquith, Mullins, and Wolff (1989), Nickell, Perraudin, and Varotto (2000), Frydman and Kadam (2004), Frydman and Schuermann (June 2005), Bangia, Diebold, Kronimus, Schagen, and Schuermann (2002), Lucas and Lonski (1992) and research summary reports published by rating agencies such as Moody's KMV, Standard & Poor's and Fitch on their web-sites.

[3]Jarrow, Lando, and Turnbull (1997) were among the first to fit a continuous time Markov chain model to observed bond prices.

[4]See Norris (1997) for an elaborate treatment of Markov Chains.

we are able to estimate for each issuer its own individual continuous time Markov chain generator. Bayesian techniques are particularly suitable for situations such as this where we have very sparse data for several rating categories and issuer types.[5] We condition the generator of the Markov chain on a joint event that can incorporate the issuer's country of domicile, industry type (sector) and potentially several other qualifiers. Since there are many possible issuer attributes and many possible values that each of these attributes take the total number of issuer profiles can be quite large. The methods we propose permit an estimate for each such profile, and in principle one can think of this as if each issuer had its own generator. Being able to explicitly recognize the heterogeneity in the issuer pool gives us a clearer picture of both Value at Risk and risk capital, as we illustrate with a hypothetical credit portfolio. It is noteworthy that we provide estimates for an arbitrary profile even if data on that profile may be a very small part of the sample we use for estimation.[6]

Our empirical results build upon the work of Nickell, Perraudin, and Varotto (2000). They offer, for each qualifier of interest (e.g. country of domicile), a conditional transition matrix (over a given time period) estimated by conditioning on values taken by that variable (e.g. USA, UK and Japan) having controlled for other sources of variation (e.g. industry type). Just as they provide conditional estimates for a discrete time Markov chain that models the evolution of ratings, we compute conditional estimates for duration times and transition rates for a continuous time Markov chain that models

[5] See footnote 23 for some evidence of sparsity in this context.

[6] For instance the rating evolution for Japanese issuers in the Utility sector can be estimated although this type of issuers comprise only 0.1% of the data. This is made possible by combining the information on Japanese issuer transitions (3% of the sample) and on Utility sector issuer transitions (10% of the sample).

the evolution of issuer ratings. However, our model and estimation methodology are both different from theirs, as are the tools we use to demonstrate the heterogeneity.

In constructing the model, Nickell, Perraudin, and Varotto (2000) employ a Probit framework to compute conditional transition probabilities in discrete time. We model in continuous time, the state durations being exponential functions and transition probabilities being logistic functions. This has two advantages. First, by modelling the duration explicitly we provide a richer understanding of rating stability.[7] Second, by using a continuous time model we capture the usual advantages emphasized in Lando and Skodeberg (2002).[8]

In estimating the model, Nickell, Perraudin, and Varotto (2000) use classical methods, we use a Bayesian approach. Bayesian inference reduces estimation error and is more suitable for sparse data situations such as this.[9]

The dataset we use is the Moodys corporate bond default database. This rich dataset provides us with issuer rating histories from several countries and industry sectors, and spanning several decades.[10]

[7] Figure 1 clearly indicates that the variability in duration times is quite high both within and across rating categories. This is ample evidence to suggest that the average stay period in any given rating is not a reliable summary statistic. A key feature of this paper vs. any other discrete time model based paper (such as Nickell, Perraudin, and Varotto (2000)) is that duration times have a model that captures this large variability.

[8] A continuous time framework is better at handling censoring, it avoids the embedding problem for Markov chains and it makes term structure modelling easier.

[9] See section 4.1 for a more elaborate discussion on this point. See footnote 23 for evidence of data sparsity in this context.

[10] See Section 3 for a detailed description of the dataset.

2 Model

We model the changes in a issuer's rating over time as a discrete space, continuous time, stationary Markov process. These Markov processes can be represented by the duration time that the process is in a state and transition probabilities or jump distributions for a transition to a new state. The duration times are independent and exponentially distributed with rate parameters that depend on the issuer's current rating. At the end of the duration, the rating jumps to a new rating. The jumps and durations are mutually independent within an issuer. The states are indexed by $k = 1$, $\ldots$, $K+1$. The states are ordered such that the first state is AAA; the second state is AA, and so on. State K corresponds to the rating being withdrawn, and state $K+1$ is default, which is absorbing. We observe the ratings process for a set of issuers where i indexes the issuer for $i = 1, \ldots, M$. The observational time period is $a \leq t \leq b$.

During the observation period, issuer i has n_i transitions or changes in its ratings. The jth transition in the rating for issuer i occurs at time $T_{i,j}$ for $j = 1, \ldots, n_i$ where $a \leq T_{i,1} < \ldots < T_{i,n_i} \leq b$. At time t such that $T_{i,j} \leq t < T_{i,j+1}$ the issuer's rating is in state $s_{i,j}$. The rating then changes at time $T_{i,j+1}$ to $s_{i,j+1} \neq s_{i,j}$.

Because the observation period is a finite interval, we need to be careful about left and right truncation of the observed rating process. If the issuer was rated before the start of the observation period, then the initial transition $T_{i,0}$ precedes a, the start of the observation period, and has rating $s_{i,0}$. The rating continues in state $s_{i,0}$ from a to $T_{i,1}$. If the issuer is first rated after a, then $T_{i,0}$ and $s_{i,0}$ are not defined. Similarly,

some of the durations are right truncated. If the issuer does not default, then the n_i+1 transition occurs at time $T_{i,n_i+1} > b$ where b is the end of the observation period. If the issuer defaults in the observation period, then the process ends at the last transition T_{i,n_i} because default is an absorbing state, and T_{i,n_i+1} is undefined. The transition times $T_{i,0}$ and T_{i,n_i+1}, when they are defined, are not observed.

The duration times are defined from the transition times. The duration time for the jth transition for issuer i is: $D_{i,j} = T_{i,j+1} - T_{i,j}$ for $j = 1, \ldots, n_i - 1$. If the issuer is rated before a, the beginning of the observation period, then the initial duration is $D_{i,0} = T_{i,1} - T_{i,0}$. However, we only observe the left truncated duration $D^*_{i,0} = T_{i,1} - a$. If the issuer is first rated after a, then $D_{i,0}$ is undefined. If the issuer does not default in the observation period, then D_{i,n_i} is right truncated, and we only observe $D^*_{i,n_i} = b - T_{i,n_i}$. If issuer i defaults before b, then D_{i,n_i} is not defined because default is an absorbing state.

In discrete space, continuous time, stationary Markov processes, the duration times are mutually independent and exponentially distributed random variables. The density for duration $D_{i,j}$ is:

$$f(t|y_{i,j}) = \exp(-y_{i,j}) \exp\left[-\exp(-y_{i,j})t\right] \text{ for } t > 0, \tag{1}$$

with rate parameter $\exp(-y_{i,j})$ and expected value $E(D_{i,j}|y_{i,j}) = \exp(y_{i,j})$. Suppose that the issuer is in state $s = s_{i,j}$ during duration $D_{i,j}$. Our model for the $y_{i,j}$ is:

$$y_{i,j} = x_i'\beta_s + \phi_{i,D} + \epsilon_{i,j} \tag{2}$$

where x_i is a p-vector of covariates[11] for the issuers; β_s is a p-vector of regression coefficients; $\phi_{i,D}$ is a random effect for issuer i; and $\epsilon_{i,j}$ are error terms. Both the random effects and the error terms are mutually independent and normally distributed with mean zero. The variance of the error terms depends on the state s: $\mathrm{var}(\epsilon_{i,j}) = \sigma_s^2$. The variance of the random effect is λ_D^2.

The impact of the random effect can be seen by the conditional expected duration given the random effect:

$$E(D_{i,j}|\phi_{i,D}) \quad = \quad E[E(D_{i,j}|\phi_{i,D}, \epsilon_{i,j})] = \exp\left(x_i'\beta_s + \phi_{i,D} + \frac{\sigma_s^2}{2}\right). \tag{3}$$

The random effect $\phi_{i,D}$ expresses the issuer's "stickiness" to remain in a rating, compared to other issuers, after adjusting for the covariate x_i. If $\phi_{i,D}$ is positive, then the issuer i tends to have longer durations, while if $\phi_{i,D}$ is negative, it changes ratings faster than most issuers with the same covariate. The unconditional expected duration integrates $\exp(y_{i,j})$ over both the random effect and error term:

$$E(D_{i,j}) \quad = \quad \exp\left(x_i'\beta_s + \frac{\lambda_D^2 + \sigma_s^2}{2}\right) \tag{4}$$

[11] The covariates used for implementation were dummy variables to capture country and industry effects.

If issuer i was rated before a, the start of the observational period, we observe the left truncated duration $D^*_{i,0} = T_{i,1} - a$. Using the memoryless property of the exponential distribution, $D^*_{i,0}$ also has an exponential distribution with rate $\exp(-y_{i,0})$. If the issuer is not in default over the observation period, then D^*_{i,n_i} is right censored, and its contribution to the likelihood function is:

$$P(D^*_{i,n_i} > b - T_{i,n_i}) = \exp\left[-\exp(-y_{i,n_i})(b - T_{i,n_i})\right]. \tag{5}$$

At time $T_{i,j}$ the issuer has a transition from state $r = s_{i,j-1}$ to state $s = s_{i,j}$ where $r \neq s$. The transition probabilities are conditional on the previous state r. If $r = K + 1$ is the absorbing (default) state, then the process ends. One way to motivate the model for the jump distribution is through the random utility framework of McFadden (1974). The rating agency has a random utility $U_{i,j,k}$ for giving issuer i a rating or k on transition j. In revaluating issuer i, the rating agency selects the utility that maximizes the random utility. We assume that the random utility has the following model:

$$U_{i,j,k} = z_i'\alpha_{r,k} + \zeta_{i,j,k} \text{ for } k = 1, \ldots, K \text{ and } k \neq r \tag{6}$$

$$U_{i,j,K+1} = \phi_{i,A} + \zeta_{i,j,K+1} \text{ for the default of absorbing state,} \tag{7}$$

where z_i is a q-vector of covariates[12] for issuer i; $\alpha_{r,k}$ is a q-vector of coefficients; $\phi_{i,A}$ is a random effect that measures propensity of the issuer to default; and $\zeta_{i,j,k}$ are mutually

[12] The covariates used for implementation were dummy variables to capture country and industry effects. The z_i covariates for model implementation were identical to the x_i covariates. This choice is by convenience, and not a restriction imposed by either the model or the estimation method.

independent error terms that have an extreme value distribution where, without loss of generality, the scale parameter is one.[13] The new rating for the issuer i is $s = \arg\max_{k \neq r} U_{i,j,k}$. The the probabilities of jumping from r to s are logistic functions:

$$P(r|r,i) = 0 \tag{8}$$

$$P(s|r,i) \propto \exp(z_i' \alpha_{r,s}) \text{ for } s = 1, \ldots, K \text{ and } s \neq r \tag{9}$$

$$P(K+1|r,i) \propto \exp(\phi_{i,A}) \text{ for the absorbing (default) state } K+1. \tag{10}$$

The random effects captures individual differences in the issuers' default rates, as can be seen by the log-odds ratio of defaulting:

$$\ln[P(K+1|r,i)] - \ln[P(s|r,i)] = \phi_{i,A} - z_i' \alpha_{r,s} \text{ for } s \neq r. \tag{11}$$

If $\phi_{i,A}$ is positive, the issuer is more likely, after adjusting for its covariates, to default than comparable issuers, while if it is negative, the issuer is less likely to default.

The random effects $\phi_i = (\phi_{i,A}, \phi_{i,D})'$ for issuer i are random samples from a mean-zero, bivariate normal distribution with covariance matrix:

$$\Lambda = \begin{bmatrix} \lambda_A^2 & \lambda_{AD} \\ \lambda_{AD} & \lambda_D^2 \end{bmatrix}.$$

[13] As a technical note, the utility for default does not include z_i in order to identify the model: preference structures are invariant to location and scale transformations of the utilities.

Given the random effects, the duration times and jump process are independent within an issuer. However, if one integrates over ϕ_i, then the duration times and jump process are correlated. A positive (negative) covariance implies that issuers that tend to remain in a rating state longer tend to have higher (lower) default rates.

Stationary Markov processes are also defined by their generators. The generator for the rating process of issuer i depends on the value of the covariates, the random effects, and the error terms. In a slight abuse of notation, $\epsilon_{i,s}$ is the error term for the ln-rate model for durations when the issuer is in state s, and $y_{i,s} = x_i'\beta_s + \phi_{i,D} + \epsilon_{i,s}$. The generator for issuer i is a K by $K+1$ matrix:

$$Q_i(\phi_i, \epsilon_i, x_i, z_i) = \begin{cases} -\exp(-y_{i,j}) & \text{for the } (j,j) \text{ element and } j = 1, \ldots, K \\ \exp(-y_{i,j})P(k|j,i) & \text{for the } (j,k) \text{ element where} \\ & j = 1, \ldots K;\ k = 1, \ldots, K+1;\ j \neq k. \end{cases} \tag{12}$$

The generator does not have a row for state $K+1$ because it is absorbing.[14] In using the generator, say in portfolio applications to compute default rates, one may not have estimates of the random effects for the issuers of interest. In this case, the random effects and error terms can be integrated out of the generator by Monte Carlo by generating G random deviates $\phi_i^{(g)}$ from a bivariate normal distribution with mean 0 and covariance matrix Λ and by generating the $\epsilon_{i,s}^{(g)}$ from normal distribution with mean 0 and variance σ_s^2. Then, the Monte Carlo approximation of the integrated generator

[14] If, for the sake of completeness, a row is included for this state then all entries in that row are zero.

is

$$Q_i(x_i, z_i) \approx \frac{1}{G} \sum_{g=1}^{G} Q_i(\phi_i^{(g)}, \epsilon_i^{(g)}, x_i, z_i).$$

3 Data Description

The dataset we use is the entire Moodys corporate bond default database available as of late 2005.[15] This rich dataset provides us rating histories from 112 countries and 14 industry sectors. The model implementation uses dummy variables for countries and industries as covariates in the duration and transition models. To improve both execution speed and output interpretation, it is desirable to have fewer countries and industry sectors for the model implementation. To this end, we eliminate those countries and industry sectors that have a very small number of rating transitions. In doing so we first merge all countries in the European Union and treat it as one country EU. This leaves us with the countries USA, UK, Japan, Canada and EU. To fit the model, we focus on the following 7 industry sectors : Banking, Utility, Insurance, Transport, Government, Finance and Real Estate Finance, eliminating the remainder which contain a very small fraction of the data. About 15% data was discarded in this process.[16]

Table 1 gives the composition of this smaller dataset i.e. the one obtained after elimination, across industry sectors and countries. We see that majority (over 80%)

[15]The earliest recorded rating transition is in 1921 although there are very few transitions up until 1970. The last recorded rating transition is in April 2005.

[16]The original data (all countries and sectors) had 27231 rating transitions. After selecting countries and industries, there were 22983 rating transitions left.

of the coverage is for US issuers.[17] Similarly majority (over 75%) of the data relates to the Industrial sector, and of the remainder, substantial parts relate to Banking and Utility sectors. For the rest of the study we focus on these three sectors.

As is customary we grouped the original ratings into eight states: Aaa, Aa, A, Baa, Ba, B, C, D and WR. In our case this grouping is also necessary on practical grounds because the probability of **each** from-to rating transition is modelled as a function of several covariates. Estimating/interpreting all these coefficients is more meaningful if there are only a handful of transitions possible. This makes it necessary to have a parsimonious state space. The ratings are ordered from the highest to the lowest with Aaa being the top ranking, D being the default state and WR denoting the state of rating withdrawal.

In general, there are very few rating category transitions per issuer and it is rare for an individual issuer to make more than three transitions in its life.[18] Table 2 shows the cross tabulation of rating category transitions. The diagonal entries in this table are zero because observations are made in continuous time.[19] The state End signifies that the observation period ended prior to making any transition so the destination state is unknown. From this table it follows that majority of the transitions are to neighboring states, and that there are substantially more downgrades than upgrades.

[17]Moody's coverage used to be largely focused on US issuers but in the recent times has become more and more international.

[18]See footnote 23.

[19]If observations were made at discrete time points, then possibly the source and destination states could have been identical, say for instance when no transition was made. In Table 2 we record a transition only when an actual transition is made. Conversely, every actual transition made does definitely get recorded in Table 2.

A large proportion (over 20%) of these transitions were to the Withdrawn state. An even larger proportion (over 30%) were to End state. These censored observations were not incorporated in the estimates of transition probabilities.[20]

Apart from the transitions themselves, a key quantity of interest is the duration of time spent in each state. The censored observations were indeed useful in enriching the estimates of duration times. Extracting this additional information from censored observations is facilitated by the fact that we employ a continuous time framework. Table 3 lists mean durations in days for each of the rating categories. It seems to indicate that higher rated issuers spend more time in their current rating category before making a transition. Figure 1 presents box plots for the duration times in each rating category. They show that not only the median duration time but also the variability in duration times is more for higher rated issuers.[21]

4 Estimation methodology

4.1 Introduction to Bayesian inference

We use Bayesian inference to estimate the proposed model for ratings migration.[22] Bayesian inference is particularly well suited in capturing random effects and parameter heterogeneity in repeated observation studies, such as ours, where there are a large

[20] As is customary in literature, a transition probability matrix estimate does not report the transition probability to End state.

[21] While commenting on the summary statistics of duration times, it is important to note that the number of transitions made from all initial rating categories is not the same. See Table 2 right hand margin column.

[22] For a more detailed introduction to Bayesian inference, see for instance Congdon (2001).

number of issuers and relative few rating transitions for each issuer.[23] In this situation, traditional estimates at the issuer level either do not exist or have large sampling variability. Bayesian inference automatically shrinks the maximum likelihood estimate (MLE), if it exists, to an aggregate or pooled estimate based on all of the data.

The amount of shrinkage depends on a variety of factors, such as the sampling variation of the issuer-specific MLE and the heterogeneity among the issuers. When the issuer-specific MLE does not exist, the Bayes estimate does by incorporating information from all of the issuers. In sparse-data situations, the issuer-specific estimates reflect the aggregate behavior of the data. As more observations are obtained for a particular issuer, the Bayes estimate reflects less on the aggregate behavior and more on the data for the specific issuer. In addition to issuer-specific estimates, Bayesian inference also estimates the heterogeneity among the issuers.[24]

An important practical advantage of using Bayesian inference is also that it becomes straightforward to obtain interval estimates such as confidence intervals.[25] Interval estimates for default probabilities are becoming increasingly popular.[26] Details of the estimation methodology are given in the Appendix.

[23] The percentage of issuers making exactly 1, 2 and 3 transitions in their entire life is roughly 30%, 20% and 10%. Furthermore, the median of the number of transitions made by issuers during their entire lifetimes is 2. The sparsity is likely to be even more pronounced when narrowing the sample to some specific cross section of issuers such as those in a particular industrial sector or country of domicile.

[24] See Allenby and Lenk (1994) and Allenby and Lenk (1995) for more detailed discussions on Bayesian inference for panel data.

[25] We do not provide such estimates here so as not to distract from our primary focus viz. heterogeneity which can be demonstrated with point estimates.

[26] See for instance Christensen, Hansen, and Lando (2004).

5 Empirical Results

5.1 Estimates for the standard profile

Of primary interest to us is the generator for the continuous time Markov chain, and a one year transition probability matrix. In Table 4 we present these estimates for US issuers in the Industrial sector. These issuers make up more than half of our data, and we treat this profile as the standard profile. It is important to note that we estimate the generator using a day as the unit of time, so the diagonal entries of the generator are to be interpreted as exit rates per day (and not per year).

Estimates for issuers from other countries (we focus on UK, Japan, Canada and EU) or other industry sectors (we focus on Industrial, Banking and Public Utility) will differ from the above standard profile estimates due to inherent heterogeneity in the rating migration behavior. The purpose of this study is to quantify and analyze that difference.

5.2 Estimates for other profiles

In the interest of brevity we do not tabulate generators and transition probability matrices for each possible profile. To illustrate the heterogeneity we choose a few prominent country-sector combinations and compare their estimates with those for the US-Industrial issuers. For the sake of illustration, one year transition probability matrices for US issuers in the industrial, banking and utility sectors (these comprise over three fourths of our data) are given in Table 5. They show strong sector effects.

For instance, Banking and Utility sector issuers have about $7 - 8\%$ lower chance of default. Diagonal entries can vary drastically (e.g. see B or C) and as do upgrade probabilities (e.g. see BAA, BA). Similarly, comparing one year transition probability matrices (not presented here in the interest of brevity) we find prominent country effects within sectors. Both country and sector effects suggest that ignoring the heterogeneity can lead to large errors in default risk computations.

To quantify the heterogeneity we compare different issuer profiles on the basis of following quantities of interest.

1. Jafry-Schuermann mobility metric proposed in Jafry and Schuermann (2004).
2. Probability that a C rated issuer will have defaulted in one year.
3. Probability that a BAA rated issuer will have been upgraded one year later.
4. Probability that a AAA rated issuer will be AAA one year later.

Figures 2 through 5 display the variation in above quantities across different sector-country profiles. It is worth noting that our approach can give estimates for any country-sector combination though such an issuer may not even exist in the dataset we use (or there may be very few issuers with that profile). This is done by aggregating the separately obtained marginal information on issuer characteristics.[27]

[27] See footnote 6.

5.2.1 The Jafry-Schuermann mobility metric

The Jafry-Schuermann mobility metric proposed in Jafry and Schuermann (2004) is the average of the singular values of the mobility matrix for that issuer profile. Here, the mobility matrix is obtained by subtracting an identity matrix from the one year transition probability matrix for that issuer profile. For Industrial issuers in the US, our standard profile, this metric is 0.2278178. Figure 2 shows deviations from this metric for different issuer profiles. A deviation to the left indicates less mobility than that for the standard profile, and vice versa. Figure 2 illustrates that compared to the standard profile, utility sector issuers are generally less mobile and banking sector issuers are generally more mobile. This may have to do with the fact that there is much less uncertainty about the revenue streams of Utility sector issuers (as they are regulated). In contrast, Banking sector issuers are generally highly leveraged and their future revenue streams have higher variance.

5.2.2 The C→D default probability

Ideally we would like to compare unconditional one year default probabilities across issuer profiles. However, the proportions of issuers across rating categories vary across profiles, and the overall default probability becomes a difficult object of comparison. In general the largest default probability is from the C rating category. Hence we compare and contrast default behavior using C→D default probability. Figure 3 shows the variation in this default probability across issuer profiles. One can easily see that compared to other issuer profiles the standard profile of US Industrial issuers shows a generally higher default probability and may lead to an overestimation of default prob-

abilities if issuer heterogeneity is ignored. Within each sector, the ordering observed for the C→D default probability is UK > Canada > US > EU > Japan. This may be a reflection of the differences in corporate bankruptcy laws across different countries. Japanese banks enjoyed a lenient accounting for bad loans in the 1990s. UK corporate bankruptcy laws are relatively highly creditor-friendly. The stance a rating agency may take in granting a C rating under such circumstances is to be overly conservative in Japan, and much less so in the UK. The true proportion of C rated defaulters may therefore be much higher in the UK than in Japan.

5.2.3 The BAA upgrade probability

Apart from default, an interesting outcome of interest is an upgrade, especially if it is from a medium to high grade. Here we illustrate the heterogeneity in the sample by examining the probability that a BAA rated issuer is upgraded to either AAA, AA or A rating category within the next year. Figure 4 shows the variation in this total upgrade probability for BAA rated issuers across different issuer profiles. It shows that the banking sector issuers are 10-15% more likely to be upgraded than issuers from other sectors. UK issuers systematically have a higher upgrade probability than US issuers. The two findings are probably related since the sample of UK issuers is dominated more by Banking sector, whereas the sample of US issuers is dominated by the Industrial sector.

5.2.4 The AAA stay probability

Finally we look at the chance that AAA issuers with negligible credit risk will remain AAA issuers after a year. Figure 5 shows the variation in this stay probability for AAA rated issuers across different issuer profiles. In general these stay probabilities are smallest for Banking issuers and largest for Utility issuers, with those for Industrial issuers lying somewhere in between. This rating stability result is consistent with the rating mobility related result obtained by using the Jafry-Schuermann metric. Again the ordering is possibly a reflection of the uncertainty in revenue streams faced by issuers.

5.3 Robustness checks

Having obtained these systematic patterns in issuer-specific estimates it is natural to wonder if they are indeed rooted in the data or is it the (continuous time) model peculiarities or the (Bayesian) estimation methods that is driving this heterogeneity. As a cross check we set forth to estimate the one year transition probability matrices in three other ways and examine them for evidence of heterogeneity. In each case the aim was to illustrate sector heterogeneity focusing on the differences between Industrial and Utility sector issuers only.

5.3.1 Continuous time logistic model, ML estimation

Firstly, to remove the effect of Bayesian estimation on the results we attempted to compute Maximum Likelihood estimates for a simpler version of our model in con-

tinuous time i.e. an equivalent model without random effects, but with exponential duration and logistic transition probabilities. There are several coefficients to estimate for issuer profile characteristics for the duration model and also for each <from,to> rating category transition. It turned out that in a large majority of cases the data was so sparse that either estimation algorithm did not converge or the converged coefficients were not statistically significant. [28] This is not the case with the main results presented in Section 5 using Bayesian inference. While we could not offer a robustness check for heterogeneity, this exercise highlighted the benefits of Bayesian estimation that is able to tackle the data sparsity.

5.3.2 Continuous time Markov chain, ML estimation

Our second attempt to retain the continuous time domain (in addition to moving away from Bayesian estimation) was simply to remove the effect of our model-specific assumptions (such as logistic functional form for transition probabilities). We do so by performing Maximum Likelihood estimation of an ordinary continuous time Markov chain. Exponentiating the generator so obtained gives the transition probability matrix. To illustrate heterogeneity we compare the transition probability matrices from two subsamples of the original data. One subsample corresponds to Industrial sector issuers, and the second subsample corresponds to Utility sector issuers. These two 1 year transition probability matrices are given in Table 6 and show significant differences across sector subsamples. This can be quantified by the difference of 1.02 between their

[28] We used the ready-made glm routine in R environment for statistical computing. Please see footnote 23 for details the sparsity of this dataset.

Jafry-Schuermann metrics. [29] Furthermore, the default rates in the last column clearly illustrate the sample heterogeneity.

5.3.3 Discrete time Markov chain, yearly observations

Lastly, to remove the influence of continuous time modelling, as well as that of our model's functional form and our Bayesian estimation approach, we computed MLEs for a discrete time Markov chain. In this context it was necessary to make discrete time observations, so we chose to do it at the end points of the ten 1 year intervals. Again, we used the two subsamples of the original data mentioned above. One subsample corresponds to Industrial sector issuers, and the second subsample corresponds to Utility sector issuers. We estimated for each sector's subsample a yearly transition probability matrix for a ten year period ending year 2000. We averaged these ten yearly matrices to obtain an average one year transition probability matrix. As shown in Table 7 this average transition probability matrix showed significant differences across sector subsamples. This can be quantified by the difference of 3.85 between their Jafry-Schuermann metrics. [30] Furthermore, the default rates in the last column clearly illustrate the sample heterogeneity.

6 Implications for Risk Capital

Risk capital is the amount of capital kept aside to cover unexpected economic losses during extreme events. We offer a small illustration of how issuer heterogeneity affects

[29] Figure 2 may help put this number in perspective.
[30] Figure 2 may help put this number in perspective.

risk capital. We construct a hypothetical "typical" credit portfolio, then compute the loss distribution on this portfolio with and without incorporating issuer heterogeneity. The two loss distributions give rise to two different estimates for risk capital, which we choose to quantify by the difference between Value at Risk (VAR) and Expected Loss (EL) for the portfolio at hand. It turns out that for this particular portfolio the risk capital is lower if heterogeneity is taken into account.

The "typical" hypothetical portfolio construction was guided by the following considerations. First, the number of obligors should be approximately 100. Second, the industry sector concentration of exposure amounts should roughly mirror the sector-wise distribution of loan amounts tabulated in Heitfield, Burton, and Chomsisengphet (2006). Third, the distribution of credit quality should roughly be 15% good, 60% medium and 25% bad. The actual portfolio constructed deviated from these considerations slightly but more or less respected all the preset criteria (e.g. it had 105 obligors instead of 100). We assumed the recovery rate to be constant at 40%. The total nominal amount of exposure does not matter as we are interested in risk capital as a percentage of that amount.

The model and method we proposed so far was to estimate the default risk is applicable at obligor level. In a portfolio setting, the dependence structure of defaults becomes crucial in determining the loss distribution of the overall portfolio. We used CreditRisk+ to model this dependence structure. We considered two scenarios. First the default rate inputs were chosen to differ across obligors depending on which industrial sector they lie in, thus explicitly incorporating heterogeneity. Second the default

rate inputs were input as if all obligors belonged to the standard profile of Industrial issuers, thus assuming homogeneity.

We found that ignoring sector heterogeneity in default rates increased the risk capital from 5.4% to 6.3% which is an increase of about 15% in proportional terms. Choosing median loss instead of expected loss to define risk capital, or choosing sector-specific recovery rates instead of a universal constant 0.4, did not change this result in any significant way. Furthermore, using a benchmark portfolio with equal weights across sectors also gave results consistent with those mentioned above.

7 Conclusion

Using a continuous time model, Bayesian estimation techniques and a sample of roughly 23000 rating transitions from the Moodys corporate bond default database we identified significant differences in rating migration behavior between issuers of different industry sectors and countries. Quantifying these differences in terms of a mobility metric and also in terms of default, upgrade and stay probabilities yielded several systematic patterns in the deviation from the standard issuer profile viz. US Industrial issuer. To summarize, we provided strong support and a tool to condition generator estimates on issuer profiles. When working in a portfolio context, such a conditioning gives a clearer picture of Value at Risk and risk capital, as we illustrated using a hypothetical credit portfolio.

References

Allenby, Greg, and Peter Lenk, 1994, Modeling Household Purchase Behavior with Logistic Normal Regression, *Journal of the American Statistical Association* 83, 1218–1231.

Allenby, Greg, and Peter Lenk, 1995, Reassessing Brand Loyalty, Price Sensitivity, and Merchandising Effects on Consumer Brand Choice, *Journal of Business and Economic Statistics* 13, 281–290.

Altman, E. I., 1998, The Importance and Subtlety of Credit Rating Migration, *Journal of Banking and Finance* 22, 1231–1247.

Asquith, P., D. W. Mullins, and E. D. Wolff, 1989, Original Issue High Yield Bonds: Aging Analyses of Defaults, Exchanges, and Calls, *Journal of Finance* 44, 923–952.

Bangia, A., F. X. Diebold, A. Kronimus, C. Schagen, and T. Schuermann, 2002, Ratings migration and the business cycle, with application to credit portfolio stress testing, *Journal of Banking and Finance* 26, 445–474.

Christensen, J. H. E., E. Hansen, and D. Lando, 2004, Confidence sets for continuous-time rating transition probabilities, *Journal of Banking and Finance* 28, 2575–2602.

Congdon, Peter, 2001, *Bayesian Statistical Modelling.* (John Wiley & Sons, New York).

Damien, P., J. C. Wakefield, and S. Walker, 1999, Gibbs sampling for Bayesian non-conjugate and hierarchical models using auxiliary variables, *Journal of the Royal Statistical Society, Series B* 61, 331–344.

Frydman, Halina, and Ashay Kadam, 2004, Estimation in the continuous time mover-stayer model with an application to bond ratings migration, *Applied Stochastic Models in Business and Industry* 20, 155–170.

Frydman, Halina, and Til Schuermann, June 2005, Credit Rating Dynamics and Markov Mixture Models, Working Paper, Wharton Financial Institutions Center.

Heitfield, Erik, Steve Burton, and Souphala Chomsisengphet, 2006, Systematic and idiosyncratic risk in syndicated loan portfolios, *Journal of Credit Risk* 2.

Jafry, Yusuf, and Til Schuermann, 2004, Measurement, Estimation and Comparison of Credit Migration Matrices, *Journal of Banking and Finance.*

Jarrow, R. A., D. Lando, and S. M. Turnbull, 1997, A Markov model for the term structure of credit risk spreads, *The Review of Financial Studies* 10, 481–523.

Lando, D., and T. M. Skodeberg, 2002, Analyzing Rating Transitions and Rating Drift with Continuous Observations, *Journal of Banking and Finance* 26, 423–444.

Lucas, D. J., and J. G. Lonski, 1992, Changes in corporate credit quality 1970-1990, *Journal of Fixed Income* pp. 7–14.

McFadden, D., 1974, Conditional Logit Analysis of Qualitative Choice Behavior, *Frontiers in Econometrics, (New York: Academic Press) Editor P Zarembka* pp. 105–142.

Nickell, Pamela, William Perraudin, and Simone Varotto, 2000, Stability of rating transitions, *Journal of Banking and Finance* 24, 203–227.

Norris, J. R., 1997, *Markov Chains.* (Cambridge University Press).

Ripley, D., 1987, *Stochastic Simulation.* (John Wiley and Sons, New York).

Appendix - Bayesian Estimation

Prior Distributions used

Bayesian analysis of the model requires prior distributions for the unknown parameters. We make common choices:

$$\beta_s \sim N_p(\mu_{0,\beta,s}, \Sigma_{0,\beta,s}) \tag{13}$$

$$\alpha_{r,s} \sim N_q(\mu_{0,\alpha,r,s}, \Sigma_{0,\alpha,r,s}) \tag{14}$$

$$\sigma_s^2 \sim IG\left(\frac{\gamma_{0,s}}{2}, \frac{\delta_{0,s}}{2}\right) \tag{15}$$

$$\Lambda \sim IW_2(\eta_0, \Omega_0) \tag{16}$$

where $N_p(\mu, \Sigma)$ is the p-variate normal distribution with mean μ and covariance matrix Σ; $IG\left(\frac{\gamma}{2}, \frac{\delta}{2}\right)$ is the inverse Gamma distribution with shape $\frac{\gamma}{2}$ and scale $\frac{\delta}{2}$; and $IW_p(\eta, \Omega)$ is the p dimensional inverted Wishart distribution with η degrees of freedom and scale matrix Ω.

In the empirical study, we used highly noninformative priors. We assumed that the prior means $\mu_{0,\beta,s}$ and $\mu_{0,\alpha,r,s}$ are zero, and the prior variances $\Sigma_{0,\beta,s}$ and $\Sigma_{0,\alpha,r,s}$ were 100 times an identity matrix. The parameters for the Inverse Gamma distribution were set so that prior mean for σ_s^2 was one, and the prior variance was 10. The prior degrees of freedom for the Inverse Wishart distribution was six, and the scale matrix was the identity.

MCMC Algorithm

We used Markov chain Monte Carlo (MCMC) (c.f. Congdon 2001) to analyze the model. MCMC sequentially generates the subsets of the parameters from the "full conditional" distribution given the data and the other sets of parameters. Except for the generation of the ln-rate parameters $y_{i,j}$, the MCMC uses standard algorithm. Each duration time $D_{i,j}$ has a ln-rate parameter $y_{i,j}$. The ln-rate parameters have a normal distribution with mean $\mu_{i,s} = x_i'\beta_s + \phi_{i,D}$ and standard deviation σ_s where the state for $D_{i,j}$ is $s = s_{i,j}$. The full conditional distribution for $y_{i,j}$ can be written as the product of exponential and normal densities:

$$f(y_{i,j}) \propto \exp(-y_{i,j}c_{i,j}) \exp[-\exp(-y_{i,j})d_{i,j}] \exp\left[-\frac{(y_{i,j}-\mu_{i,j})^2}{2\sigma_s^2}\right]$$

where $c_{i,j} = 1$ if the duration time $D_{i,j}$ is observed, and $c_{i,j} = 0$ if the duration time is right truncated, which occurs if the bond does not default before the end of the observation interval. We generate $y_{i,j}$ by using the "slice sampling" method of Damien, Wakefield, and Walker (1999). This method introduces an auxiliary random variable V and defines the joint distribution of V and $y_{i,j}$ as:

$$f(y_{i,j}, v) \propto \chi\left\{v \leq \exp[-\exp(-y_{i,j})d_{i,j}]\right\} \exp(-y_{i,j}c_{i,j}) \exp\left[-\frac{(y_{i,j}-\mu_{i,j})^2}{2\sigma_s^2}\right]$$

where χ is the indicator function. One can verify that integrating over V in the joint distribution gives the full conditional distribution of $y_{i,j}$. Given $y_{i,j}$, the conditional distribution of V is uniform on zero to $\exp[-\exp(-y_{i,j})d_{i,j}]$. Given V, the conditional

distribution of $y_{i,j}$ has a truncated normal distribution:

$$f(y_{i,j}|V) \propto \exp\left[-\frac{(y_{i,j} - [\mu_{i,j} - c_{i,j}\sigma_s^2])^2}{2\sigma_s^2}\right] \chi\left\{y_{i,j} > -\ln[-\ln(v)]\right\}.$$

These facts are used in the MCMC to generate $y_{i,j}$. Given the current value of $y_{i,j}^o$, generate $y_{i,j}$ from a truncated normal distribution with mean $\mu_{i,j} - c_{i,j}\sigma_s^2$; variance σ_s^2, and lower truncation:

$$y_{i,j} > -\ln[-\ln(v)] \tag{17}$$

$$y_{i,j} > y_{i,j}^o - \ln\left[d_{i,j} - \exp(y_{i,j}^o)\ln(u)\right] \tag{18}$$

where u is a uniform $[0,1]$ random deviate. We used the inverse cumulative distribution function for the normal distribution to generate the truncated normal (c.f. Ripley (1987)).

Given the ln-rate parameters $\{y_{i,j}\}$, the full conditional distributions for β_s, $\phi_{i,D}$ and σ_s^2 are standard, closed-form distributions. The full conditional density for β_s is:

$$\beta_s \sim N_p(\mu_{n,\beta,s}, \Sigma_{n,\beta,s}) \tag{19}$$

$$\Sigma_{n,\beta,s} = \left(\sum_{i,j:s_{i,j}=s} \frac{1}{\sigma_s^2} x_i x_i' + \Sigma_{0,\beta,s}^{-1}\right)^{-1} \tag{20}$$

$$\mu_{n,\beta,s} = \Sigma_{n,\beta,s}\left(\sum_{i,j:s_{i,j}=s} \frac{1}{\sigma_s^2}(y_{i,j} - \phi_{i,D})x_i + \Sigma_{0,\beta,s}^{-1}\mu_{0,\beta,s}\right). \tag{21}$$

The random effects ϕ_i have a bivariate normal distribution. The full conditional distribution of $\phi_{i,D}$ involves the conditional distribution of $\phi_{i,D}$ given $\phi_{i,A}$:

$$\phi_{i,D}|\phi_{i,A} \sim N\left(\frac{\lambda_{AD}}{\lambda_A^2}\phi_{i,A}, \lambda_D^2 - \frac{\lambda_{AD}^2}{\lambda_A^2}\right). \tag{22}$$

The full conditional distribution of $\phi_{i,D}$ is:

$$\phi_{i,D} \sim N_p(\mu_{i,D}, \Sigma_{i,D}) \tag{23}$$

$$\Sigma_{i,D} = \left(\sum_{s=1}^{K}\sum_{j:s_{i,j}=s}\frac{1}{\sigma_s^2} + \frac{\lambda_A^2}{\lambda_D^2\lambda_A^2 - \lambda_{AD}^2}\right)^{-1} \tag{24}$$

$$\mu_{i,D} = \Sigma_{i,D}\left(\sum_{s=1}^{K}\sum_{j:s_{i,j}=s}\frac{1}{\sigma_s^2}(y_{i,j} - x_i'\beta_s) + \frac{\lambda_{AD}\phi_{i,A}}{\lambda_D^2\lambda_A^2 - \lambda_{AD}^2}\right). \tag{25}$$

The full conditional distribution of σ_s^2 is:

$$\sigma_s^2 \sim IG\left(\frac{\gamma_{n,s}}{2}, \frac{\delta_{n,s}}{2}\right) \tag{26}$$

$$\gamma_{n,s} = \gamma_{0,s} + \sum_{i,j:s_{i,j}=s} 1 \tag{27}$$

$$\delta_{n,s} = \delta_{0,s} + \sum_{i,j:s_{i,j}=s}\left(y_{i,j} - x_i'\beta_s - \phi_{i,D}\right)^2 \tag{28}$$

We use random walk, Metropolis-Hastings to generate the parameters $\alpha_{r,s}$ and $\phi_{i,A}$ for the jump distributions or transition probabilities. We generate a candidate value $\alpha_{r,s}^c$ from a random walk:

$$\alpha_{r,s}^c \sim N_q(\alpha_{r,s}, \tau_1^2 I)$$

where τ_1 is a tuning parameter for the algorithm. This candidate is accepted with probability:

$$\rho(\alpha_{r,s}, \alpha^c_{r,s}) = \min\left\{\frac{\pi(\alpha^c_{r,s})}{\pi(\alpha_{r,s})}, 1\right\}$$

where π is proportional to the posterior distribution of $\alpha_{r,s}$:

$$\pi(\alpha_{r,s}) = \prod_{i,j:s_{i,j-1}=r} P(s_{i,j}|r,i) \exp\left[-\frac{1}{2}(\alpha_{r,s} - \mu_{0,\alpha,r,s})'\Sigma^{-1}_{0,\alpha,r,s}(\alpha_{r,s} - \mu_{0,\alpha,r,s})\right],$$

and $s_{i,j}$ are the observed ratings for all issuers and transitions. The current value $\alpha_{r,s}$ is retained with probability $1-\rho(\alpha_{r,s}, \alpha^c_{r,s})$. Similarly, random walk, Metropolis-Hastings is used to generate $\phi_{i,A}$. Generate a candidate from:

$$\phi^c_{i,A} \sim N(\phi_{i,A}, \tau_2^2),$$

and τ_2 is a tuning parameter for the algorithm. The candidate is accepted with probability

$$\rho(\phi_{i,A}, \phi^c_{i,A}) = \min\left\{\frac{\pi(\phi^c_{i,A})}{\pi(\phi_{i,A})}, 1\right\}$$

where

$$\pi(\phi_{i,A}) = \prod_{j:s_{i,j-1}=r} P(s_{i,j}|r,i) \exp\left[-\frac{\lambda_D^2\left(\phi_{i,A} - \frac{\lambda_{AD}}{\lambda_D^2}\phi_{i,D}\right)^2}{2\left(\lambda_A^2\lambda_D^2 - \lambda_{AD}^2\right)}\right],$$

and the current $\phi_{i,A}$ is retained with probability $1 - \rho(\phi_{i,A}, \phi^c_{i,A})$.

The initial "burn-in" period of our MCMC chains consisted of 100,000 iterations. We then generated another 100,000 iterations for estimation. To conserve memory, we

thinned the chained by only using every tenth iterations, for a total of 10,000 iterations to compute posterior means and posterior standard deviations of the parameters. We also computed the generator for various values of the covariates on each of the 10,000 iterations that we used for estimation. On each of these iterations, we generated 100 random effects ϕ_i, and 100 error terms $\epsilon_{i,s}$ for each state, and computed the generator for each draw of the random effects and error terms. In totaly, we computed the generator 1,000,000 times for each setting of covariates.

Table 1: Cross Tabulation : Countries vs. Sectors

	Banking	Industrial	Utility	Other	Total
Canada	47	546	46	224	863
EU	569	685	71	312	1637
Japan	168	469	23	89	749
UK	232	537	113	131	1013
US	1200	12848	2115	2558	18721
Total	2216	15085	2368	3314	22983

Table 2: Cross tabulation of rating category transitions

	End	AAA	AA	A	Baa	Ba	B	C	WR	D	Total
End	0	0	0	0	0	0	0	0	0	0	0
AAA	130	0	256	20	0	0	0	0	98	0	504
AA	667	102	0	828	15	3	4	0	375	0	1994
A	1067	10	537	0	1199	49	12	1	742	0	3617
Baa	965	7	40	804	0	976	75	12	784	6	3669
Ba	440	1	9	49	715	0	1250	68	965	16	3513
B	696	1	9	25	54	611	0	1234	897	112	3639
C	318	0	1	1	12	24	228	0	344	848	1776
WR	3294	32	59	171	171	232	238	74	0	0	4271
D	0	0	0	0	0	0	0	0	0	0	0
Total	7577	153	911	1898	2166	1895	1807	1389	4205	982	22983

Table 3: Mean Duration Times

Rating Category	AAA	AA	A	Baa	Ba	B	C	WR
Mean Duration in Days	2807	2192	2463	2112	1449	1143	1002	1140

Table 4: Daily Generator (Q) & Yearly Transition Probability Matrix (TPM) Estimates for the Standard Profile i.e. US Industrial Issuers

Q	AAA	AA	A	BAA	BA	B	C	WR	D
AAA	−3.236E−04	2.138E−04	1.240E−05	4.578E−15	1.472E−13	2.957E−14	3.475E−16	9.743E−05	1.051E−09
AA	3.016E−05	−4.020E−04	2.410E−04	3.506E−06	2.491E−06	1.993E−06	4.525E−08	1.227E−04	1.789E−07
A	1.088E−06	3.991E−05	−3.340E−04	1.702E−04	1.116E−05	2.678E−06	1.357E−07	1.088E−04	5.168E−08
BAA	2.947E−07	4.020E−06	8.942E−05	−4.183E−04	1.657E−04	1.498E−05	4.825E−07	1.426E−04	8.381E−07
BA	8.035E−08	2.015E−06	9.691E−06	1.142E−04	−6.990E−04	3.062E−04	1.217E−05	2.505E−04	4.133E−06
B	2.053E−07	2.003E−06	5.151E−06	1.153E−05	1.375E−04	−7.398E−04	3.133E−04	2.397E−04	3.029E−05
C	1.465E−10	1.346E−09	3.879E−07	7.418E−06	1.511E−05	1.499E−04	−1.083E−03	2.715E−04	6.384E−04
WR	2.957E−05	9.249E−05	3.800E−04	4.554E−04	8.916E−04	9.936E−04	2.848E−04	−3.128E−03	7.207E−08
D	0	0	0	0	0	0	0	0	0

TPM	AAA	AA	A	BAA	BA	B	C	WR	D
AAA	0.8891	0.0688	0.0088	0.0023	0.0039	0.0044	0.0013	0.0213	0.0001
AA	0.0098	0.8650	0.0791	0.0061	0.0058	0.0061	0.0017	0.0262	0.0002
A	0.0006	0.0133	0.8886	0.0566	0.0093	0.0060	0.0016	0.0238	0.0002
BAA	0.0003	0.0021	0.0311	0.8634	0.0553	0.0135	0.0025	0.0312	0.0006
BA	0.0004	0.0017	0.0079	0.0395	0.7876	0.0971	0.0112	0.0516	0.0030
B	0.0004	0.0016	0.0058	0.0094	0.0483	0.7782	0.0854	0.0500	0.0208
C	0.0003	0.0011	0.0045	0.0074	0.0147	0.0503	0.6790	0.0494	0.1933
WR	0.0063	0.0199	0.0815	0.0997	0.1790	0.2009	0.0627	0.3398	0.0102
D	0	0	0	0	0	0	0	0	1

Table 5: Estimation Results : Transition Probability Matrices (over one year horizon) for US Issuers in various industry sectors. Table entries are 100 times the probability values.

Banking Sector

	AAA	AA	A	BAA	BA	B	C	WR	D
AAA	85.06	11.44	0.64	0.25	0.07	0.01	0.01	2.53	0.00
AA	0.43	90.07	5.91	0.39	0.09	0.04	0.01	3.04	0.00
A	0.15	5.58	88.32	2.95	0.23	0.04	0.01	2.72	0.00
BAA	0.27	2.03	15.61	72.01	5.54	0.48	0.05	3.97	0.04
BA	0.17	1.10	5.95	14.92	61.71	10.57	1.28	4.24	0.06
B	0.31	0.68	1.61	4.90	13.98	56.97	14.14	6.45	0.95
C	0.07	0.18	0.25	0.41	0.33	0.75	85.00	2.02	11.00
WR	4.54	11.21	14.58	12.38	3.36	0.63	0.23	53.04	0.02
D	0	0	0	0	0	0	0	0	100

Industrial Sector

	AAA	AA	A	BAA	BA	B	C	WR	D
AAA	88.91	6.88	0.88	0.23	0.39	0.44	0.13	2.13	0.01
AA	0.98	86.50	7.91	0.61	0.58	0.61	0.17	2.62	0.02
A	0.06	1.33	88.86	5.66	0.93	0.60	0.16	2.38	0.02
BAA	0.03	0.21	3.11	86.34	5.53	1.35	0.25	3.12	0.06
BA	0.04	0.17	0.79	3.95	78.76	9.71	1.12	5.16	0.30
B	0.04	0.16	0.58	0.94	4.83	77.82	8.54	5.00	2.08
C	0.03	0.11	0.45	0.74	1.47	5.03	67.90	4.94	19.33
WR	0.63	1.99	8.15	9.97	17.90	20.09	6.27	33.98	1.02
D	0	0	0	0	0	0	0	0	100

Utility Sector

	AAA	AA	A	BAA	BA	B	C	WR	D
AAA	94.17	3.78	1.83	0.07	0.01	0.00	0.00	0.15	0.00
AA	0.34	90.63	7.45	0.54	0.03	0.02	0.01	0.98	0.00
A	0.03	1.79	92.22	4.47	0.20	0.02	0.01	1.27	0.00
BAA	0.02	0.10	3.72	90.97	3.16	0.25	0.11	1.65	0.02
BA	0.02	0.10	0.50	9.93	82.87	3.84	0.95	1.66	0.14
B	0.01	0.06	0.33	3.14	15.35	66.63	9.18	3.39	1.90
C	0.04	0.09	0.64	2.97	3.78	11.68	62.27	6.39	12.14
WR	0.01	1.90	11.27	19.20	3.63	1.00	0.57	62.37	0.06
D	0	0	0	0	0	0	0	0	100

Table 6: Robustness Checks : Continuous time Markov Chain

Industrial	Aaa	Aa	A	Baa	Ba	B	C	WR	D
Aaa	89	6	1	0	0	0	0	3	0
Aa	1	86	8	0	0	0	0	4	0
A	0	1	88	6	1	0	0	4	0
Baa	0	0	3	84	6	1	0	6	0
Ba	0	0	0	4	77	9	1	8	0
B	0	0	0	1	4	74	9	9	3
C	0	0	0	0	1	5	63	9	22
WR	0	0	1	2	3	3	1	90	0
D	0	0	0	0	0	0	0	0	100

Utility	Aaa	Aa	A	Baa	Ba	B	C	WR	D
Aaa	95	3	2	0	0	0	0	0	0
Aa	0	91	8	1	0	0	0	1	0
A	0	2	91	5	0	0	0	2	0
Baa	0	0	4	89	4	0	0	2	0
Ba	0	0	0	10	82	4	1	2	0
B	0	0	0	3	18	60	11	5	3
C	0	0	0	2	4	13	57	8	15
WR	0	0	3	4	1	0	0	91	0
D	0	0	0	0	0	0	0	0	100

Transition Probability Matrices (over one year horizon) for Industrial and Utility Issuers. Each table above shows 100 times the probability values. The matrices have been estimated using an ordinary continuous time Markov chain and the entire dataset. They illustrate the heterogeneity between Industrial and Utility sector issuers.

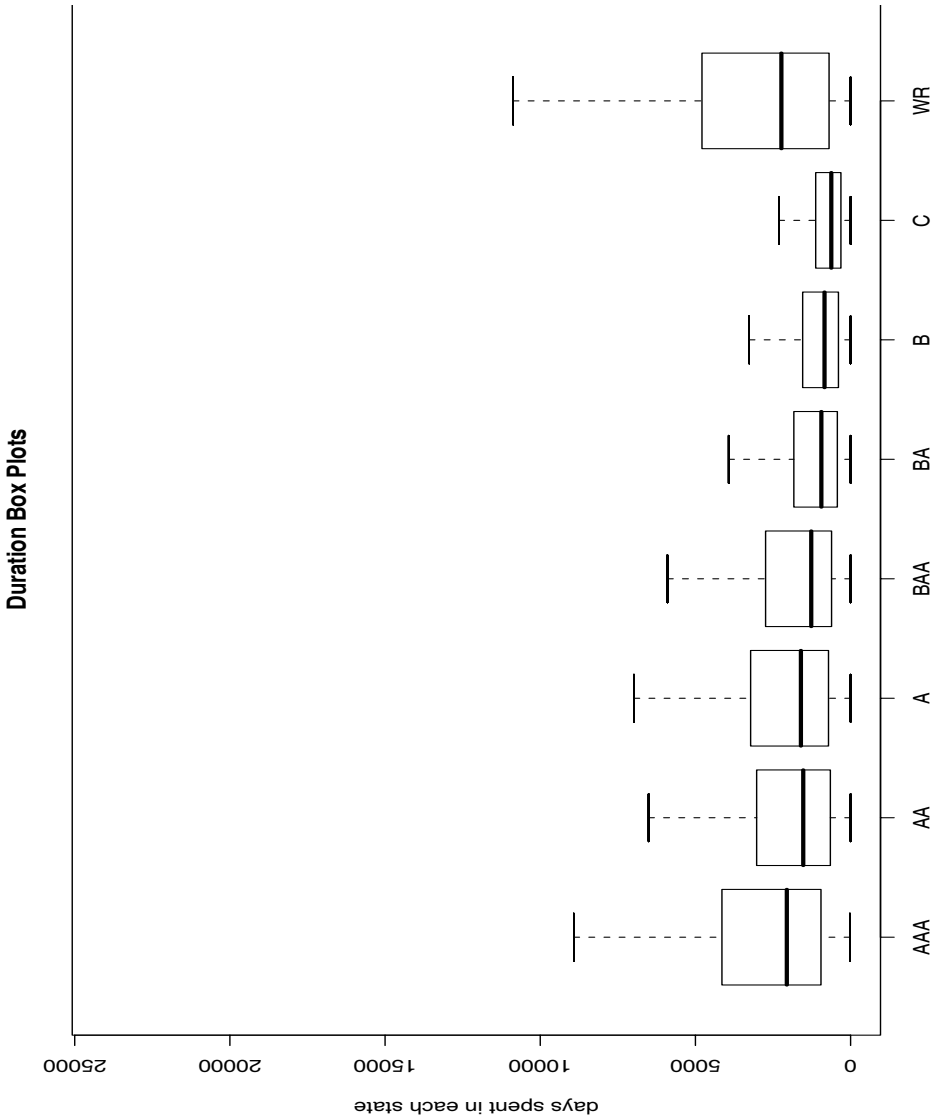
Table 7: Robustness Checks : Discrete time Markov Chain

Industrial	Aaa	Aa	A	Baa	Ba	B	C	WR	D
Aaa	84	8	2	0	0	0	0	5	0
Aa	0	83	9	1	0	0	0	6	0
A	0	1	84	7	1	0	0	6	0
Baa	0	0	4	83	4	1	0	7	0
Ba	0	0	1	5	70	10	1	12	1
B	0	0	0	0	5	71	9	11	4
C	0	0	0	0	1	5	63	10	22
WR	0	0	1	2	3	7	2	85	0
D	0	0	0	0	0	0	0	0	100

Utility	Aaa	Aa	A	Baa	Ba	B	C	WR	D
Aaa	91	5	4	0	0	0	0	0	0
Aa	0	83	12	2	0	0	0	4	0
A	0	2	88	5	1	0	0	3	0
Baa	0	0	4	87	3	1	0	5	0
Ba	0	0	1	5	83	5	1	4	0
B	0	0	1	10	11	70	1	5	3
C	0	0	0	0	0	11	65	10	14
WR	0	1	3	6	2	1	0	86	0
D	0	0	0	0	0	0	0	0	100

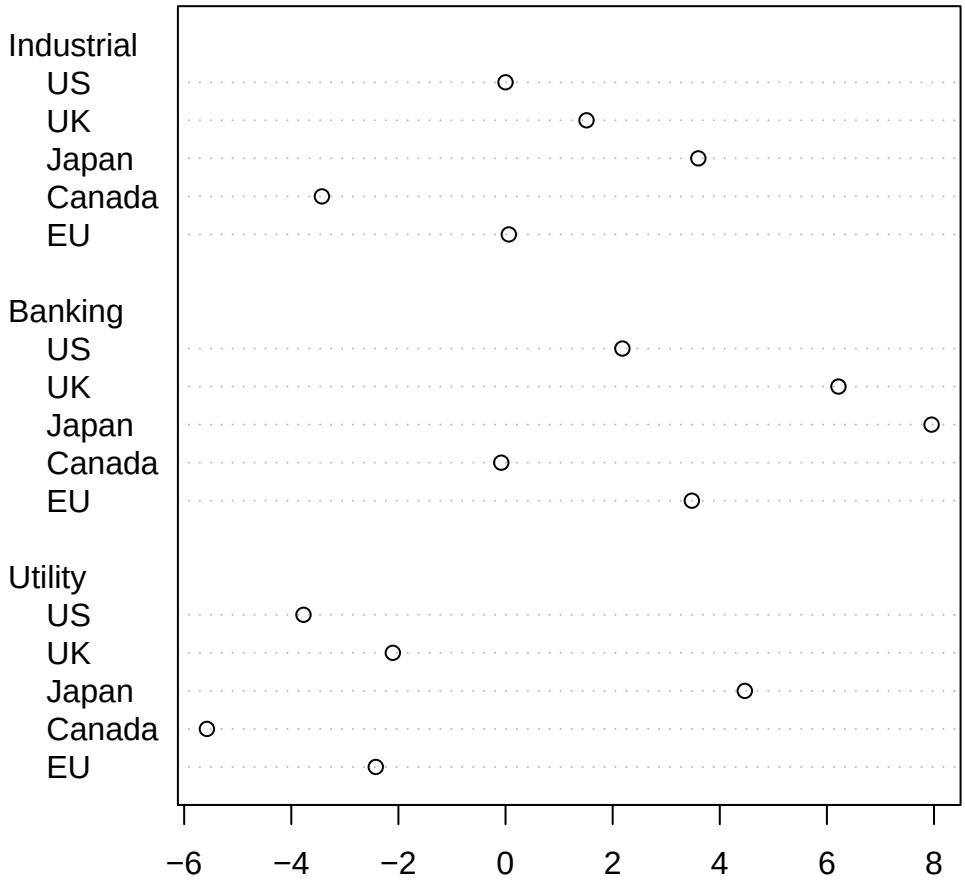
Transition Probability Matrices (over one year horizon) for Industrial and Utility Issuers. Each table above shows 100 times the probability values. The matrices have been estimated using a discrete time Markov chain and discrete (yearly) observation times over the ten year period 1991-2000. They illustrate the heterogeneity between Industrial and Utility sector issuers.

Figure 1: Summary of duration times spent in each rating category



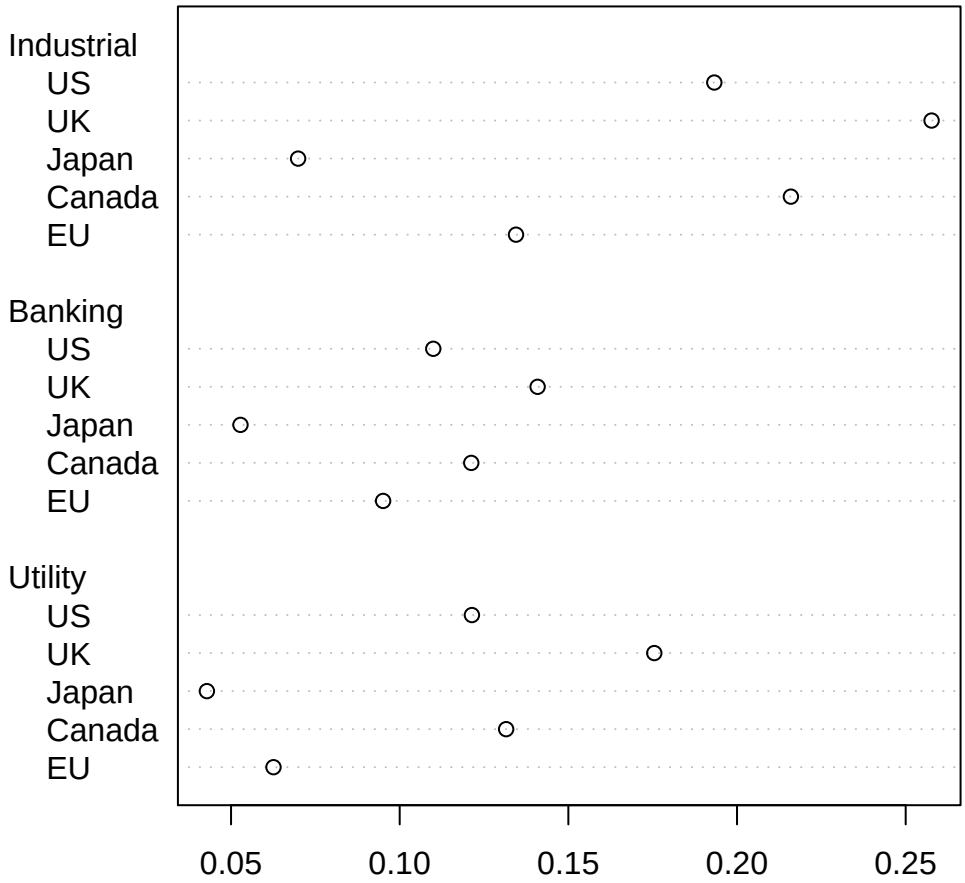
This figure shows the variation of duration times across rating categories. Higher rated issuers have longer and more variable duration times.

Figure 2: Variation of the Jafry Schuermann metric across issuer profiles



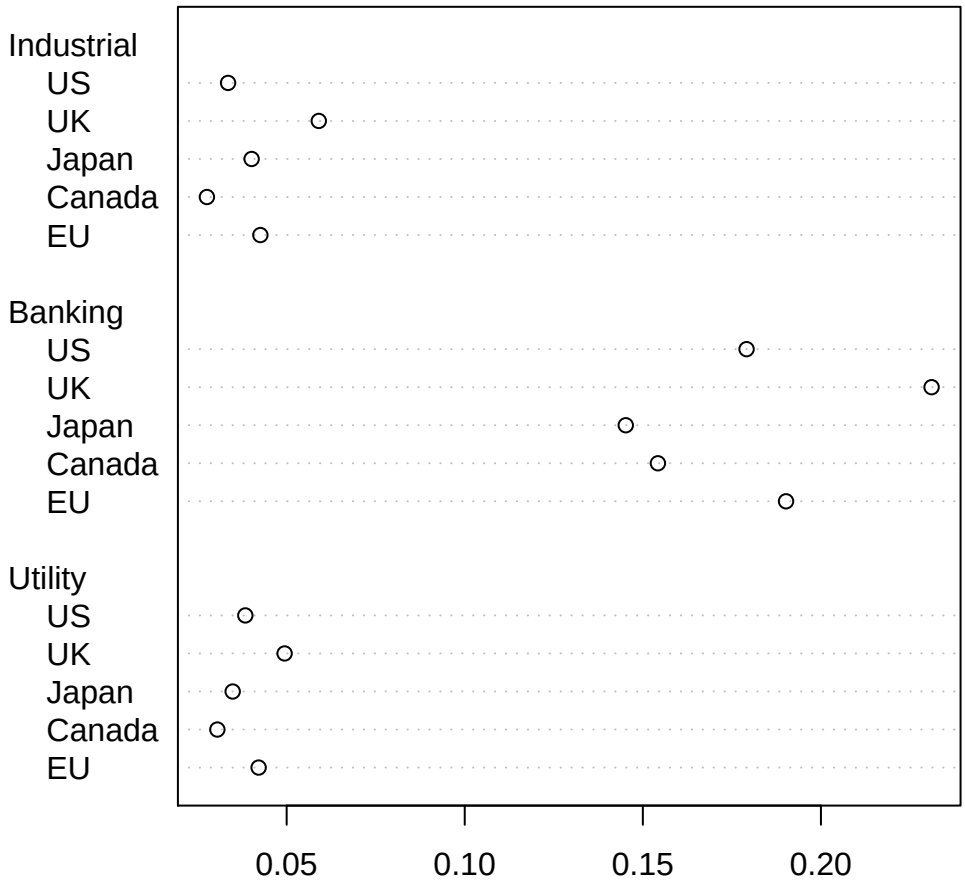
This figure shows the variation of the Jafry-Schuermann metric (proposed in Jafry and Schuermann (2004)) across country-sector profiles on a relative scale. US Industrial issuers make up the standard profile. The figure shows for other important issuer profiles, (100 times) the deviation of the mobility metric from this standard. Compared to US Industrial issuers, US as well as non-US issuers from the utility sector have generally lower mobility and US as well as non-US issuers from the banking sector have generally higher mobility.

Figure 3: Variation of the C→D default probability across issuer profiles



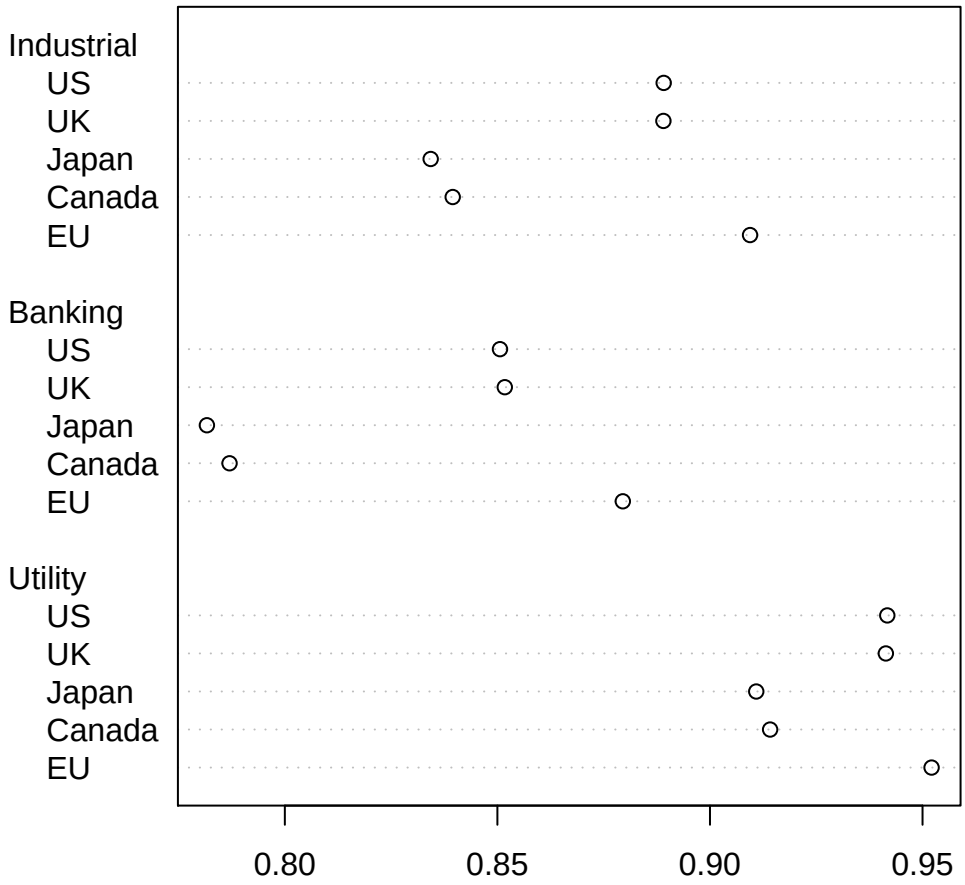
This figure shows the variation in C→D default probability across issuer profiles. The standard profile of US Industrial issuers shows a generally higher default probability. Within each sector, the EU issuers show systematically lower default probabilities than their US and UK counterparts.

Figure 4: Variation of the BAA upgrade probabilities across issuer profiles



This figure graphs the probability that a BAA rated issuer is upgraded to either AAA, AA or A rating category within the next year. It shows the variation in this total upgrade probability for BAA rated issuers across different issuer profiles. It shows that the banking sector issuers are 10-15% more likely to be upgraded than issuers from other sectors. Within each sector depicted, UK issuers systematically have the highest upgrade probability.

Figure 5: Variation of the AAA stay probabilities across issuer profiles



This figure shows the variation in the stay probability for AAA rated issuers across different issuer profiles. This is the chance that AAA issuers will remain in AAA rating category after a year. In general these stay probabilities are smallest for Banking issuers and largest for Utility issuers, with those for Industrial issuers lying somewhere in between. Within each sector, the EU issuers show systematically higher AAA stay probabilities than other issuers in that sector.