From default rates to default matrices: an application to Brazilian consumer credit

Preliminary summarized version

1. Introduction

Default rate is a term frequently used in financial or economic circles to designate the percentage of borrowers of a given universe (for example, a specific bank portfolio) that have not or will not comply with their credit obligations. Based on past default data, expectations of future delinquency is one of the components that usually explains the level of bank spreads (see BCB, 1999). Also, the monitoring of default rate time series makes it possible to draw relationships with economic cycles (e.g., Bangia et al, 2002) and may assist in constructing anti-cyclical regulations dealing with bank provision or capital (e.g., Jiménez & Saurina, 2006). However, despite the varied utilities of default rates, one may have different concepts of default in mind. Consequently, measurement of default rates (and its implications) may vary sharply depending on the definition employed.

In this study, we exemplify this point by investigating consumer default rates in large Brazilian banks. We relate the concept of delinquency mainly to the idea of arrears in a significant portion of the credit obligation. However, instead of proposing a precise definition (for example, 90 days of arrears), we work simultaneously with different past-due ranges by means of transition matrices (to be named hereafter default matrices). In this way, delinquency is depicted in a multidimensional way, including several percentage figures that designate migrations among various brackets of arrears. Though this provides a more complete picture of the behavior of delinquency, comparisons between different portfolios is no longer an easy task and requires specific metrics.

Differences in default matrices of different banks can be explained mainly by two factors: different market niches and variations in the efficacy of recovering credits in arrears. The first factor has a greater weight in explaining why performing obligations of different banks have significantly different chances of being included in a specified bracket of arrears, while the second factor would explain why obligations already in arrears of different banks have different chances of shifting into a bracket of lesser arrears.

2. Data

The database used in this study is based on data drawn from the Brazilian Public Credit Register. It consists of time series of regulatory credit risk classifications of consumer loans at four large Brazilian banks from January 2003 through January 2008. The database includes loans started before January 2003, but still in effect during the time span of the study, or started within that period. Almost all of the borrowers with operations initiated in one bank do not remain in that bank until the end of the time period, in January 2008, for a series of reasons discussed in detail in the complete version of this article (several of which are related to the policies adopted by the banks themselves).

In order to increase comparability between classifications and, therefore, among default matrices of different banks, we restrict this study to small operations, that, according to Brazilian regulation, can be subject to review solely as a result of arrears. Besides, we carry out reclassifications of the original regulatory classifications in order to increase the interpretation of the former as occurrences of brackets of arrears, according to the table below[1].

Table – Interpretation of classifications as arrears

Classification	A	AR	B	C	D	E	F	G	H
Arrears (days)	-	renegotiated	15-30	31-60	61-90	91-120	121-150	151-180	>180 or written-off

[1] More on that on the complete version of the paper.

3. Methodology

3.1 Estimating matrices

This study focuses on the estimation of banks' consumer credit default matrices. To accomplish that, the time series of classifications of each bank is seen as a realization of a Markov chain of nine states ("A" through "H", according to the previous table) in discrete or continuous time, depending on the estimation technique employed.

The simplest and most used estimation technique is the cohort method, based on discrete time. The technique is widely employed by rating agencies and the academic literature. Given N_i borrowers with a given classification i at the start of the time horizon considered, suppose that N_{ij} of these present classification j at the horizon end T. Then the transition probability is estimated by $P_{ij}=N_{ij}/N_i$.

If the transition process is also assumed homogeneous, one can use the multinomial estimator, in which N_i and N_{ij} are collected over the course of various sample periods of duration T. In this case, $P_{ij}= \Sigma N_{ij}/\Sigma N_{i.}$

Estimators of the discrete type permit the construction of analytical confidence intervals for the elements of the default matrices. Due to the significant number of borrowers upon which this study is based, we safely adopt the normal approximation for the construction of the intervals.

On the other hand, the estimation based on survival analysis, in continuous time, makes use of the transitions observed at shorter frequencies than horizon T, assuming that these transitions are generated by a process homogeneous or not. In the homogeneous case, estimation by survival analysis turns into estimation of the generator matrix **G** of the chain, which allows the production of transition matrices for any forecasting horizon $t\neq T$, according to the equation below.

$\mathbf{P}(t)=\exp(\mathbf{G}t)$, in which $\mathbf{P}(t)\equiv(P_{ij}(t))$ is the transition probability matrix for horizon t.

The elements of **G** satisfy $g_{ij} \geq 0$ for $i\neq j$, $g_{ii}=-\sum g_{ij}$ and are estimated by maximum likelihood by:

$$g_{ij} = N_{ij}\times M/\int_0^T Y_i(t)dt,$$

where M is the number of months in horizon T, N_{ij} is the total number of transitions from i to j observed in the base and $Y_i(t)$ is the number of borrowers of classification i in month t.

The non-homogeneous continuous time case is equivalent to applying the cohort method for the shortest observation frequency, monthly in the case at hand, in order to estimate monthly transition matrices. Then, a horizon-T matrix is formed by multiplying various estimated monthly matrices. This is in fact an application of the Aalen-Johansen estimator and the resulting matrix so obtained is specific to the corresponding period T.

By using all of the information available on the database, the continuous time estimations have three major advantages in relation to the discrete methods, as documented by Lando & Skodeberg (2002). First, non-null probabilities are generated for transitions that have not occurred for any fixed set of borrowers, but that are plausible through intermediate transitions that occurred for different sets of borrowers. Second, transitions of borrowers that do not remain in the base during all of the months, either due to withdrawal prior to the final month or entry subsequent to the initial month, are used in the method, producing more efficient estimations. Third, transition matrices are generated for arbitrary time horizons with greater ease, particularly in the homogeneous case.

Finally, Gagliardini & Gouriéroux (2005) propose a procedure that is somewhat different from the estimators described above. In a context in which the horizon T matrices are assumed stochastic, albeit i.i.d., the authors demonstrate that it is the average of the various sample matrices of the different consecutive periods of duration T that produces the appropriate estimator. In particular, when each of those is estimated by cohort, the simple average, instead of the weighted average given by the multinomial estimator, is the appropriate estimator.

3.2 Comparing matrices

In order to compare how different are delinquencies and their dynamics among various banks, metrics for transition matrices must be considered. Jafry & Schuermann (2004) examine alternative proposals of metrics, with the goal of measuring the average "quantity" of mobility embedded into the matrices (mobility being understood as the probability of migration to a classification different from the original classification) and suggest two metrics: M_{SVD} based on singular values and $M_{EUC,}$ based on the Euclidean distance.

$$M_{SVD} \equiv \frac{\sum_{i=1}^{N} \sqrt{\lambda_i\left((P-I)^T(P-I)\right)}}{N} \qquad M_{euc} \equiv \frac{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(P_{ij}-I_{ij}\right)^2}}{N/\sqrt{N-1}},$$

where N is the number of classifications, λ_is are the autovalues of the matrix in parentheses and P is the transition matrix under consideration.

Two matrices P_1 and P_2 can then be compared through[2]:

$$\Delta M_{SVD} \equiv |M_{SVD}(P_1) - M_{SVD}(P_2)| \;\text{ or }\; \Delta M_{euc} \equiv |M_{euc}(P_1) - M_{euc}(P_2)|$$

However, even for the selected metrics, it is difficult to capture all the dimensions of the concept of mobility in a single scalar. Indeed the metrics suggested are not able to distinguish between migrations to better classifications and migrations to worse classifications, or between extreme migrations and shorter migrations. In the first case, note that both M_{SVD} and M_{euc} generate the same value for P and P^t, while, in the second case, Jafry & Schuermann (2004) demonstrate that both M_{SVD} and M_{euc} may fail to generate larger values for matrices with migration probability distributed further from the diagonal. This study proposes solutions for both problems. In the first case, it seeks to work with triangular matrices. In the second, it makes use of the concept of expected utility to be measured by the average opportunity cost of the operations in arrears[3].

4. Results and conclusions

Preliminary results follow:

a) The credit risk applied literature on transition matrices usually focuses on matrices of rating agencies. Compared to them, our proposed default matrices display much less probability on the diagonal and strong probability concentration on the extreme columns. Also, probabilities of delinquency and recovery are usually monotone in the arrears classification, as expected.

b) Compared to the homogeneous survival estimation, the multinomial estimation implies lesser probabilities of delinquency and recovery, but greater probabilities of migration between distinct brackets of arrears, for all of the banks analyzed. One explanation for the latter fact may be the presence of downward momentum.

c) The differences found in the pair of estimations mentioned above are significantly greater than those found between the homogeneous and non-homogeneous versions of survival estimation, so that, as in Jafry & Schuerman (2004) the efficient gains of survival estimation are more important than a possibly false hypothesis of homonegeneity. However, the differences between the two survival estimators are greater than in the case of rating agencies (see, Lando & Skodeberg, 2002). This suggest that time specific shocks, for example related to changes in the credit policy of the institution, have a material

---

[2] To be precise, it is ΔM, not M, that represents a metric (or better yet, a pseudo-metric) in the space of the transition probability matrices.

[3] For greater detail, see the complete version of this article.

impact on the results and demonstrate that the Aalen-Johansen estimator may be a useful tool for closely monitoring the behavior of delinquencies on a bank level.

d) Based on the estimated analytical confidence intervals the analysis indicates that, even in the optimistic case without assumption of time variation of the matrices and making use of a large number of observations, certain brackets of large arrears cannot be statistically discriminated with respect to both delinquency and recovery migrations. Depending on the bank, this phenomenon occurs for the pairs (E,F), (G,H) or (F,G). As a policy implication, this means that the goal of risk discrimination (e.g. by requiring different provisions) for loans with significant past due may be unfeasible.

d) The comparative analysis among banks identifies, through the use of default matrices, a pair of banks with similar risk behaviors and another pair with distinct risk behaviors. Also, tests on equality of means indicate 54 different matrix elements for the first pair and 68 for the second. The results also indicate whether the similarity or distinction is derived mainly from delinquency, recovery or opportunity cost.

e) The estimation of default matrices over time indicates that: I) time variations are pronounced (in particular, for the migrations A to >=D and A for >=E) so that the hypothesis of time constant matrices is rejected, II) time variations are much greater than the difference between the estimators of simple and weighted average, III) metrics of default matrices, as well as particular transitions, show sharp differences between banks, as the only element common to them is the downward direction from early 2006 to the end of 2007. These results indicate that differences in market niches, growth strategies, renegotiation policies, among others, are important in any analysis of the four banks during the period extending from 2003 to 2007.

In conclusion, one should remark that the series of results produced, coupled with specific knowledge about the banks themselves by the supervisory authority, should allow it to better understand the behavior of realized delinquency over time, without missing relevant information. It could also permit a kind of subjective validation of the statistical techniques themselves, based on other sources of information available to the supervisory authority, such as on-site supervision.

5. References

BANCO CENTRAL DO BRASIL (1999). **Juros e Spread Bancário no Brasil**.

BANGIA, A.; DIEBOLD, F. X.; KRONIMUS. A.; SCHAGEN C.; SCHUERMANN T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. **Journal of Banking and Finance,** n. 26, p. 445-474.

CHRISTENSEN, J. H. E.; HANSEN E.; LANDO D. (2004). Confidence sets for continuous-time rating transition probabilities. **Journal of Banking and Finance**, n. 28, p. 2575-2602.

GAGLIARDINI, P.; GOURIÉROUX C. (2005). Migration correlation: Definition and efficient estimation. **Journal of Banking and Finance**, n. 29, p. 865-894.

JAFRY, Y.; SCHUERMANN, T. (2004). Measurement, estimation and comparison of credit migration matrices. **Journal of Banking and Finance**, n. 28, p. 2603-2639.

JIMÉNEZ G.; SAURINA J. (2006). Credit Cycles, Credit Risk and Prudential Regulation. **International Journal of Central Banking**, vol.2, no. 2, p. 65-98.

LANDO D.; SKODEBERG T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. **Journal of Banking and Finance**, n. 26, p. 423-444.