Modeling Ratings Migration for Credit Risk Capital and Loss Provisioning Calculations

by Jorge Sobehart and Sean Keenan

Reliable loss prediction requires both robust estimation methods and accurate data. This article presents a way to leverage ratings agency data that can provide greater flexibility and stability of results in simulation-based estimates of future portfolio losses.

Based on a simple behavioral model that quantifies the structural relationships in historical default frequencies and transition rates for different ratings, this technique leads analysts to hypothetical transition matrices for portfolio loss simulations that preserve the basic relationships observed in the historical transition and default rates reported by the ratings agencies, allowing for unlimited sampling. The matrices can also be linked to macroeconomic factors to mimic the dynamics of credit cycles and economic shocks, allowing for richer descriptions of plausible future scenarios and what-if scenario analysis that goes beyond the limitations of historical data.

© 2004 by RMA. Jorge Sobehart and Sean Keenan are vice presidents and senior analysts at Citigroup Risk Architecture, Greenwich, New York. The analysis and conclusions in this article are those of the authors. Citigroup is not responsible for any statement or conclusion herein, and opinions or theories presented herein do not necessarily reflect the position of the institution.
Methodologies, analysts continue to rely on data from the major ratings agencies for default rates, ratings migration rates, and other key statistics. Despite recurring and somewhat troubling issues regarding the meaning and consistency of ratings, regulators tend to be more accepting of methodologies based on agency data because of the agencies’ long and well-documented ratings histories. This data may indeed be deeper and may conform better to an accepted standard than banks’ own internal ratings histories, yet the depth of agency data generally falls short of what’s needed for the Monte Carlo-based economic risk capital estimation techniques in widespread use today.

The Shortcomings

The simplest portfolio loss model assumes that ratings transition probabilities are stable across obligor types and across the business cycle, and that a single set of average historical ratings transition and default rates is all that’s needed to characterize potential future losses. However, there is ample evidence that credit migration and the ratings process depend on a number of factors, such as the state of the economy—for example, the probability of downgrades and defaults is greater in a downturn than in an upturn. Moreover, historical data is volatile; thus, the average-rate approach will underestimate potential tail loss—the very thing we want to measure with precision. A slightly more sophisticated alternative is to use observed annual historical-ratings transition rates as a sample from which to draw plausible future credit migration scenarios to simulate the forward loss distribution. The main drawback of this method is the small number of historical-ratings scenarios on which to draw. Accurate Monte Carlo simulations for large portfolios usually require tens—or even up to hundreds of thousands—of random draws. However, because historical scenarios number only in the tens, the simulated loss distribution will tend to be lumpy as tail losses bunch up around the worst year from the historical period. Clearly, this problem cannot be overcome by increasing the number of Monte Carlo simulations.

A Behavioral Model of Risk Perception

A different approach is to directly model the relationship between transition probabilities and macroeconomic factors and then simulate plausible ratings migration patterns over time by generating various macroeconomic conditions. To do this, we need a behavioral model of how risk ratings are assigned. Let’s begin with the observation that ratings are opinions of credit quality, representing different degrees of belief in the credit quality of the firm. Agency statistics, such as default and transition frequencies, are merely by-products of this rating-assignment process, rather than properties inherent to the ratings themselves. Analysts’ judgments, meanwhile, are based on a combination of qualitative and quantitative comparisons of the credit risk they perceive. Even if specifically attempting to arrive at a default-probability calculation, the analyst cannot be sure of the precise relationship between the risk factors affecting the obligor and his or her own mental model of risk perception, which may lead to errors in risk assessment. Thus, even with complete and perfect information on the obligor’s risk exposure, the analyst would still face “model risk” because of judgment. Any qualitative comparison between two risk exposures is clearly probabilistic in nature since it relates to uncertain events. Unfortunately, analysts’ perceptions of the probability of default, expected losses, and future ratings revisions are not publicly available and therefore cannot be tested. However, we can construct a behavioral model for the average perceived risk that can be calibrated with historical default and transition rates associated with a given risk perception (rating at a given point in time) assuming that the ratings are unbiased estimates of the average (ex-ante) analyst’s perception of...
the risk criterion.

The basic argument underlying the model presented here is that, at the fundamental level, the risk-assessment process is based on a relative comparison between perceived risk severities for pairs of risk exposures—the obligor’s risk exposure and that of its peers, or a mental estimate constructed by the analyst. More precisely, let $E$ be the obligor’s risk exposure with severity $p(E)$ (for example, the expected probability of default) and $R$ be the resultant average risk perception (risk rating). If the absolute perception of two risk exposures differs by a just noticeable amount when separated by a given relative increment of risk severity, then when the risk exposures are increased, the perceived risk increment must be proportionally increased for the difference in perception to remain just noticeable. From this relationship we find that the relation between the severity $p(E)$ and the risk perception $R$ becomes approximately:

$$\log\left(\frac{p}{1-p}\right) = \frac{1}{a}(R - b) + c$$

Equation 1

Here, the parameter $a$ is the psychological sensitivity to variations of the risk exposure, and $b$ is a reference value for the maximum risk severity corresponding to the maximum risk exposure. The term $c$ reflects judgment errors for individual risk exposures. In the following we focus only on the average risk perception, neglecting the error term $c$.

Equation 1 depends on the time horizon over which the risk is being assessed. For a given time horizon $T$, the parameter $b(T)$ in Figure 1 provides the reference risk rating used in the comparison of risk exposures, and the parameter $a(T)$ provides the sensitivity of the risk perception to changes in risk severity. If we assign a cardinal value to agency ratings (e.g.: AAA=0, AA+=1, AA=2, etc.), then the parameter $b(T)$ provides a snapshot of the perceived risk rating at various time horizons, and $a(T)$ reflects the sensitivity of this perception to changes in risk severity over time.

Figure 1
Logarithm of the Odds of Default for Different Ratings and Time Horizons for 1983-2003, and Parameters of the Curve Fitting
a(T) provides the number of notches required to increase the odds nearly threefold (actually an increase in a factor e = 2.73).

For the moment, let’s assume that the probability of default is an unbiased estimate of the analyst’s perception of risk severity and that the historical default rates are unbiased estimates of the (ex-ante) probability of default for a given risk perception (in the next section we introduce a different risk perception criterion and analyze the consistency between the two). The upper panel of Figure 1 shows the empirical relation between the risk perception R (rating) and the average default rate for corporate issuers for different time horizons during the period January 1983 to December 2003 for Moody’s and S&P ratings, expressed in terms of the logarithm of the odds of default: log(Pd/(1-Pd)) given the default rate Pd.

The quasi-linear trend between ratings (perceived risk) and the logarithm of the odds of default is an example of the Weber-Fechner law observed in psychology and physiology, which indicates that intuitive human sensations tend to be measured in relative terms leading to logarithmic or power functions of the stimulus. For example, normal conversation may appear two times as loud as a whisper, whereas its true acoustic intensity is actually hundreds of times greater. The familiar decibel scale used in audio equipment relates perceived loudness to the objective concept of intensity in the same way that the risk perception R (rating) measures the likelihood of default.

Notice that if Equation 1 were strictly true, this would indicate that the separation between ratings grades had a consistent meaning in terms of the relative change in the likelihood of default. For example, a one-notch downgrade would indicate an e-fold increase in the likelihood of default from the previous rating, and a one-notch upgrade would indicate an e-fold reduction in the likelihood of default. In general, this is not the case, but the approximation holds reasonably well.

The lower panels of Figure 1 show the parameters of the fitting for Equation 1 for both ratings agencies. For short time horizons, the linear fitting provides a rating dispersion in the order a(T=1 year) ≈ 1.5 notches for an e-fold increase in the odds of default, and a rating reference b(T=1 year) ~ 18-19 notches (roughly a CCC/Caa rating). These values are consistent with the notions that agency ratings are reasonably accurate within one or two notches for short-term horizons, and that CCC/Caa ratings show similar characteristics to defaulters. The proximity of fitted parameters for both leading ratings agencies shown in Figure 1 is remarkable. Another important aspect of Figure 1 is that the linear relationship between the empirical log of the odds and ratings breaks down at the top of the rating scale where defaults are extremely rare.

This casts doubt on whether investment-grade analysts are responding primarily to changes in perceived default risk, as discussed in the following section.

Transition Risk as Perceived Risk

Obviously, given the sheer number of ratings assignments relative to the number of defaults, it might be unreliable to infer the behavioral underpinnings of the rating scale on the basis of default risk alone. In fact, the rating scale may have a more consistent basis in transition risk. From the investor’s perspective, the likelihood that a rating may be raised or lowered over a particular time horizon creates risk primarily associated with the obligor’s performance. But from the analyst’s perspective, ratings revisions may also reflect the inaccuracy of the initial rating. In noncontroversial cases, revisions may result from catastrophic events, changes in regulations, or unanticipated corporate actions. In other cases, the true value of new information is subject to interpretation, and ratings revisions (their timing and magnitude) can be interpreted as evidence that the previous rating had been assigned in error.

Common signals of inaccuracy include complaints from issuers and investors, persistent inconsistencies between ratings and credit
spreads, and most importantly, frequent revisions of ratings that appear to be reactive instead of anticipatory.

Agencies and financial institutions measure ratings volatility over time with ratings transition matrices. The elements of the transition matrix represent the likelihood of either remaining in the same rating or moving up or down to a new ratings category. Transition matrices give us an independent set of frequencies with which to calculate the odds-ratios for ratings revisions needed to test our behavioral model of perceived severity using this new risk perception criterion. The test is straightforward. Instead of assuming that the perceived severity is the obligor’s default probability, let’s assume that it is the transition risk. Given a time horizon \( T \), for each initial rating \( R \) we simply calculate the odds-ratio of a downgrade of \( W \) notches to a worse rating (higher value in notches) \( W \) using the empirical transition frequencies as proxies for the transition probabilities (rating revisions) \( P_{RW}(T) \). We then approximate the logarithm of the odds-ratio of ratings revisions for each position as a linear function of the magnitude of ratings revisions using the following extension to Equation 1:

\[
\log \left( \frac{P_{RW}}{1 - P_{RW}} \right) \approx \frac{1}{a_d} (R - W) - b_d
\]
The model estimates and the actu-

Notice the divergence between 

rating including the default state.

from a given rating to any other 

the transition element (1-

Equation 2. Figure 2 also shows 

lines represent the model in 

transition frequencies and the 

bols represent the empirical rating 

data. In Figure 2 the small sym-


any other rating for the period 

S&P one-year historical transition 

creditworthiness.

severe two-notches revision of 

three times as frequent as a more 

example, a one-notch revision of 

severity of ratings transitions. For 

of the relative change in the 

ship between the default frequen-

tures of Figure 2 is the relation-

per the line for rating upgrades).

better the credit quality, the steep-

as credit quality improves (the 

large variability of the transition 

rates over time. For the BBB/Baa 

rating, the inverted V-shaped pat-

tern seems symmetric. However, 

for investment-grade firms, analysts 

seem to be more reluctant to 

upgrade ratings than to down-

grade them creating asymmetries 

in the scale, (a_u=a_d ). In contrast, 

for the segment of very low credit 

quality, analysts seem to be less 

biased to either downgrade or 

upgrade firms, although the fit is 

relatively poor. From Figure 2 we 
can infer that the sensitivity to rat-

ings downgrades is roughly con-

stant, while ratings upgrades 

become increasingly more difficult 
as credit quality improves (the 

better the credit quality, the steeper 

the line for rating upgrades).

One of the most striking fea-

ures of Figure 2 is the relation-

ship between the default frequen-

cy and the line implied by the 

behavioral model in Equation 2.

In each case, the empirical default 

rate (open triangles) lies above 

the line and represents one of the 
largest deviations from the model. 

That is, the expected default rate 
based on the credit migration pat-

tern is significantly lower than the 

observed default rate. One expla-
nation is that transition rates 

reflect expected losses as opposed 
to default risk. Note, however, 

that the recovery value implied 

from the extrapolated transition 

rates is much lower than the aver-

ger reported by the agencies. 

Another plausible explanation is 

that analysts have full control over 

the assigned ratings, and therefore 

transition rates between ratings are 

affected mainly by analysts’ deci-

sions. For example, the decision to 

maintain a rating may not neces-
sarily reflect a stable credit quality 

outlook for the issuer, but its pur-

pose might be a reluctance to fur-

ther limit the firm’s access to cred-

it markets. In contrast, analysts 
have no control over which issuers 

actually default on their obliga-

tions, and therefore inconsist-

encies, between the assigned ratings 

and the transition rate to the 
default state could easily arise.4

Credit Migration and Portfolio 

Risk

The model introduced here 

provides a simple yet sound 

means of constructing ratings 

transition matrices that preserve 

the basic relationships observed in 

the historical transition and 

default rates reported by the rat-

ings agencies. As discussed above, 

this is critical since there is a lim-

ited number of historical transition 

matrices—a data set inade-

quate for providing the wide spec-
trum of scenarios required to obtain robust Monte Carlo simulations for economic capital and reserves at the high confidence levels required. The ability to construct ratings migration scenarios with internally consistent structural relationships and variability for different economic conditions (obtained, for example, from the distribution of fitting errors for Equations 1 and 2) can help to analyze situations beyond the limitations of historical data. Moving one step further, by associating the time paths of the key parameters \( a, b, a_d(R,T), b_d(R,T), a_u(R,T) \) and \( b_u(R,T) \) with macroeconomic and credit market data, transition matrices generated using Equation 2 can be linked to the dynamics of credit cycles and economic shocks, allowing for the kind of what-if scenario analysis and stress testing required for the active management of credit risk.

The steps for the simulation of ratings migrations can be summarized as follows:
1. Obtain historical time series of ratings migration matrices and macroeconomic variables.
2. Construct the models in Equation 1 and 2 for each ratings transition matrix.
3. Regress the structural coefficients of the model and the standard deviation of unexplained errors on lagged macroeconomic variables to identify systematic drivers of credit migration.
4. For each macroeconomic scenario, simulate random ratings migration patterns using the structural models 1 and 2 and the distribution of unexplained errors.

Simple econometric models for the coefficients in Equations 1 and 2 and the standard deviation of unexplained errors using lagged multiple regression analysis illustrate this process. The selected descriptive variables are the annual GDP growth rate, relative changes in short-term lending rates, and the ratio of the number of speculative grade issuers over the total number of issuers lagged two years. The latter variable describes the relationship between debt issuance and credit quality, and the overall dynamics of the credit cycle. Even this limited set of variables can allow for the analysis of regional diversification in credit portfolios by simulating ratings migration for individual countries conditioned on their current position in their economic cycle.

Figure 3 shows the evolution of the historical and estimated

![Figure 3](image-url)
one-year default rates for S&P, and the standard deviation of the scenarios produced by the model for selected ratings for U.S. issuers. Similar results are obtained for Moody’s ratings. From the viewpoint of forecasting default rates, the accuracy of this simple model may seem relatively modest. However, its purpose is not to forecast specific outcomes, but to generate large numbers of plausible scenarios consistent with given macroeconomic conditions. From the portfolio simulation viewpoint, the additional uncertainty allows for a realistically wide spectrum of alternative what-if scenarios for given values of the economic drivers through the credit cycle or for stress scenarios. This is exactly what is really needed for economic capital simulations, since the main goal of portfolio simulation is to study the tail end of the loss distribution, and the characteristics of this tail are often driven by the irregularities and small sample size effects found in published transition data.

Conclusions

Current financial institutions’ credit risk assessment processes often include inputs both from quantitative, statistical models and from traditional fundamental analysis. Frequently, however, models for estimating portfolio loss provisions and economic risk capital retain a dependence on certain key agency statistics such as default rates and transition rates. Unfortunately, this historical data is still generally insufficient for robust estimation of portfolio losses through Monte Carlo techniques.

The behavioral model offered in this article is capable of quantifying the empirical distribution of default rates and transition rates for different ratings categories in a sensible and parsimonious way. The model can be used to construct a rich set of ratings transition scenarios that go beyond the limitations of historical data, while preserving the empirically observed structural relations between ratings transitions and default rates. It also provides a link between economic conditions and credit migration scenarios defined in terms of transition and default frequencies. These capabilities allow institutions to conduct more robust and more controlled Monte Carlo simulations for risk capital calculations, portfolio loss provisioning, and what-if and stress scenario analysis.

The authors may be contacted by email at jorge.r.sobehart@citigroup.com; sean.keenan@citigroup.com.

Notes

1 Agency studies have recently acknowledged that ratings themselves may be insufficient for predicting future migrations and default rates (Hamilton & Cantor 2004).

2 Agencies typically direct researchers to the historical rating statistics, but do not go so far as to say that these performance statistics were what the ratings originally intended to express.

3 This covers the entire history of ratings on the numerically modified scale.

4 However, a rating change can produce additional financial distress through ratings trigger covenants in the firm’s obligations.