Ratings-Based Models of Credit Risk

Haipeng Xing

Outline

1 Rating agencies, rating factors and processes

2 Markov chain for rating migrations
Rating agencies I

Rating agencies have a long tradition in the United States. For example, S&P traces its history back to 1860 and began rating the debt of corporate and government issuers more than 75 years ago. The Securities and Exchange Commission (SEC) has currently designated several agencies as “nationally recognized statistical rating organizations”, including Moody’s KMV, Standard & Poor’s, Fitch, or Thomson BankWatch. Although methodologies and standards differ from one rating agency to another, regulators generally do not make distinctions among the agencies. Furthermore, although there is a high congruence between the rating systems of Moody’s and S&P, different agencies might assign slightly different ratings for the same bond. The ratings given by rating agencies can be served as input data for several credit risk software models such as CreditMetrics of JP Morgan, a system that evaluates risks individually or across an entire portfolio.

Rating agencies II

The rating agencies generally provide two different sorts of ratings: Issue-specific credit ratings and Issuer credit ratings. Issue-specific credit ratings are current opinions of the creditworthiness of an obligor with respect to a specific financial obligation, a specific class of financial obligations, or specific financial program. Issue-specific ratings also take into account the recovery prospects associated with the specific debt being rated. On the other hand, issuer credit ratings give an opinion of the obligor’s overall capacity to meet its financial obligations — that is, its fundamental creditworthiness. These so-called corporate credit ratings indicate the likelihood of default regarding all financial obligations of the firm. The practice of differentiating issues in relation to the issuer’s overall creditworthiness is known as notching. Issues are notched up or down from the corporate credit rating level in accordance with established guidelines.
Some of the rating agencies have historically maintained separate rating scales for long-term and short-term instruments. Long-term credit ratings, i.e., obligations with an original maturity of more than one year, are divided into several categories ranging such as AAA, AA, A, BBB, BB, B, CCC, and D. Ratings in the four highest categories, AAA, AA, A and BBB, generally are recognized as being investment grade, whereas debts rated BB or below generally are regarded as having significant speculative characteristics and are also called noninvestment grade. Ratings from AA to CCC may be modified by the addition of “+” or “-” to show the relative standing within the major rating categories. The symbol R is attached to the ratings of instruments with significant noncredit risks. It highlights risks to principal or volatility of expected returns that are not addressed in the credit rating. In case of default, the symbol SD (Selective Default) is assigned when an issuer can be expected to default selectively, that is, continues to pay certain issues or classes of obligations while not paying others. The issue rating definitions are expressed in terms of default risk and the protection afforded by the obligation in the event of bankruptcy.
The rating process I

Most corporations approach rating agencies to request a rating prior to sale or registration of a debt issue. For instance, S&P assigns and publishes ratings for all public corporate debt issues over $50 million, with or without a request from the issuer; but in all instances, S&P analytical staff will contact the issuer to call for cooperations. Generally, rating agency analysts concentrate on one or two industries only, covering the entire spectrum of credits within those areas. Such specialization allows accumulation of expertise and competitive information better than if, e.g., speculative grade issuers were monitored separately from investment-grade issuers. For basic research, analysts expect financial information about the company consisting of five years of audited annual financial statements, the last several interim financial statements, and narrative descriptions of operations and products. The meeting with corporate management can be considered an important part of an agency’s rating process. The purpose is to review in detail the company’s key operating and financing plans, management policies, and other credit factors that have an impact on the rating. Additionally, facility tours can take place to convey a better understanding of a company’s business to a rating analyst. Shortly after the issuer meeting, the industry analyst convenes a rating committee in connection with a presentation. It includes analysts of the nature of the company’s business and its operating environment, evaluation of the company’s strategic and financial management, financial analysts, and a rating recommendation. Once the rating is determined, the company is notified of the rating and the major considerations supporting it. Is is usually the policy of rating agencies to allow the issuer to respond to the rating decision prior to its publication by presenting new or additional data. In the case of a decision to change an existing rating, any appeal must be conducted as quickly as possible, i.e., within one or two days. The rating committee reconvenes to consider the new information. After the company is
The rating process III

notified, the rating is published in the media — or released to the company for publication in the case of corporate credit ratings. Corporate ratings on publicly distributed issues are monitored for at least one year. For example, the company can then elect to pay the rating agency to continue surveillance. Ratings assigned at the company’s request have the option of surveillance, or being on a “point-in-time” basis. Where a major new financing transaction is planned such as, e.g., acquisitions, an update management meeting is appropriate. In any event, meetings are routinely scheduled at least annually to discuss industry outlook, business strategy, and financial forecasts and policies. As a result of the surveillance process, it sometimes becomes apparent that changing conditions require reconsideration of the outstanding debt rating. After a preliminary review, which may lead to a so-called CreditWatch listing of the company or outstanding issue, a presentation to the rating committee follows to arrive at a rating decision. Again, the

The rating process IV

company is notified and afterwards the agency publishes the rating. The process is exactly the same as the rating of a new issue. Reflecting this surveillance, the timing of rating changes depends neither on the sale of new debt issues nor on the agency’s internal schedule for reviews. CreditWatch and rating outlooks focus on scenarios that could result in a rating change. Rating appear on CreditWatch lists when an event or deviation from an expected trend has occurred or is expected and additional information is necessary to take a rating action. For instance, an issue is placed under such special surveillance as the result of mergers, recapitalizations, regulatory actions, or unanticipated operating developments. Such rating reviews normally are completed with 90 days, unless the outcome of a specific even is pending. However, a listing does not mean a rating change is inevitable, but in cases, the rating change is certain and only the magnitude of the change is unclear. A rating outlook also assesses potential for change, but has a longer time frame
The rating process

than CreditWatch listings and incorporates trends or risks with less certain implications for credit quality. For example, S&P regularly publishes CreditWatch listings with the corresponding designations and rating outlooks to notify both the issuer and the market of recent developments whose rating impact has not yet been determined.

Credit rating factors

Table 1 exemplarily illustrates possible business risk and financial risk factors that enter the rating process of S&P. All categories mentioned there are scored in the rating process and there are also scores for the overall business and financial risk profile. The company’s business risk profile determines the level of financial risk appropriate for any rating category. S&P computes a number of financial ratios and tracks them over time. Generally all of the major rating agencies agree that a rating is, in the end, an opinion and considers both quantitative and qualitative factors. In the world of emerging markets, rating agencies usually also incorporate country and sovereign risk to their rating analysis. Both business risk factors such as macroeconomic volatility, exchange-rate risk, government regulation, taxes, legal issues, etc., and financial risk factors such as accounting standards, potential price controls, inflation, and access to capital are included in the analysis. Additionally, the
Credit rating factors II

anticipated ups and downs of business cycles — whether industry-specific or related to the general economy — are factored into the credit rating.

**Table 1**: Corporate credit analyst factors (Source: S&P Corporate Rating Criteria (2000))

<table>
<thead>
<tr>
<th>Business risk</th>
<th>Financial risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry characteristics</td>
<td>financial characteristics</td>
</tr>
<tr>
<td>Competitive position</td>
<td>financial policy</td>
</tr>
<tr>
<td>Marketing</td>
<td>Profitability</td>
</tr>
<tr>
<td>Technology</td>
<td>Capital structure</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Cash flow protection</td>
</tr>
<tr>
<td>Regulation</td>
<td>Financial flexibility</td>
</tr>
<tr>
<td>Management</td>
<td></td>
</tr>
</tbody>
</table>

Types of rating systems I

An important classification of rating systems is the decision whether a rating system is *point-in-time* or *through-the-cycle*. A point-in-time probability of default describes the actual creditworthiness within a certain time horizon, whereas through-the-cycle probability of default also take into account possible changes in the macroeconomic conditions. The latter will not be affected when the change of the creditworthiness is caused only by a change of macroeconomic variables which more or less describe the state of the economy and which more or less affect the creditworthiness of all borrowers in a similar way. These two types of rating systems are considered to be extreme types of possible rating methodologies. Most rating systems are somewhere in between these two methods.
Credit scoring systems I

Credit scoring systems can be found in virtually all types of credit analysis, from consumer credit to commercial loans. The idea is to pre-identify certain key factors that determine the PD and combine or weight them into a quantitative score. This score can be either directly interpreted as a probability of default or used as a classification system. The first research on bankruptcy prediction goes back to the 1930s (Fitzpatrick, 1932); however, two of the seminal papers in the area were published in the 1960s by Altman (1960) and Beaver (1966). Since then an impressive body of theoretical and especially empirical research concerning this topic has evolved. The most significant reviews can be found in Zavgren (1986), Altman (1983), Jones (1987), Altman and Narayanan (1997), Altman and Saunders (1998), and Balcaena and Ooghe (2006). The latter provide a detailed survey of credit risk measurement approaches. The major methodologies for credit scoring includes linear probability models, logit models, probit models, discriminant analysis models, and more recently, neural networks.

In general, in bankruptcy prediction, two streams of research can be distinguished: the most often investigated research question has been the search for the optimal predictors or financial ratios leading to the lowest misclassification rates. Another stream of literature has been concentrated on the research for statistical methods that would also lead to improved prediction accuracy.

Credit scoring systems II
Discrete-time Markov chain for rating migrations I

Suppose that the state space \( S = \{1, \ldots, K\} \) represents the different rating classes. Hereby, state 1 denotes the best credit rating and state \( K \) represents the default case. Then in the discrete case, the \( K \times K \) one-period probability transition of a Markov chain is

\[
P = \begin{pmatrix}
  p_{11} & p_{12} & \cdots & p_{1K} \\
  p_{21} & p_{22} & \cdots & p_{2K} \\
  \vdots & \vdots & \ddots & \vdots \\
  p_{K-1,1} & p_{K-1,2} & \cdots & p_{K-1,K} \\
  0 & 0 & \cdots & 1
\end{pmatrix},
\]

in which \( p_{ij} \geq 0 \) for all \( 1 \leq i \neq j \leq K \) and \( p_{ii} = 1 - \sum_{j=1, j \neq i} p_{ij} \) for all \( i \). The variable \( p_{ij} \) represents the actual probability of going to state \( j \) from initial rating state \( i \) in one time step.

The estimates of credit transition matrices published by rating agencies use a discrete-time setting. Suppose that there are \( N_i \) firms in a given rating category \( i \) at the beginning of the year and that out of this category \( N_{ij} \) firm migrate to the category \( j \), then one year transition rate is estimated as \( \hat{p}_{ij} = N_{ij}/N_i \) for \( j \neq i \). There are several concerns regarding this discrete estimate. First, if no transition from rating category \( i \) to rating category \( j \) is observed, then the estimated transition probability \( \hat{p}_{ij} \) become 0. Second, though one year is usually the shortest time interval with which a transition matrix is estimated, one frequently needs a transition matrix for a period shorter than one year. Third, it is questionable to assume time homogeneity for credit transition matrices over the long run.
Markov and modulated Markov processes I

Suppose that study subjects move among a number $K > 1$ for discrete states over the course of the study. For the time being, we only consider a homogeneous population with no covariates. Let $A(t)$ be the state occupied at time $t$, $t \geq 0$, and $A(t)$ follows a Markov process. If a randomly chosen individual is in state $i$ at time $t^-$, the transition rate or intensity from $i$ to $j$ at time $t$ is given by

$$d\Lambda_{ij}(t) = P[A(t^- + dt) = j|A(u), 0 \leq u < t, A(t^-) = i] = P[A(t^- + dt) = j|A(t^-) = i], \quad t > 0,$$

which holds for all $A(u), 0 \leq u < t$ with $A(t^-) = i$ and $i, j \in \{0, 1, \ldots, K - 1\}, j \neq i$. The process is memoryless in that only the current state occupied is relevant in specifying the transition rates.

The rates themselves are allowed to depend on the time $t$ since the beginning of the study. It is convenient to define

$$d\Lambda_{ii}(t) = -\sum_{j \neq i} d\Lambda_{ij}(t)$$

so that the row sums of the matrix

$$d\Lambda(t) = [d\Lambda_{ij}(t)]_{K \times K},$$

are all 0.

In the discrete case, there exists a set of times $\{a_k; k = 1, 2, \ldots\}$, where $0 < a_1 < a_2, < \ldots$, at which transitions can occur and

$$P_k = I + d\Lambda(a_k) = \left( P(A(a_k) = j|A(a_k^-) = i) \right)_{K \times K},$$
Markov and modulated Markov processes III

where \( I \) is the \( K \times K \) identity matrix, is the usual one-step probability transition matrix of a nonhomogeneous Markov chain. Let 
\[
P^{(r)} = (p_{ij}^{(r)})_{K \times K}, \text{ where } p_{ij}^{(r)} = P[A(a_r) = j | A(0) = i] \text{ is the } r\text{-step transition probability, } r = 0, 1, \ldots.\]

It is well known that 
\[
P^{(r)} = \prod_{k=1}^{r} P_k = P_1 P_2 \ldots P_r, \quad r = 0, 1, 2, \ldots, \tag{4}
\]

where an empty product is interpreted as \( I \). It should be noted that the order of multiplication matters here since the \( P_k \) matrices will generally not commute. If the chain is homogeneous, so that \( P_k = P \) for all \( k = 1, 2, \ldots \) and \( P^{(r)} = P^r \).

In the continuous case, \( d\Lambda_{ij}(t) = \lambda_{ij}(t) dt \) for all \( i, j = 0, \ldots, q-1 \), so that \( \lambda_{ij}(t), i \neq j \) is the continuous-time intensity function for \( i\)-to-\( j \) transitions, and \( \lambda_{ii}(t) = -\sum_{j \neq i} \lambda_{ij}(t) \). In the homogeneous special case, \( \lambda_{ij}(t) = \lambda_{ij} \) independent of \( t \), and in this case the distribution of the sojourn time in the \( i \)th state is exponential with rate \(-\lambda_{ii}, i = 0, 1, \ldots, K-1 \). Let \( P_{ij}(t) = P(A(t) = j | A(0) = i), i, j = 0, 1, \ldots, K-1 \), and let \( P(t) = [P_{ij}(t)]_{K \times K} \). Analogous to (4), we can write 
\[
P(t) = \mathcal{P}_0 \left[ I + d\Lambda(t) \right] = \mathcal{P}_0 \left[ I + \lambda(u) du \right] \tag{5}
\]

where \( \lambda(u) = [\lambda_{ij}(u)]_{K \times K} \). This product integral is defined in the obvious way:
\[
\mathcal{P}_0 \left[ I + d\Lambda(t) \right] = \lim_{M \to \infty} \prod_{i=1}^{M} \left[ I + \Lambda(u_i) - \Lambda(u_{i-1}) \right], \tag{6}
\]

where \( u_0 = 0 < u_1 < \cdots < u_M = t \) and the limit is taken as \( M \to \infty \) and \( \delta u_i = u_i - u_{i-1} \to 0 \).
Markov and modulated Markov processes V

Estimation of the cumulative intensity functions $\Lambda_{ij}(t)$ and the related transition probabilities $P_{ij}(t)$ proceeds in a straightforward way. Consider a possibly right-censored and/or left-truncated sample of $n$ individuals from the model (2). For $l = 1, \ldots, n$, let $N_{ij}(t)$ be the right continuous process that counts the number of observed direct $i$-to-$j$ transitions for $l$th individual, $i, j = 0, \ldots, K - 1; i \neq j$. Let $Y_i(t)$ be the corresponding at risk process and $Y_{il}(t) = 1\{Y_i(t)=1, A_l(t^-)=i\}$ indicate the $l$th individual is in state $i$ and under observation at time $t^-$, $i = 0, \ldots, K - 1$. Define the filtration or history process as

$$\mathcal{F}_t = \{N_{ijl}(t), Y_{il}(u^+), 0 \leq u \leq t, l = 1, \ldots, n; i, j = 0, \ldots, K - 1\}$$

and suppose that the censoring (and/or truncation) is independent so that

$$P[dN_{ijl}(t) = 1|\mathcal{F}_t^-] = Y_{il}(t)d\Lambda_{ij}(t),$$

which must hold for all $i, j, l$ and $t > 0$. In a manner completely analogous to earlier applications, the Nelson-Aalen estimator of $\Lambda_{ij}(t)$ is given by

$$d\hat{\Lambda}_{ij}(t) = dN_{ij}(t)[Y_i(t)]^{-1}$$

for all $i \neq j$. Let $\hat{\Lambda}_{ij}(t) = -\sum_{j \neq i} \hat{\Lambda}_{ij}(t)$ and $\hat{\Lambda}(t) = [\hat{\Lambda}_{ij}(t)]_{K \times K}$ be the matrix of cumulative intensity estimators. The corresponding estimate of the probability transition matrix is

$$\hat{P}(t) = P_0^t[I + d\hat{\Lambda}(u)].$$

Suppose no that there is a vector of possibly time-dependent basic coariates $x(t)$, which, for convenience, we assume to include $A(t)$, and let $X(t) = \{x(u) : 0 \leq u < t\}$. Continuous-time modulated Markov models can be specified for the underlying (possibly random) intensity function

$$\lambda_{ijl}(t) = \lim_{h \to 0} h^{-1}P[A_l(t^- + h) = j|A_l(t^-) = i, X_l(u), 0 < u < t],$$

which can be estimated by a straightforward extension of the above method.
where the argument $t$ on the left side indicates the basic time scale to be time since on study, as in (2). Parametric or semiparametric models for $\lambda_{ijl}$ can be specified using any of the approaches discussed in previous chapters. For example, one could specify a parametric model depending on a vector of unknown parameters $\theta$. Alternatively, a semiparametric relative risk model could be specified. In the latter case, a natural model to consider is

$$
\lambda_{ijl}(t) = \lambda_{0ij}(t) \exp \left[ Z_l(t)' \beta \right],
$$

for all $i, j, l$ with $i \neq j$ and $t > 0$, where $\lambda_{0ij}$ is an unknown baseline intensity function and the regression parameter vector is taken to be separate for each possible $i, j$ transition. In some instances, only one or two of the transitions $i$ to $j$ may be of interest, whereas in other instances, we may wish to model simultaneously all transition intensities.

Consider a possibly right-censored and/or left-truncated sample of $n$ individuals from the model (8) and define the filtration or history process as

$$
\mathcal{F}_t = \{ N_{ijl}(t), X_l(t^+), Y_l(u^+), 0 \leq u \leq t, l = 1, \ldots, n; i, j = 0, \ldots, K-1 \}.
$$

Suppose that the censoring (and/or) truncation is independent, so that

$$
P(dN_{ijl}(t) = 1|\mathcal{F}_{t^-}) = Y_{il}(t) \lambda_{ijl}(t),
$$

which must hold for all $i \neq j, l$, $\mathcal{F}_{t^-}$ and $t > 0$. Under a parametric model, the full log-likelihood on data over the interval $[0, \tau]$ is

$$
\log L_M = \sum_{i \neq j} \left\{ \int_0^\tau \sum_{l=1}^n \left[ \log \lambda_{ijl}(t; \theta) dN_{ijl}(t) - Y_{il}(t) \lambda_{ijl}(t; \theta) dt \right] \right\},
$$

(9)
Markov and modulated Markov processes IX

which for each fixed \( i, j \) \((i \neq j)\) is of the same form as for a single sample, except that individuals come in and out of observation depending on the risk indicator \( Y_{il}(t) \). This likelihood is again constructed by following the entire study cohort over time as discussed in Section ???. In the parametric case, asymptotic properties for the score function and for the MLE \( \hat{\theta} \) can again be deduced under regularity conditions either with independence assumptions across individuals or through martingale arguments as in Chapters ???.

The relative risk model (8) can be analyzed using partial likelihood arguments based on conditional probabilities of \( dN_{ijl}(t), l = 1, \ldots, n \) given \( \{\mathcal{F}_{t^-}, dN_{ij}(t), i, j \in [0, \ldots, K - 1], i \neq j; t > 0\} \). At each time \( t \in [0, \tau] \) where a transition from \( i \) to \( j \) occurs, the contributing term to the partial likelihood is

\[
P[dN_{ijl}(t) = 1 | dN_{ij}(t) = 1, \mathcal{F}_{t^-}] = \frac{Y_{il}(t) \exp \left[Z_l(t) \beta_{ij}\right]}{\sum_{u=1}^{n} Y_{iu}(t) \exp \left[Z_u(t) \beta_{ij}\right]}.
\]

The log partial likelihood is then

\[
\sum_{all \ i, j} \left[ \int_{0}^{\tau} \sum_{l=1}^{n} Z_l(t) \beta_{ij} dN_{ij}(t) - \log \left[ \sum_{l=1}^{n} Y_{il}(t) \exp \left[Z_l(t) \beta_{ij}\right] dN_{ij}(t) \right] \right],
\]

and the \((i, j)\)th term in this sum is the partial likelihood that arises from a model for time-dependent strata but generalized to accommodate the competing risks or event types indexed by \( j \).
Time-homogenous Markov chain and its generator I

The definition of a continuous-time Markov chain is obtained from the discrete-time definition simply by making the time parameter continuous. Hence for a stochastic process \( \eta \) indexed by a continuous parameter \( t \) with values in a finite state space, the Markov property holds if

\[
P(\eta_t = j | \eta_{s_0} = i_0, \eta_{s_1} = i_1, \ldots, \eta_{s_n} = i_{n-1}, \eta_s = i) = P(\eta = j | \eta_s = i),
\]

whenever \( s_0, s_1, \ldots, s_{n-1} < s \). For a homogeneous Markov chain, the family of transition matrices satisfies

\[
P(u - s) = P(t - s)P(u - t), \quad \text{for } s < t < u.
\]

A time-homogeneous Markov chain on a finite state space can be described by an associated generator matrix, i.e., a \( K \times K \) matrix \( \Lambda \) for which

\[
P(t) = \exp(\Lambda t),
\]

where the exponential of the matrix \( \Lambda t \) is defined as

\[
\exp(\Lambda t) = I + \sum_{k=1}^{\infty} \frac{\Lambda^k t^k}{k!}.
\]

The elements of a generator matrix \( \Lambda = (\lambda_{ij}) \) satisfy

\[
\lambda_{ii} = \sum_{j \neq i} \lambda_{ij}, \quad \lambda_{ij} \geq 0 \quad \text{for } i \neq j. \quad (11)
\]

A row consisting of all zeros corresponding to an absorbing state. If the chain hits such a state, it remains there and never makes another transition. It is usually to work with the default states as absorbing states even if firms may, in practice, recover and leave the default state. If we ask what the probability is that a firm will default before time \( T \), then this can be read from the transition matrix \( P(T) \) when we have defined...
default to be an absorbing state. The transition matrix for time $T$ will only give the probability of being in a default at time $T$ and this probability is typically not the one people are interested in for risk-management purposes.

For the transition matrix (1), the generator matrix $\Lambda$ satisfies (11) and $\lambda_{kj} = 0$ for $1 \leq j \leq K$. One the other hand, under certain regularity conditions (see details in Israel, Rosenthal and Wei, 2000), given a discrete-time one-year transition matrix, the corresponding generator matrix can be calculated using the following expression:

$$\Lambda = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(P - I)^k}{k}.$$  

(12)

If we find a generator matrix with negative off-diagonal entries in a row, we will have to correct this. The result may lead to a generator not providing exactly $P = e^\Lambda$ but only an approximation, though ensuring that from an economic viewpoint the necessary condition that all off-diagonal row entries in the generator are nonnegative is guaranteed. The literature suggests different methods to deal with the problem; see e.g., Jarrow et al. (1997), Israel et al. (2000), Araten and Angbazo (1997), Kreinin and Sidelnikova (2001).