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: September 7, 2007

JEL Classification : C11, C13, C41, G12

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

*We appreciate feedback from the audience/discussants at the following conferences : Symposium on Credit Risk (Gutmann Center, Vienna, 2007), Annual Derivatives Securities and Risk Management Conference (FDIC, Washington D.C., 2007), Credit Risk Conference (ISCTE, Lisbon, 2007), Credit Risk Workshop (Cass Business School, London, 2006) and Quantitative Methods in Finance (UT Sydney, 2005). In particular, we would like to thank Ana-Maria Fuertes, M. Nazmul Hasan, Shelagh Heffernan, Wenying Jiangli, Aneel Keswani, Meziane Lasfer, Gulnur Muradoglu, Min Qi, Alexander Reisz, Anthony Saunders, Til Schuermann, Anand Srinivasan and Stuart Turnbull for specific helpful comments. Excellent research assistance was provided by Helder Palaro. We are grateful to two anonymous referees for their valuable suggestions on improving the paper. We accept responsibility for all remaining errors.

†Corresponding author. Address: 106 Bunhill Row, London EC1Y 8TZ, United Kingdom. Email:A.Kadam@city.ac.uk Phone: +44-20-7040-8632

Bayesian Inference for Issuer Heterogeneity in Credit Ratings Migration

Abstract

Rating transition matrices for corporate bond issuers are often based on fitting a discrete time Markov chain model to homogeneous cohorts. Literature has documented that rating migration matrices can differ considerably depending on the characteristics of the issuers in the pool used for estimation. However, it is also well known in literature that a continuous time Markov chain gives statistically superior estimates of the rating migration process. It remains to verify and quantify the issuer heterogeneity in rating migration behavior using a continuous time Markov chain. We fill this gap in literature. We provide Bayesian estimates to mitigate the problem of data sparsity. Default data, especially when narrowing down to issuers with specific characteristics, can be highly sparse. Using classical estimation tools in such a situation can result in large estimation errors. Hence we adopt Bayesian estimation techniques. We apply them to the Moodys corporate bond default database. Our results indicate strong country and industry effects on the determination of rating migration behavior. Using the CreditRisk+ framework, and a sample credit portfolio, we show that ignoring issuer heterogeneity can give erroneous estimates of Value-at-Risk and a misleading picture of the risk capital. This is consistent with some recent findings in literature. Therefore, given the upcoming Basel II implementation, understanding issuer heterogeneity has important policy implications and warrants further research.

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] However, it is also well known in literature that a continuous time Markov chain gives statistically superior

[1]Norris (1997) gives an elaborate treatment of Markov Chains. Having accepted this model, the actual reported transition probability matrices can vary considerably 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 Frydman and Schuermann (2007), Chava, Stefanescu, and Turnbull (2006), Frydman and Kadam (2004), Bangia, Diebold, Kronimus, Schagen, and Schuermann (2002), Nickell, Perraudin, and Varotto (2000), Lucas and Lonski (1992), Asquith, Mullins, and Wolff (1989) and research summary reports published by rating agencies such as Moody's KMV, Standard & Poor's and Fitch on their web-sites.

estimates of the rating migration process.[3] It remains to verify and quantify the issuer heterogeneity in rating migration behavior using a continuous time Markov chain. We fill this gap in literature.[4]

In that sense our modelling framework is similar to Frydman and Kadam (2004) and Frydman and Schuermann (2007). Both of these apply continuous time Markov chain based mixture models to ratings data. The discrete time model of Chava, Stefanescu, and Turnbull (2006), explicitly addressing issuer heterogeneity, has a similar motivation. However, all of these use Maximum Likelihood Estimation for model calibration.

We provide Bayesian estimates to mitigate the problem of data sparsity. Default data, especially when narrowing down to issuers with specific characteristics, can be highly sparse. Using classical estimation tools in such a situation can result in large estimation errors. In contrast, Bayesian methods have at least the following two major advantages.[5]

The first advantage is that of estimation accuracy. Stefanescu, Tunaru, and Turnbull (2007), who also advocate Bayesian methodology for calibrating models for rating

[3]Jarrow, Lando, and Turnbull (1997) were among the first to fit a continuous time Markov chain model to observed bond prices. Lando and Skodeberg (2002) give very clear arguments in favor of using a continuous time Markov chain. Christensen, Hansen, and Lando (2004) re-iterate these advantages and provide further extensions by allowing for latent excited states in the state space, as well as by reporting interval estimates.

[4]Such an exercise would be redundant if every observed rating transition matrix of a discrete time Markov chain could result unambiguously from a continuous time Markov chain rating migration process. However, in general, neither existence nor uniqueness of solution to this embedding problem is assured. A good summary of important results relevant to this problem, in the ratings migration context, can be found in Schonbucher (2003), Bluhm, Overbeck, and Wagner (2002) or Israel, Rosenthal, and Wei (2001).

[5]A third side-benefit of using Bayesian inference is also that it becomes straightforward to compute the transition or default probability interval estimates which are becoming increasingly popular; see for instance Christensen, Hansen, and Lando (2004). 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.

transition probabilities using historical data, assert "Model calibration for this type of application is difficult in a classical frequentist estimation framework, because the sparsity of data often leads to unrealistic transition probabilities".[6] Because of the nature of estimation procedure, we are able to provide estimates for an arbitrary issuer profile even if data on that profile may be a very small part of the sample we use for estimation.[7] The estimation error in doing this using a frequentist approach may be quite large.

The second advantage is the incorporation of expert opinion or subjective beliefs (such as those held by regulators) via prior distributions for rare events (such as rating transitions or defaults). As pointed out by McNeil and Wendin (2007), who also advocate Bayesian estimation for portfolio credit risk applications, this "could, in a sense, allow us to draw stronger conclusions about default risk than is possible from an analysis of empirical defaults alone". In our empirical analysis, we used highly noninformative priors but given default data sparsity, incorporating prior beliefs is a valuable tool that can be potentially prove quite useful.

Our empirical results build upon the work of Nickell, Perraudin, and Varotto (2000), who made a significant contribution to literature by fitting a Probit model to discrete rating data. Their model-based approach allows for each qualifier of interest (e.g. country of domicile), a conditional transition matrix (over a given time period), estimated

[6]We experience this weakness of the frequentist approach when performing robustness checks for our results. See Section 5.3 for details.

[7] 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).

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). Their ordered probit model assumes that rating changes when an unobserved, latent measurement falls into disjoint, adjacent intervals. An advantage of this approach is that a common set of parameters for the latent measure is used for each rating state. In our model, the transition parameters depend on the current state of the process, which provides a more flexible model than the ordered-probit at the cost of a significantly larger parameter space. Fortunately, the issuer rating dataset we use is large, and Bayesian inference enables the estimation of a large number of parameters. Furthermore, we explicitly model duration viz. the time spent by an issuer in the current state before making a transition to the future state. Modelling the duration explicitly allows us to provide a richer understanding of rating stability.[8]

Given the use of Bayesian techniques in ratings migration context, our work shares some similarities with McNeil and Wendin (2006) who allow for serially correlated unobserved risk factors that affect the rating migrations process.[9] An important difference between the two models is that their model is for discrete-time migration counts whereas we work in a continuous-time framework.[10]

[8] 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. other discrete time Markov chain model based papers (such as Nickell, Perraudin, and Varotto (2000)) is that duration times have a model that captures this large variability.

[9] This serial correlation gives joint migration distributions in terms of high dimensional integrals, which are awkward for standard maximum likelihood procedures; Bayesian estimation circumvents this problem.

[10] As mentioned before, this is to benefit from some of the statistical advantages mentioned in Lando and Skodeberg (2002).

Our results indicate strong country and industry effects on the determination of rating migration behavior. For instance, issuer default probability shows a clear ordering across countries: UK > Canada > US > EU > Japan. Utility sector issuer ratings are generally more stable and whereas Banking sector issuer ratings are generally less stable; Industrial sector issuer ratings lie somewhere in between. A possible explanation for country heterogeneity is the cross-country variation in bankruptcy codes, corporate governance and accounting standards. A possible explanation for sector heterogeneity is the cross-sector variation in the uncertainty of future revenue streams.

Using the CreditRisk+ framework, and a sample credit portfolio, we show that ignoring issuer heterogeneity can give erroneous estimates of Value-at-Risk and a misleading picture of the risk capital. This is consistent with some recent findings in literature. Using a different portfolio composition, a different dependency model between obligors, different data and different (classical) estimation methods, Hanson, Pesaran, and Schuermann (2007) show that "heterogeneity in the ... probability of default, measured for instance by a credit rating, is of first order importance in affecting the shape of the loss distribution". Being able to explicitly recognize the heterogeneity in the issuer pool gives us a clearer picture of both Value at Risk and risk capital, both of which depend crucially on the loss distribution. Therefore, given the upcoming Basel II implementation, understanding issuer heterogeneity has important policy implications and warrants further research.

The rest of the paper is organized as follows: Sections 2 and 3 summarize the model and estimation procedure used, relegating the details to the Appendix. Sections 4 and 5 describe the data and empirical results obtained. Section 6 demonstrates some implications for holding risk capital and section 7 presents the conclusion.

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.

We index the states by $k = 1, \ldots, K + 1$. The states $1, \ldots, K - 1$ are ordered such that as the index increases, credit quality deteriorates. 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 j^{th} 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 $s_{i,j+1} \neq s_{i,j}$. Duration times are defined from the transition times. The duration time for the j^{th} transition for issuer

i is: $D_{i,j} = T_{i,j+1} - T_{i,j}$ for $j = 1, \ldots, n_i - 1$. Since the observation period is a finite interval, we need to carefully deal with the left and right truncation of the observed rating process. These truncation technicalities are explained in Appendix I.

In discrete space, continuous time, stationary Markov processes, the duration times are mutually independent and exponentially distributed random variables. Suppose the density for duration $D_{i,j}$ is $f(t|y_{i,j}) = \exp(-y_{i,j}) \exp\left[-\exp(-y_{i,j})t\right]$ for $t > 0$ 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:

$$y_{i,j} = x_i' \beta_s + \phi_{i,D} + \epsilon_{i,j} \tag{1}$$

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.

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. We choose to model

[11] The covariates used for implementation were dummy variables to capture country and industry effects. In a different context, Aretz and Pope (2007) decompose the systematic variation of shocks to firms default probabilities into (a global,) a country and an industry effect and identify strong country effects.

the probabilities of jumping from r to s as logistic functions[12]

$$\begin{aligned} P(r|r,i) &= 0 && (2) \\ P(s|r,i) &\propto \exp(z_i'\alpha_{r,s}) \text{ for } s = 1,\ldots,K \text{ and } s \neq r && (3) \\ P(K+1|r,i) &\propto \exp(\phi_{i,A}) \text{ for the absorbing (default) state } K+1 && (4) \end{aligned}$$

where z_i is a q-vector of covariates[13] 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.

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 $\Lambda = \begin{bmatrix} \lambda_A^2 & \lambda_{AD} \\ \lambda_{AD} & \lambda_D^2 \end{bmatrix}$. Further explanation of the random effects used in both duration and transition models is given in Appendix I.

Stationary Markov processes can also be compactly represented by their generators. The generator for the rating migration process of issuer i in our framework 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 can now be written

[12] Appendix I gives a motivation for the jump probability distribution model.

[13] 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.

as the $K+1$ by $K+1$ matrix

$$
Q_i(\phi_i, \epsilon_i, x_i, z_i) = \begin{cases} -\exp(-y_{i,j}) & \text{for the } (j,j) \text{ element and } j = 1, \ldots, K \\ \exp(-y_{i,j}) P(k|j,i) & \text{for the } (j,k) \text{ element where} \\ & j = 1, \ldots K;\ k = 1, \ldots, K+1;\ j \neq k \\ 0 & \text{for the } (K+1,k) \text{ element; } k = 1, \ldots, K+1. \end{cases} \quad (5)
$$

3 Estimation methodology

We use Bayesian inference to estimate the proposed model for ratings migration.[14] 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 number of issuers and relative few rating transitions for each issuer. For the dataset we used, 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. 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.

[14]For a more detailed introduction to Bayesian inference, see for instance Congdon (2001). For more detailed discussions on Bayesian inference for panel data, see Allenby and Lenk (1994) and Allenby and Lenk (1995).

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.

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.[15] 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. On each of these iterations, we generated 100 random effects ϕ_i, and computed the generator for each draw of the random effects and error terms. In total, we computed the generator 1,000,000 times for each setting of covariates. We used highly noninformative priors. Further details on the prior distributions and the estimation methodology we used are given in Appendix II.

[15]We generate $y_{i,j}$ by using the "slice sampling" method of Damien, Wakefield, and Walker (1999).

4 Data Description

The dataset we use is the entire Moodys corporate bond default database available as of late 2005.[16] 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.[17]

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%) of the coverage is for US issuers.[18] 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.

[16]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.

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

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

As is customary we grouped the original ratings into eight states: Aaa, Aa, A, Baa, Ba, B, C, D and WR. 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. 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. 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

[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.

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

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]

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.

[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.

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.

In order to formalize these systematic differences a first approach could be to consider some distance measure between generators (or transition probability matrices) of two issuer profiles and check if this distance is significantly different from zero. In a Bayesian context, the distance measure will have its own posterior probability distribution. Hence a distance significance check translates to checking if the mean of the distance is significantly larger than the standard deviation of the distance. We found that for any issuer profile (characterized by a country-sector combination) its distance from all other issuer profiles was quite large. Without exception the mean distance was at least thrice the standard deviation of distance, usually much larger. This was true for generators as well as transition probability matrices, and for L1 as well as L2 measures of distance. Thus our first approach indicates that rating migration behavior is statistically significantly different across the issuer profiles we considered.

In order to further quantify the heterogeneity we now compare different issuer profiles on the basis of the following specific 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.

5.2.1 Overall mobility vs. AAA stability

The Jafry-Schuermann mobility metric proposed in Jafry and Schuermann (2004) is a measure of overall mobility for a rated issuer. It 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 0.1552092. 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. We see a pattern consistent with this result when examining the AAA stability i.e. the chance that AAA issuers, which have negligible credit risk, will remain in the AAA rating category after one 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. Both these observations may have to do with the fact that there is usually much less uncertainty about the revenue streams of Utility sector issuers (especially if they are regulated monopolies). Banking sector issuers are generally highly leveraged and their future revenue streams usually have higher variance.

5.2.2 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 probabilities if issuer heterogeneity is ignored. Within each sector, the ordering observed for the C$\rightarrow$D default probability is UK $>$ Canada $>$ US $>$ EU $>$ Japan.

It is possible that this clear ordering is a reflection of the differences in corporate bankruptcy environment across different countries. Davydenko and Franks (2006) claim that “despite significant adjustments in lending practices, bankruptcy codes still sharply affect default outcomes”. Furthermore, Mayer (1998) notes that “there are important differences in corporate systems across countries”. In fact, the ordering of countries we observe is also consistent with the ordering of countries by their accounting standards as obtained from Table 5 of Porta, de Silanes, Shleifer, and Vishny (1998). One possible explanation for this ordering is therefore that the stance a rating agency may take in granting a C rating to a debt issuer is to be overly conservative in a country with lenient accounting standards (as was the case with Japan in the past), thus artificially inflating the C rated issuer base. Thus the observed proportion of C rated defaulters would be lower. In contrast, it could be much more lenient in a country with conservative accounting standards (e.g. UK), resulting in the observed proportion of C rated defaulters to be much higher.

5.2.3 BAA upgrade probability

Similarly, one can examine 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. Furthermore, UK issuers systematically have a higher upgrade probability than US issuers. The latter observation may simply be a consequence of the higher mobility of UK issuers, since they also demonstrate a higher C→D default probability than US 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 continuous 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. [22] 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 Jafry-Schuermann metrics.[23] 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

[22]We used the ready-made glm routine in R environment for statistical computing.
[23]Figure 2 may help put this number in perspective.

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. [24] 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 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 higher if heterogeneity in default rates is ignored.

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

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).[25] 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.

[25] Admittedly, the loan default rates may differ significantly from those for corporate bond issuers, but a similar sector-wise decomposition for a 'typical' bank's corporate bond portfolio was not available from publicly available sources. Given this shortcoming, we report results for other portfolio constructions as well. We confirm that unless the bank's portfolio comprises entirely of issuers in the standard issuer profile, heterogeneity matters even if the portfolio does not mirror this 'typical' portfolio.

We found that ignoring sector heterogeneity in default rates increased the risk capital from 5.3% to 6.1% 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. Upon choosing a well diversified benchmark portfolio with equal weights across sectors, the impact of heterogeneity on risk capital increased further (to 20% in proportional terms).[26] Upon choosing a portfolio with high concentration, the impact depended on the sector it was concentrated in. For concentration in the US Utility sector, which has much lower default rates than the standard profile, the impact was to increase the risk capital by 24% in proportional terms. Overall, it seems that default rate heterogeneity impacts risk capital and the magnitude of impact is not negligible. This result may have policy implications and could be worth further examination.

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 strong differences in rating migration behavior between issuers of different industry sectors and countries. Quantifying these differences via rating mobility/stability and also via default/upgrade probabilities yielded several systematic deviations (from estimates for the standard issuer profile viz. US Industrial issuer). This provides strong support and a tool to condition generator estimates on issuer profiles. In a portfolio

[26] The weight of US Industrial issuers in the equal-weights portfolio was lower than in the 'typical' portfolio. Hence a larger component of the portfolio deviated from the standard profile.

context, such a conditioning gave a clearer picture of Value at Risk and risk capital. In particular, it highlighted the possibility that ignoring heterogeneity may increase risk capital by a large percentage. This may be an issue worth examining further, given the upcoming Basel II implementation.

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.

Aretz, Kevin, and Peter F. Pope, 2007, Common Factors in Default Risk Across Countries and Industries, Lancaster University Management School working paper.

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.

Bluhm, Christian, Ludger Overbeck, and Christoph Wagner, 2002, *An Introduction to Credit Risk Modeling.* (Chapman & Hall/CRC).

Chava, S., C. Stefanescu, and S. M. Turnbull, 2006, Modeling Expected Loss, Working Paper, London Business School.

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.

Davydenko, S., and J. Franks, 2006, Do bankruptcy codes matter? A study of defaults in France, Germany and the UK, *working paper, London Business School.*

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, 2007, Credit Rating Dynamics and Markov Mixture Models, Working Paper, Wharton Financial Institutions Center.

Hanson, Samuel G., M. Hashem Pesaran, and Til Schuermann, 2007, Firm Heterogeneity and Credit Risk Diversification, 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.

Israel, Robert B., Jeffrey S. Rosenthal, and Jason Z. Wei, 2001, Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings, *Mathematical Finance* 11, 245–265.

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.

Mayer, C., 1998, Financial Systems and Corporate Governance: A Review of the International Evidence, *Journal of Institutional and Theoretical Economics* pp. 144–165.

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

McNeil, Alexander, and Jonathan Wendin, 2006, Dependent credit migrations, *Journal of Credit Risk* 2, 87–114.

McNeil, Alexander, and Jonathan Wendin, 2007, Bayesian inference for generalized linear mixed models of portfolio credit risk, *Journal of Empirical Finance* 14, 131–149.

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).

Porta, Rafael La, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W. Vishny, 1998, Law and Finance, *Journal of Political Economy* 106, 1113–1155.

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

Schonbucher, Philipp J., 2003, *Credit Derivatives Pricing Models: Models, Pricing, Implementation.* (Wiley Finance).

Stefanescu, Catalina, Radu Tunaru, and Stuart Turnbull, 2007, The credit rating process and estimation of transition probabilities: a Bayesian Approach, Working Paper, London Business School.

Appendix I - Model

Truncation technicalities

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.

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.

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:

$$P(D^*_{i,n_i} > b - T_{i,n_i}) = \exp\left[-\exp(-y_{i,n_i})(b - T_{i,n_i})\right]. \tag{6}$$

Motivation of the transition model

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 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:

$$U_{i,j,k} = z_i'\alpha_{r,k} + \zeta_{i,j,k} \text{ for } k = 1, \ldots, K \text{ and } k \neq r \tag{7}$$

$$U_{i,j,K+1} = \phi_{i,A} + \zeta_{i,j,K+1} \text{ for the default of absorbing state} \tag{8}$$

where $\zeta_{i,j,k}$ are mutually independent error terms that have an extreme value distribution where, without loss of generality, the scale parameter is one. The new rating for the issuer i is $s = \arg\max_{k \neq r} U_{i,j,k}$. 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.

Additional notes on Random Effects

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

$$E(D_{i,j}|\phi_{i,D}) = E[E(D_{i,j}|\phi_{i,D}, \epsilon_{i,j})] = \exp\left(x_i'\beta_s + \phi_{i,D} + \frac{\sigma_s^2}{2}\right). \quad (9)$$

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:

$$E(D_{i,j}) = \exp\left(x_i'\beta_s + \frac{\lambda_D^2 + \sigma_s^2}{2}\right) \quad (10)$$

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

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

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.

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.

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. Then, the Monte Carlo approximation of the integrated generator is

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

Appendix II - Bayesian Estimation

Prior Distributions used

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

$$\beta_s \sim N_p(\mu_{0,\beta,s}, \Sigma_{0,\beta,s}) \quad (12)$$

$$\alpha_{r,s} \sim N_q(\mu_{0,\alpha,r,s}, \Sigma_{0,\alpha,r,s}) \quad (13)$$

$$\sigma_s^2 \sim IG\left(\frac{\gamma_{0,s}}{2}, \frac{\delta_{0,s}}{2}\right) \quad (14)$$

$$\Lambda \sim IW_2(\eta_0, \Omega_0) \quad (15)$$

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

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:

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

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:

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

These facts are used in the MCMC to generate $y_{i,j}$. Given the current value of $y^o_{i,j}$, 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:

$$y_{i,j} > -\ln[-\ln(v)] \tag{16}$$

$$y_{i,j} > y^o_{i,j} - \ln\left[d_{i,j} - \exp(y^o_{i,j})\ln(u)\right] \tag{17}$$

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:

$$\begin{aligned}
\beta_s &\sim N_p(\mu_{n,\beta,s}, \Sigma_{n,\beta,s}) && (18)\\
\Sigma_{n,\beta,s} &= \left(\sum_{i,j:s_{i,j}=s} \frac{1}{\sigma_s^2} x_i x_i' + \Sigma_{0,\beta,s}^{-1} \right)^{-1} && (19)\\
\mu_{n,\beta,s} &= \Sigma_{n,\beta,s} \left(\sum_{i,j:s_{i,j}=s} \frac{1}{\sigma_s^2} (y_{i,j} - \phi_{i,D}) x_i + \Sigma_{0,\beta,s}^{-1} \mu_{0,\beta,s} \right). && (20)
\end{aligned}$$

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 $\phi_{i,A}$:

$$\phi_{i,D}|\phi_{i,A} \sim N\left(\frac{\lambda_{AD}}{\lambda_A^2} \phi_{i,A}, \lambda_D^2 - \frac{\lambda_{AD}^2}{\lambda_A^2} \right). \qquad (21)$$

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

$$\begin{aligned}
\phi_{i,D} &\sim N_p(\mu_{i,D}, \Sigma_{i,D}) && (22)\\
\Sigma_{i,D} &= \left(\sum_{s=1}^{K} \sum_{j:s_{i,j}=s} \frac{1}{\sigma_s^2} + \frac{\lambda_A^2}{\lambda_D^2 \lambda_A^2 - \lambda_{AD}^2} \right)^{-1} && (23)\\
\mu_{i,D} &= \Sigma_{i,D} \left(\sum_{s=1}^{K} \sum_{j:s_{i,j}=s} \frac{1}{\sigma_s^2} (y_{i,j} - x_i' \beta_s) + \frac{\lambda_{AD} \phi_{i,A}}{\lambda_D^2 \lambda_A^2 - \lambda_{AD}^2} \right). && (24)
\end{aligned}$$

The full conditional distribution of σ_s^2 is:

$$
\begin{aligned}
\sigma_s^2 &\sim IG\left(\frac{\gamma_{n,s}}{2}, \frac{\delta_{n,s}}{2}\right) & (25)\\
\gamma_{n,s} &= \gamma_{0,s} + \sum_{i,j:s_{i,j}=s} 1 & (26)\\
\delta_{n,s} &= \delta_{0,s} + \sum_{i,j:s_{i,j}=s} \left(y_{i,j} - x_i'\beta_s - \phi_{i,D}\right)^2 & (27)
\end{aligned}
$$

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:

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

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

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

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

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

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

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

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 $1 - \rho(\phi_{i,A}, \phi_{i,A}^c)$.

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	−2.977E−04	1.967E−04	1.141E−05	4.226E−15	1.358E−13	2.724E−14	3.205E−16	8.962E−05	9.690E−10
AA	2.846E−05	−3.793E−04	2.273E−04	3.307E−06	2.350E−06	1.881E−06	4.269E−08	1.157E−04	1.687E−07
A	1.022E−06	3.748E−05	−3.136E−04	1.598E−04	1.048E−05	2.515E−06	1.274E−07	1.022E−04	4.843E−08
BAA	2.782E−07	3.793E−06	8.438E−05	−3.947E−04	1.564E−04	1.414E−05	4.553E−07	1.345E−04	7.892E−07
BA	7.691E−08	1.929E−06	9.280E−06	1.093E−04	−6.692E−04	2.932E−04	1.165E−05	2.398E−04	3.953E−06
B	1.987E−07	1.937E−06	4.980E−06	1.115E−05	1.330E−04	−7.154E−04	3.030E−04	2.318E−04	2.925E−05
C	1.409E−10	1.295E−09	3.740E−07	7.151E−06	1.457E−05	1.445E−04	−1.043E−03	2.618E−04	6.150E−04
WR	1.571E−07	4.988E−07	2.064E−06	2.461E−06	4.851E−06	5.386E−06	1.550E−06	−1.697E−05	4.080E−10
D	0	0	0	0	0	0	0	0	0

TPM	AAA	AA	A	BAA	BA	B	C	WR	D
AAA	0.8974	0.0635	0.0064	0.0002	0.0001	0.0001	0.0000	0.0324	0.0000
AA	0.0092	0.8716	0.0732	0.0032	0.0010	0.0007	0.0001	0.0410	0.0001
A	0.0004	0.0121	0.8932	0.0514	0.0047	0.0011	0.0001	0.0370	0.0000
BAA	0.0001	0.0014	0.0272	0.8676	0.0473	0.0068	0.0005	0.0487	0.0004
BA	0.0000	0.0007	0.0035	0.0332	0.7863	0.0836	0.0076	0.0826	0.0025
B	0.0001	0.0006	0.0017	0.0043	0.0382	0.7745	0.0805	0.0806	0.0195
C	0.0000	0.0000	0.0002	0.0022	0.0050	0.0387	0.6854	0.0814	0.1871
WR	0.0001	0.0002	0.0007	0.0009	0.0016	0.0018	0.0006	0.9940	0.0001
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	86.09	10.43	0.32	0.01	0.00	0.00	0.00	3.15	0.00
AA	0.31	90.36	5.30	0.09	0.01	0.03	0.00	3.90	0.00
A	0.05	5.06	88.69	2.55	0.15	0.03	0.00	3.46	0.00
BAA	0.13	1.59	14.50	72.95	5.23	0.43	0.04	5.10	0.03
BA	0.02	0.71	5.25	14.15	62.83	10.30	1.20	5.48	0.06
B	0.07	0.10	0.82	4.11	13.60	57.96	13.83	8.62	0.89
C	0.00	0.00	0.03	0.20	0.27	0.72	85.50	2.65	10.62
WR	0.04	0.10	0.14	0.12	0.03	0.01	0.00	99.56	0.00
D	0	0	0	0	0	0	0	0	100

Industrial Sector

	AAA	AA	A	BAA	BA	B	C	WR	D
AAA	89.74	6.35	0.64	0.02	0.01	0.01	0.00	3.24	0.00
AA	0.92	87.16	7.32	0.32	0.10	0.07	0.01	4.10	0.01
A	0.04	1.21	89.32	5.14	0.47	0.11	0.01	3.70	0.00
BAA	0.01	0.14	2.72	86.76	4.73	0.68	0.05	4.87	0.04
BA	0.00	0.07	0.35	3.32	78.63	8.36	0.76	8.26	0.25
B	0.01	0.06	0.17	0.43	3.82	77.45	8.05	8.06	1.95
C	0.00	0.00	0.02	0.22	0.50	3.87	68.54	8.14	18.71
WR	0.01	0.02	0.07	0.09	0.16	0.18	0.06	99.40	0.01
D	0	0	0	0	0	0	0	0	100

Utility Sector

	AAA	AA	A	BAA	BA	B	C	WR	D
AAA	94.62	3.49	1.67	0.04	0.00	0.00	0.00	0.17	0.00
AA	0.32	91.12	6.99	0.38	0.01	0.01	0.00	1.16	0.00
A	0.02	1.67	92.57	4.07	0.15	0.01	0.00	1.50	0.00
BAA	0.02	0.07	3.41	91.24	2.96	0.22	0.10	1.96	0.02
BA	0.02	0.07	0.35	9.37	83.47	3.70	0.90	1.98	0.13
B	0.01	0.02	0.07	2.59	14.92	67.44	9.00	4.12	1.83
C	0.04	0.00	0.11	2.01	3.48	11.38	63.29	7.89	11.80
WR	0.00	0.02	0.09	0.16	0.03	0.01	0.01	99.68	0.00
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.

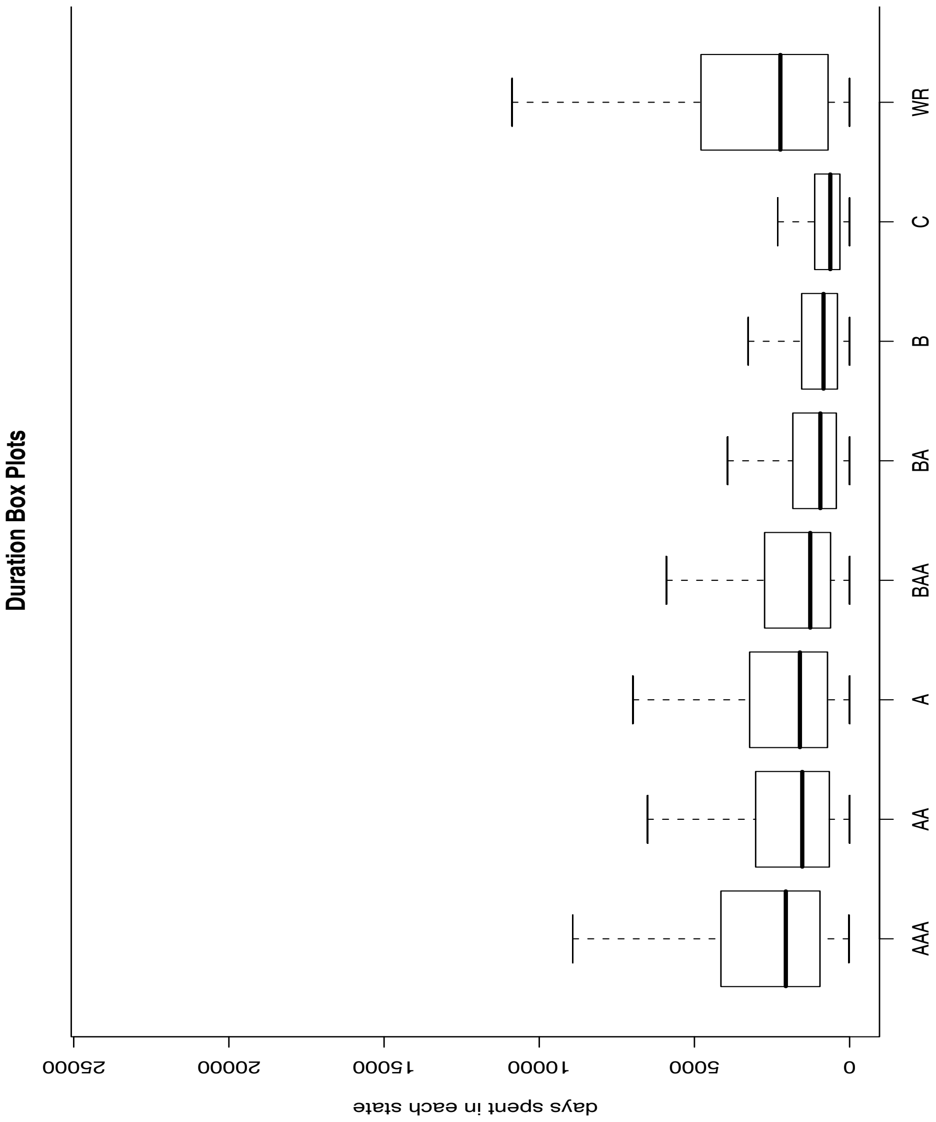
Table 7: Robustness Checks : Discrete time Markov Chain

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

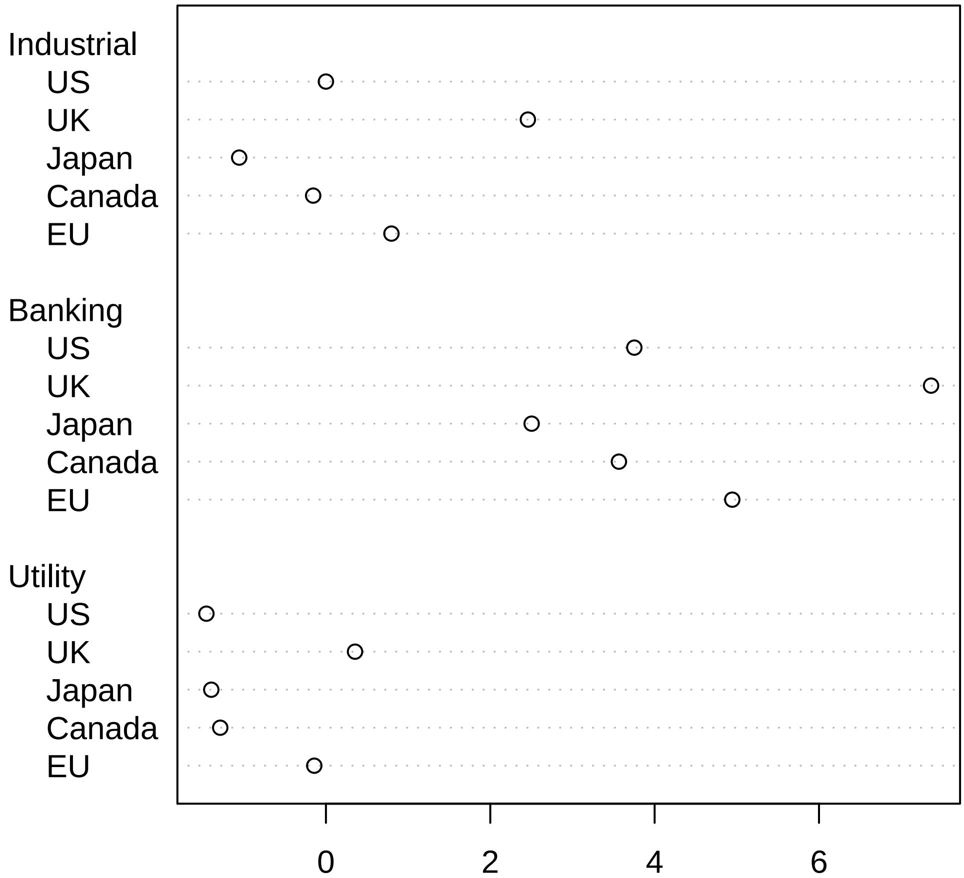
Transition Probability Matrices (over one year horizon) for Industrial and Utility Issuers. Each table above shows 100 times the probability values. The matrices have been estimated using a discrete time Markov chain and discrete (yearly) observation times over the ten year period 1991-2000. They illustrate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

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



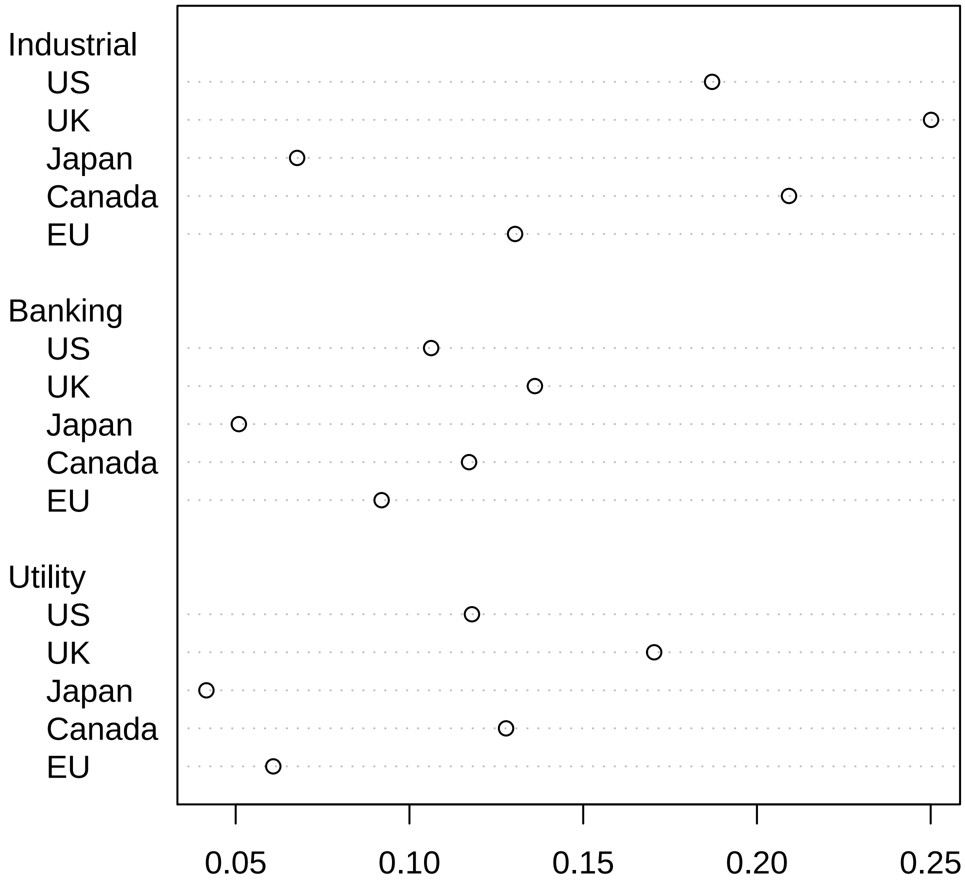
This figure shows the variation of duration times across rating categories. Higher rated issuers have longer and more variable duration times.

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



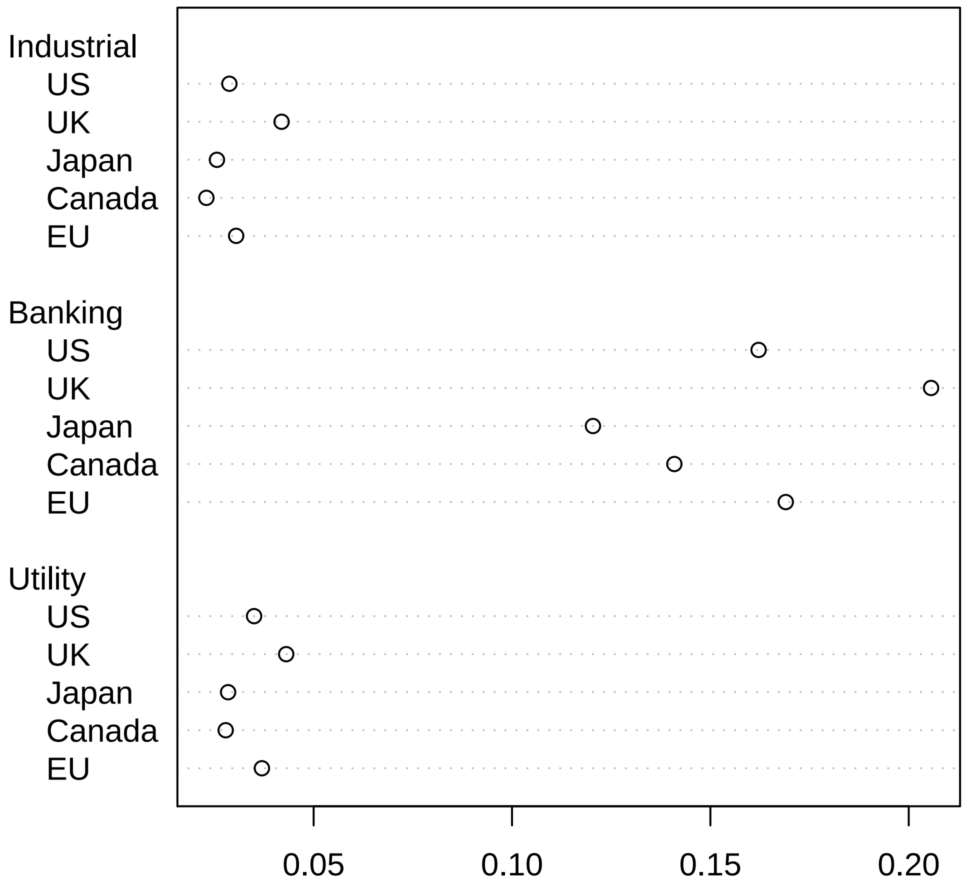
This figure shows the variation of the Jafry-Schuermann metric (proposed in Jafry and Schuermann (2004)) across country-sector profiles on a relative scale. US Industrial issuers make up the standard profile. The figure shows for other important issuer profiles, (100 times) the deviation of the mobility 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



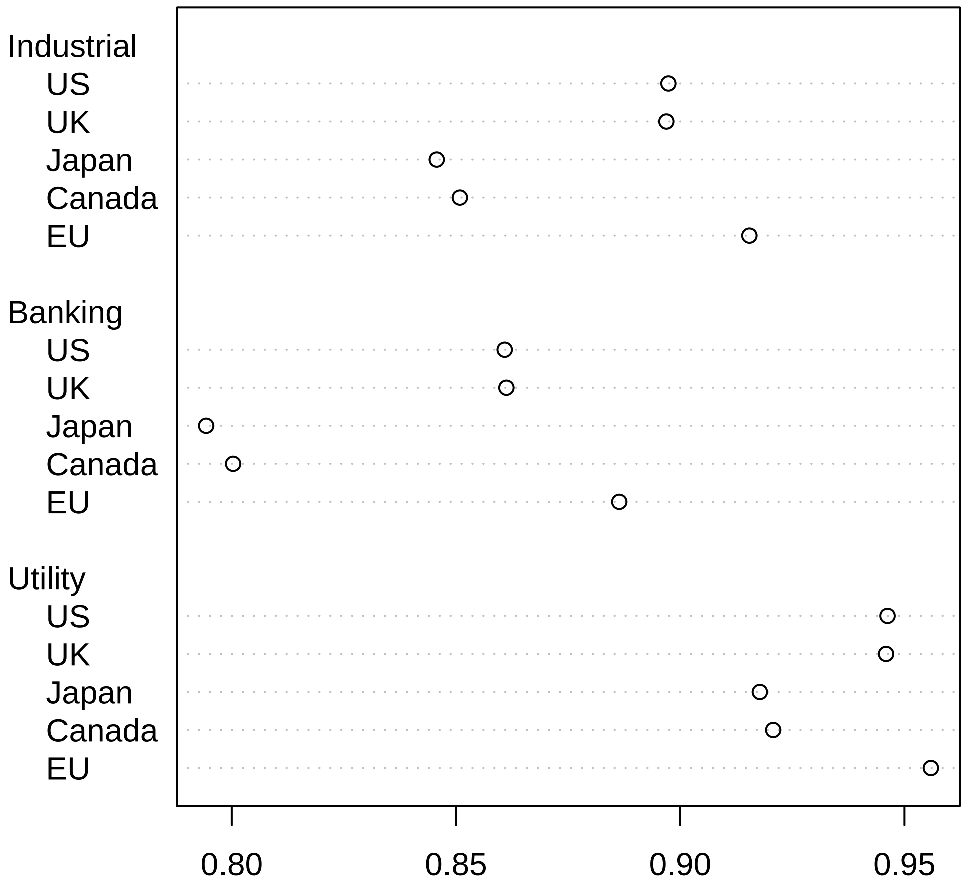
This figure shows the variation in C→D default probability across issuer profiles. The standard profile of US Industrial issuers shows a generally higher default probability. Within each sector, the EU issuers show systematically lower default probabilities than their US and UK counterparts.

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.