Point-in-Time versus Through-the-Cycle Ratings [1]

Authors:

Scott D. Aguais, Lawrence R. Forest, Jr., Elaine Y. L. Wong, Diana Diaz-Ledezma [2]

[1] The authors would like to acknowledge the many Basel and credit risk related discussions they have had with various members of the Barclays Risk Management Team over the last year. The authors also want to thank, Tim Thompson, Julian Shaw and Brian Ranson for helpful comments. Moodys|KMV also deserves thanks for providing an historical data set of five-year KMV EDF term structures for all of their counterparties. Finally, we thank Zoran Stanisavljevic, for his un-ending credit data management support for our credit research and modelling efforts. All errors remain the responsibility of the authors. The views and opinions expressed in this chapter are those of the authors and do not necessarily reflect the views and opinions of Barclays Bank PLC.

[2] Scott Aguais is Director and Head of Credit Risk Methodology. Lawrence Forest and Elaine Wong are Associate Directors, and Diana Diaz-Ledezma is a Manager, all on the Credit Risk Methodology Team. All four work at Barclays Capital, the Investment Banking unit of Barclays.

I. Introduction:

To qualify for the advanced internal ratings-based (AIRB) status under the Basel II Capital Accord, "internationally active banks" ("banks") must document the validity not only of their probability of default (PD) estimates but also of the Loss Given Default (LGD) and Exposure at Default (EAD) measures that they plan to use as alternatives to the foundation internal ratings-based (FIRB) assumptions.[3] The PD, LGD, and EAD building blocks are central to most aspects of credit risk management – structuring, monitoring, pricing transactions, establishing loan loss provisions, and assessing credit portfolio risk. The likely reduction in regulatory capital arising from qualifying for the AIRB status increases the incentives for banks to improve these fundamental measures used in credit risk management.

Banks' internal ratings provide the PD indicators under the IRB approach. These ratings typically summarize information coming from internal financial analyses, vendor credit models, and agency ratings. Some of these sources focus on the current situation and others attempt to look at likely developments over several years. Basel's Third Consultative Paper (CP3), as well as other documents, refers to these contrasting approaches as Point-in-Time (PIT) and Through-the-Cycle (TTC).

The goal of this chapter is to:

- provide operational definitions of PIT and TTC ratings by focusing on the horizon involved in the credit assessment
- introduce an approach for translating agency ratings at different points in the cycle to one-year (PIT) PDs; and
- describe tests and initial empirical results measuring the accuracy of existing ratings as either PIT or TTC indicators of default risk.

By defining the relevant horizon for PIT or –TTC, we support a definition that can then be empirically tested as to the accuracy of a rating's default prediction. By converting agency ratings into PIT representations of one-year default rates, we provide an approach for integrating PIT ratings (such as one-year KMV EDFs) with agency ratings.[4] Finally, by focusing specifically on what makes a rating system PIT or TTC, we hope to provide a foundation for assessing and validating credit ratings.

[3] PD, EAD and LGD are the standard Basel II definitions for: probability of default, exposure at default and loss given default, respectively.

[4] Barclays utilizes what is called the Agency Read-Across Matrix as the master-scale in determining one-year default probabilities by internal ratings grades. The Matrix combines mappings of Agency Ratings to Barclays Business Grades (BBG) and median one-year default probabilities to BBGs. Successful conversion of through-the-cycle Agency Ratings to one-year point-in-time representations of default rates as outlined in this chapter, provides a means for comparing Agency Ratings with KMV EDFs to consistently derive BBGs within the Matrix.

II. A Brief Review of Internal Credit Ratings

Within the credit risk literature, a key focus over the last five to ten years has been on default prediction and the development of internal rating measures of borrower creditworthiness. Before turning to the issue of PIT versus TTC, we highlight some of the salient characteristics required of sound internal credit rating systems to provide some context for the PIT-TTC discussion to follow.

Banks use internal ratings as critical inputs in approving and pricing loans, establishing reserves, setting limits, and managing the portfolio of credit exposures. For banks to perform these functions well, internal credit ratings must discriminate accurately between borrowers with greater and lesser chances of defaulting over the varying time frames used in the analysis.

From an enterprise perspective, credit risk comes in many different colours, styles and shapes, and is traditionally managed in silos. For example, large banks have different types of obligors – retail customers, large corporate borrowers, SMEs, and sovereigns – requiring varying approaches for estimating creditworthiness. See for example Aguais and Rosen (2001), who outline an enterprise credit risk management framework that recognizes the need for different approaches across the banking enterprise.

Looking back on the evolution of internal ratings, earlier non-statistical approaches focused primarily on deriving ordinal rankings of risk, executed on a "yes-or-no" basis. Credit officers used qualitative factors and, later, financial data in assessing a borrower's willingness and ability to repay a credit obligation. As statistical and behavioural credit scoring models were being developed for consumer markets, fundamental credit analysis was becoming more quantitative for corporate borrowers.

Focusing on approaches applied to the corporate sector, the evolution from qualitative to quantitative (and then ordinal to cardinal) has followed a natural progression. Ed Altman's (1968) approach for the Z-Score stands out as one of the first statistical approaches. In the early 1990s, KMV Corporation provided one of the first commercially available models based on the Merton framework for predicting one-year PDs (also known as Expected Default Frequency or EDF). See Ranson, (2002) Chapter 3, for a review of some of the evolving approaches to risk rating and default measurement. [5]

With regard to Basel II today, there are some basic characteristics that broadly define a successful internal risk rating system. To start, ratings systems need to

[5] Also, see Treacy and Carey (1998) and Basel (2000) for a more detailed discussion of corporate risk rating systems.

distinguish different levels of credit risk with enough resolution (granularity) so that the bank avoids adverse selection in competing for customers. In addition, the ratings system must imply default risk measures that are cardinal (i.e. numeric) as in KMV EDFs and not just ordinal. Thus, for each separate internal risk rating, whether labelled BB+, 3-, or whatever, one must be able to identify an acceptably narrow range of values (e.g. 50 to 75 bps) for the corresponding PD over one or more standard horizons. The different ratings need to span the entire credit risk range consisting of from 0 to 100%. Rating systems that are derived from statistical default modelling typically satisfy these criteria.

Banks also use ratings in the pricing of illiquid, credit risky instruments. Thus, a bank may also need each ratings level to reference a benchmark credit spread or equivalently, a risk-neutral PD at one or more maturities.[6] Credit spreads, in principle, could involve ratings somewhat different from those related to PDs. Suppose that companies A and B have the same one-year PD, but that A's PD has a greater propensity to increase in recessions. The two companies would share the same one-year rating related to a real-world default risk, but A's one-year rating related to risk-neutral risk would be inferior to B's. In practice, however, analysts customarily use the same rating in gauging real-world and risk-neutral PDs. The state-of-the-art has not progressed far enough to permit differentiation between real-world and risk-neutral ratings.

With regard to the horizon for calibrating an internal ratings system, banks most often evaluate their ratings systems over one-year intervals. However, credit exposures often have maturities greater than one year and proper pricing and portfolio management usually involves analysis over longer horizons. This makes it important to consider the possibility of distinct PIT and TTC ratings.

III. Point-in-Time vs. Through-the-Cycle Ratings – An Overview

History of the Terminology:

In the January 2001 Consultative Document on the proposed IRB Approach for the New Basel Capital Accord, the Basel Committee on Banking Supervision (Basel, 2001) provides a formal distinction between PIT and TTC credit ratings. While it doesn't define the two terms explicitly, Basel evidently believes that there are PIT ratings that measure default risk over a short horizon of perhaps a year,

[6] Keeping the discussion simple, risk-neutral spreads are required in pricing to incorporate risk premiums, which compensate for uncertainty around expected credit losses.

and there are TTC ratings that measure it over a longer horizon of perhaps five or more years.

Specifically, Paragraph 53, page 12, of the 2001 Consultative Document for the Internal Ratings-Based Approach, says:

53. Some banks distinguish their rating system on the basis of whether it estimates the probability of a borrower's default on a "point in time" or "through the cycle" approach. In a "point in time" process, an internal rating reflects an assessment of the borrower's current condition and/or most likely future condition over the course of the chosen time horizon. As such, the internal rating changes as the borrower's condition changes over the course of the credit/business cycle. In contrast, a "through the cycle" process requires assessment of the borrower's riskiness bases on a worst-case, "bottom of the cycle scenario" (i.e., its condition under stress). In this case, a borrower's rating would tend to stay the same over the course of the credit/business cycle. [7]

While the January 2001 Basel Consultative Document discusses PIT and TTC approaches more broadly, the first Basel reference seems to be a year earlier in January 2000. In, the Basel Committee's Discussion Paper describing a G-10 survey of internal ratings systems (Basel 2000), the PIT-TTC distinction is first referenced under a discussion of risk rating time horizons. The comments made there are simple, but consistent with the January 2001 comments made above. In addition, the paper raises concerns expressed by some banks during the survey of potential inconsistencies created when mapping between external agency ratings (usually thought of as TTC) and internal PIT ratings.

So the general debate surrounding PIT versus TTC in the context of Basel II was initiated. But where actually do these terms for PIT and TTC actually come from? [8] Since agency ratings provided early examples of TTC ratings, it is natural that early references to "through-the-cycle" ratings first appeared in Moody's and S&P discussions of their approach to corporate ratings. One of the first examples of "through the cycle" ratings discussions can be found in a November 1995 Moody's report on the copper industry (Moody's, 1995). [9] In 1996, a discussion appearing in S&P's *Corporate Ratings Criteria* entitled, "Factoring Cyclicality Into Corporate Ratings", provides S&P's perspective on the TTC issue:

"Standard & Poor's credit ratings are meant to be forward-looking; that is, their time horizon extends as far as is analytically foreseeable. Accordingly, the anticipated ups and downs of business cycles – whether industry specific or

[7] See Basel 2001, Page 12.

[8] In writing this chapter, we have undertaken an initial search of the literature to find the origins of PIT-TTC and any related analysis. We have not, however, had the time to-date to conduct a complete review of the literature so the analysis is preliminary.

[9] Moody's (1995) page 7.

related to the general economy – should be factored into the credit rating all along. This approach is in keeping with Standard & Poor's belief that the value of its rating products is greatest when its ratings focus on the long-term, and do not fluctuate with near-term performance. Ratings should never be a mere snapshot of the present situation. There are two models for how cyclicality is incorporated in credit ratings. Sometimes, ratings are held constant throughout the cycle. Alternatively, the rating does vary – but within a relatively narrow band. . . . . . The ideal is to rate "through the cycle". " [10]

Following on the rating agencies discussions, the first use of the "point in time" terminology can be found in an analysis and survey of large US bank's risk rating systems conducted by two researchers from the Federal Reserve, William Treacy and Mark Carey (1998). [11] In this 1998 article, the first reference to point-in-time ratings seems to have been born in a juxtaposition of PIT and TTC ratings approaches. They write that, *"Rating the current condition . .[of the borrower] . .is consistent with the fact that rating criteria at banks do not seem to be updated to take account of the current phase of the business cycle. Banks we interviewed do vary somewhat in the time period they have in mind producing ratings, with about 25 percent rating the borrower's risk over a one-year period, 25 percent rating over a longer period, such as the life of the loan, and the remaining 50 percent having a specific period in mind. . . . In contrast to bank practice, both Moodys and S&P rate through the cycle."* [12]

Examining the Concepts of PIT-TTC

Having developed the history of the PIT-TTC terminology, we now examine these concepts in more detail below. We start by providing working definitions. We then identify conditions under which one can meaningfully distinguish between PIT and TTC ratings. Finally, we introduce the idea of testing the extent to which existing ratings can be regarded as good PIT or TTC indicators.

We start with a working definition of PIT and TTC ratings. Consider the following: A PIT rating measures default risk over a short horizon, often considered a year or less. A TTC rating measures it over a horizon long enough for business-cycle effects mostly to go away. As one convention, one could regard default risk over a period of five or more years as TTC.

With this definition, we can easily imagine a company's PIT and TTC ratings differing. Suppose we expect a company's creditworthiness to trend up or down

---

[10] See Page 65, Standard & Poor's (1996). This is the first example of S&P use of through the cycle terminology.
[11] See Treacy and Carey (1998).
[12] See Treacy and Carey (1998) page 899. In discussions with Mark Carey, to the best of his knowledge, he believes their 1998 article was the first to use the point-in-time terminology

atypically over several years. In this case, its PIT and TTC ratings would differ, with the disparity reflecting the anticipated but unusual evolution in the company's status.

Observe that we have not said that the pending developments reflect the general business cycle. A rating with a long horizon needs to account for all of the things that may occur over several years, not only the business cycle. In making decisions on multi-year exposures, a bank can't afford to ignore any of the events that seem likely or just possible over an extended period. Thus, we probably should think of TTC as denoting "long term" and PIT "short term."

While we can always imagine separate PIT and TTC ratings, the distinction may be unimportant in practice. We need to consider the following question:

When does a TTC rating provide information not already in a PIT one?

In answering this, it is easier to describe when a TTC rating provides *no* additional information. Suppose that the relationship between short- and long-term default risks is always the same – say a one-year PD of x invariably implies a 5-year PD of $f(x)$. Then, the TTC rating is redundant, since the PIT one already implies it. (Alternatively, we could regard the PIT rating as redundant, implied by the TTC one.) However, if the relationship between short- and long-term default risk varies across firms or time – meaning that a one-year PD of x implies a 5-year PD of $f(x) \pm z$, where z is an economically significant random variable – then the TTC rating could add information beyond that already provided by the PIT one. In the example, the TTC rating would add information if it helped predict z.

We could also express this by saying that the PIT-TTC distinction does not matter unless at least two factors influence credit risk. To grasp this point, consider the analogous situation for interest rates. In a single-factor interest rate model, the short rate provides all of the information needed for determining the entire yield curve. In other words, the short rate serves not only as a definitive indicator of where interest rates are and will be shortly (PIT), but also as a best predictor of where they might be over any extended time interval (TTC). If, however, it takes more than one factor to describe the predictable patterns in rates, one often selects a long-term interest rate (TTC) as a second factor that together with the short rate explains the entire yield curve.

For credit, this point becomes most transparent when applied to spreads. Imagine modelling credit spreads in the same way as interest rates and suppose that one needs values for two risk factors to establish a full term structure. In this case, one could think of the PIT rating as determining the one-year spread and the TTC rating as determining the 5-year one and those two spreads together as establishing the full term structure. However, if spreads derive from a one-factor model, then only one of the two ratings would suffice to pin down the entire term structure. If the term structure arises from a 3-factor model, one would need even more information than ratings at 2 different horizons.

So, the concept of distinct PIT and TTC ratings makes sense if one needs at least two risk factors to explain a typical company's credit risk term structure (see **Panel 1** for an example). We think a meticulous modelling of default term structures would involve more than one factor, so we continue to talk of PIT and TTC ratings. However, before rushing to develop a two-part rating system, a bank must ask whether, with the information now available, would it significantly improve the accuracy of its credit evaluations and the quality of its credit decisions to justify the effort? For the most part, we think not. Basel II looks ahead to having banks validate 1-factor ratings systems. For most banks, two-factor systems have not yet reached the drawing board. Some day we think they will.

[To Typesetter: Please insert Panel 1 somewhere here.]

We now consider the question whether one can classify any of the existing ratings systems as PIT or TTC. We often observe the KMV's one-year EDFs described as PIT and agency ratings as TTC, and wonder if this is a valid description. Those who believe so observe that (i) the one-year EDF has a PIT horizon; (ii) the rating agencies describe their long-term ratings as involving an analysis of conditions over several years; and (iii) the agency ratings exhibit considerably less volatility than KMV EDFs. However, one can counter with two observations. If, as is often assumed in pricing and portfolio modelling, credit risk reflects a series of independent shocks, then the long-term rating need not be less volatile than the short-term one. Additionally, the relatively low volatility of agency ratings could simply reflect errors – as would occur if the ratings are merely rank-ordered – or the relatively high volatility of KMV EDFs could reflect errors – as would occur if stock prices were "excessively volatile" compared with more fundamental factors revealed in credit events. Indeed, those who presume that a TTC rating must be less volatile than a PIT one evidently believe that credit risk exhibits mean reversion. However, we are unaware of any evidence that justifies this view.

Rather than trying to classify ratings systems *a priori* as PIT or TTC, we are instead seeking to answer the following question:

> *Which one or which combination of existing ratings and other indicators provides the best predictions of default events over a one-year horizon and which provides the best predictions over a 5-year horizon?*

A good PIT rating system could outperform a poor TTC one in trials over a 5-year horizon. In this case, we would accept the PIT system as a better TTC indicator. In the end, predictive power matters.

We now turn to two related matters. In the next section, we describe a process of developing dynamic mappings of agency ratings to one-year PDs. In the section after that, we describe some ongoing testing of the accuracy of KMV EDFs and agency ratings as default predictors over both one-year and 5-year horizons. These empirical tests consider the concepts of PIT and TTC as relating to the horizon over which the default rate prediction accuracy is best. Since our internal ratings derive in part from those two sources, we see this effort as integral to the process of establishing the validity of our internal ratings.

IV. Using a Single Credit Factor Framework in Mapping Agency Ratings to one-year PIT PDs

As overall credit conditions improve or deteriorate, agency ratings move up and down, but less so than default rates. We reach this conclusion after observing that the default rates, spreads, and median KMV EDFs for each agency rating exhibit cyclical variations. This implies that to estimate one-year default losses accurately for purposes of setting appropriate reserves, one must use dynamic mappings of agency ratings to default rates. We describe an approach for accomplishing that next.

A "static map" translates a particular rating always to the same PD at each relevant horizon as determined from historical average experience (see **Table 1**).

Moody's Mappings *		
Grade	1-year PD	5-year PD
Aaa	0.000%	0.11%
Aa	0.015%	0.33%
A1	0.036%	0.51%
A2	0.039%	0.74%
A3	0.040%	1.01%
Baa1	0.132%	1.88%
Baa2	0.150%	2.73%
Baa3	0.367%	4.89%
Ba1	0.642%	8.05%
Ba2	0.876%	11.61%
Ba3	2.550%	19.51%
B1	3.702%	25.06%
B2	8.391%	35.23%
B3	11.957%	44.43%
Caa	15.926%	50.48%
Ca	19.576%	57.36%

Table 1: Illustration of Static Mappings of Moody's Grades to PDs

* Derived from data in the most recent default rate study published by Moody's(2003). The above values reflect the smoothing that arises in fitting a transition matrix to the historical average default term structures for the different grades and then producing the mappings from that matrix.

A "dynamic map" translates ratings to PDs that vary with the credit cycle. In adjusting for the cycle, we start with the static map, which we regard as establishing the unconditional default rates for each grade. We then use a single-factor, CreditMetrics model for determining default rates conditional on the state of the credit cycle. [13] [14]

In applying this approach, we begin by estimating monthly series of latent factor values for each major (alpha, not alpha-numeric) agency grade. This involves comparing the monthly, median, KMV one-year EDFs for each major grade with the respective historical average values of those medians. The latent factor values arise from back solving the CreditMetrics model. For that purpose, we determine the default point for each grade by applying the inverse normal distribution function to the historical average median EDF for each grade. We estimate the correlation parameter for each grade simultaneous with the extraction of latent factors. We set the correlation parameter for each grade so that the extracted factor values have a mean near 0 and standard deviation close to 1.

We next average the latent factors across the major grades, thereby obtaining a single, summary series measuring the credit cycle. We now determine the default points for all of an agency's (alpha-numeric) grades by applying the inverse normal function to the long-run PDs in the static map. By substituting the series of single factor values into the CreditMetrics model, with the correlation parameter set to the average estimated in the factor extractions above, we obtain monthly PDs for each grade adjusted for current credit conditions. (See **Panel 2** for a more detailed description of this dynamic mapping approach.) This approach can produce quite different PD translations over time (see **Figure 1**).

[To Typesetter: Please insert Panel 2 somewhere here.]

---

[13] See Belkin, Suchower and Forest (1998 A and B) for a discussion of one-factor credit models in the context of the CreditMetrics framework.

[14] Barclays Capital is also developing new LGD models that incorporate the kind of cyclical credit factor described here, making these also point in time. Developing more complex LGD models with these types of cyclical credit factors, however, implies a more complex PD model as well – because there are partially offsetting impacts between LGD and PD due to variable default points in the PIT LGD model.

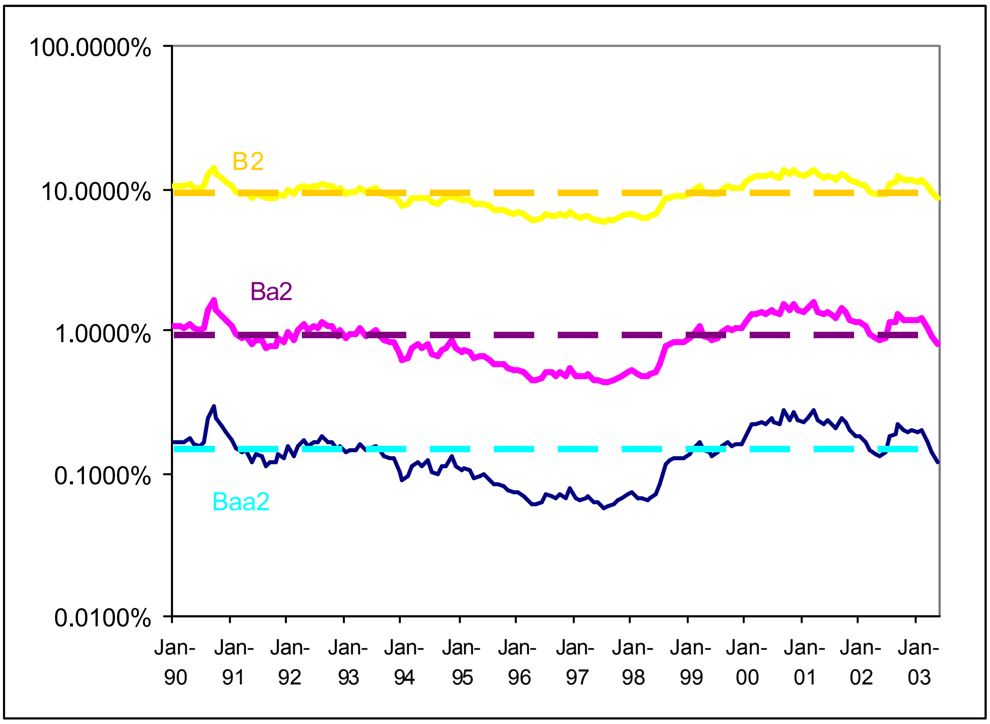


Figure 1: Illustration of Dynamic (solid) and Static (dashed) one-year EDF Mappings for Selected Moody's Grades

Source: Developed from Moody's KMV data on ratings and EDFs.

In using this dynamic mapping approach, we take the one or more agency grades for a company and translate them to PDs using the current mapping. In effect, we retain the rank ordering of default risk determined by agency ratings, but attach a default rate calibration drawn from using KMV data. Thus, the resulting 1-year PDs represent a hybrid of agency and KMV information.

V. A Framework for Assessing and Validating Credit Ratings

Under the proposed Basel II, banks need to ensure that their credit ratings accurately predict PDs. To support these efforts, there is a growing literature on credit model validation.

Several analysts have worked on evaluating credit ratings focusing on ways of gauging goodness of fit both in- and out-of-sample for discrete variables such as the default/no-default realizations. The rating agencies and credit scoring firms have long presented Power Curves - plots of (percent bad, percent good) points at different exclusion thresholds -- in illustrating the performance of their indicators in discriminating between good and bad outcomes. Alternatively, van

Deventer and Wang (2004), a chapter in this book, examine the properties of ROC curves and related metrics for judging such predictive performance.

Papers by Stein (2003) and Kurbat and Korablev (2002) look at skewness in small default samples and its effect on tests of the accuracy of default models and ratings. Friedman and Sandow (2003) present a utility based rationale for using likelihood ratios in judging predictive performance. Most work in this area has focused almost exclusively on goodness of fit, paying little attention to the conceptual soundness of the underlying model. However, KMV in many places has presented an extensive theoretical rationale for its empirical approach. [15]

We view the process of validating credit ratings and the other key Basel II components (LGD and EAD) of credit risk quantification as involving both a theoretical and empirical component. We see a model, ratings or otherwise, as "valid" only if one can give affirmative answers to each of the following questions:

- Is the model conceptually sound, so that, under conditions of ample data for estimation, one would expect it to provide reliable predictions?

- Is the model as currently implemented demonstrably reliable, as indicated by statistical tests of significance and precision?

These two conditions mirror academic and industry research journals acceptance criteria for recognition of research results.

To make a case for statistical significance and precision, one typically must:

- show that all of the model coefficient estimates fall within the range of plausible values and achieve significance as indicated by T-statistics or analogous measures;

- give evidence that the model is robust by obtaining close to the same parameter values in estimating over varying samples or time periods; and,

- demonstrate by means of both in- and out-of-sample (including "out-of-time"[16]) simulations that the estimated model produces results superior to an applicable naïve model such as one that forecasts values as remaining constant or falling on a trend line.

In cases of limited data, statistical tests may have little power. Thus, with new models or old ones newly applied in areas with scarce data for estimation and

[15] For example, see Crosbie and Bohn (2002).
[16] "Out of sample" denotes an observation not used in the model estimation. "Out of time" indicates that an out-of-sample observation occurred within a time period not covered by the observations used in estimation.

testing, the first criterion often weighs heavily in evaluating validity. Observe that this condition provides some assurance that a model will perform well in forecasting outside the range of past experience. With enough data, however, one needs to verify this with "out of sample" and "out of time" simulations.

We are currently refining the documentation of new and existing models as part of the validation process. We now turn to some empirical tests being conducted on the *components* that go into our internal ratings. Presumably if the components pass muster and we can show that the process of using those components in establishing a single best rating is sound, our internal ratings will also pass muster as well.

Whether a ratings system is accurate enough to be considered good enough or "valid" for some purpose such as under Basel II, usually involves human judgment. We can compare the accuracy of different ratings systems reasonably and objectively, but often cannot say with strict objectivity what level of accuracy corresponds to validity. Thus, we use the following criteria in this chapter:

> *In several places in its consultative papers, Basel endorses the use of agency ratings.* [17] *Thus, if we find that another rating system outperforms agency ratings in predicting defaults, we will take that as indicating that the other rating system is valid for use in implementing the IRB approach.*

In following up on this, we are currently conducting empirical testing designed to address the following question:

Which of the following best predicts defaults over a horizon of 1 year and over a horizon of 5 years?

- *KMV 1-year or 5-year EDFs*
- *KMV 1-year or 5-year EDFs grouped and given group-average EDFs*
- *S&P ratings mapped statically to 1-year and 5-year EDFs*
- *Moody's ratings mapped statically to 1-year and 5-yea) EDFs*
- *S&P ratings mapped dynamically to 1-year and 5-year EDFs*
- *Moody's ratings mapped dynamically to 1-year and 5-year EDFs*
- *Weighted averages of the above.*

These tests will help us determine the best credit ratings over 1- and 5-year horizons. This will also help us evaluate the advantage from having separate PIT and TTC ratings.

We are measuring accuracy in each case by using the log-likelihood of the default/no-default outcomes, conditional on the particular rating system or

[17] Basel II's term for accepted agency ratings is "external credit assessment institution" (ECAI). The criteria for acceptance, includes, objectivity, independence, international access/transparency, disclosure, resources and credibility.

combination of systems used in gauging default risk. We are also assessing rank-ordering performance using Lorenz curves and the related metrics.

We compute the log-likelihood (*LL*) across companies and time for the *j*th rating system as below:

$$LL_j = ln\left(\prod_i (1-PD_{ij})^{1-D_{it}} PD_{ij}^{D_i}\right)$$
$$= \sum_i D_i ln(PD_{ij}) + (1-D_i) ln(1-PD_{ij})$$

$$j \equiv index\ of\ the\ rating\ system \qquad (1)$$
$$ln \equiv natural\ logarithm$$
$$i \equiv index\ of\ the\ observation$$
$$PD_{ij} \equiv PD\ predicted\ by\ the\ rating\ system\ j\ for\ the\ observation\ i$$
$$D_i \equiv default\ indicator\ for\ the\ ith\ observation\ (1 = default,\ 0 = no\ default)$$

Friedman and Sandov (2003) provide a theoretical rationale based on utility functions for using the log-likelihood measures in choosing among competing models. They show that rank-ordering measures, such as the Gini coefficient, have no rationale justification. A ratings system could rank order perfectly, yet bankrupt a lender due to poor calibration. The likelihood measures account for both rank-ordering and calibration accuracy. Nonetheless, we will also examine the pure, rank-ordering performance of the different ratings.

Preliminary Assessment and Validation Results for PIT and TTC Ratings

Our initial tests support the previous findings (Crosbie and Bohn, 2002) that KMV 1-year EDFs predict default at a 1-year horizon better than agency ratings (see Tables 2 through 5 and Figure 2 through 5). This result comes out most strongly in the use of regression analysis in finding the combined measure that explains best (as gauged by maximum likelihood) the observed default and no-default outcomes. The best blended indicator drawing on KMV and Moody's data assigns weights of 84 per cent to the KMV EDF, 16 per cent to the dynamically mapped Moody's rating, and 0 to the statically mapped agency rating. Conducting the same analysis with KMV and S&P data, we get close to the same relative weighting.

On a stand-alone basis, we also see that the KMV EDFs best explain the observed sample (likelihood value of -1007 compared with -1100 for dynamically mapped Moody's and -1120 for statically mapped Moody's and based on a more limited sample we see similar results comparing KMV with S&P). Even if we eliminate the greater resolution of KMV EDFs by placing them in bins that parallel the Moody's or S&P ratings grades and assigning to each the average PD of the bin, we find that the KMV indicators predict better at one year. By examining the power curves, we see that KMV's rank-ordering performance at one year exceeds that of Moody's and S&P. The KMV 1-year power curve lies mostly above those of Moody's and S&P.

At a 5-year horizon, the Moody's and S&P ratings perform relatively better, supporting the view that they embody a multi-year analysis. On a standalone basis, the Moody's ratings mapped (statically) to unconditional 5-year PDs perform best, with the dynamically mapped ratings second and KMV's 5-year EDF last. We get similar results in comparing KMV with S&P over a five-year horizon. In the regression analysis, we find that the best composite indicator places puts approximately equal weight on the KMV 5-year EDF and the statically mapped Moody's rating. Combining KMV and S&P data, we also find that the best blended indicator puts about equal weights on the KMV 5-year EDF and the statically mapped S&P ratings.

These preliminary results offer evidence that credit analysis for short- and longer-term exposures need to differ. Thus, banks may need to look ahead to developing separate short and long-term ratings, with KMV EDFs when available playing a dominant role in the former, and agency ratings and other longer horizon indicators having a substantial role in the latter.

Table 2: Likelihood Values for Moody's and KMV Default Predictors

Horizon	Intervals	Moody's		KMV	
		Static Mapping	Dynamic Mapping	Binned	Full Resolution
1 year	Jan 90 to Jan 91 to Jan 02 to Jan 03	-1120	-1100	-1013	-1007
5 years	Jan 90 to Jan 95 to Jan 98 to Jan 03	-1562	-1582	-1651	-1645

Source: Moodys|KMV for Moodys ratings and defaults and Credit Monitor EDFs

a. "Static" denotes a Moody's rating mapped to a PD using a historical average default rate.
b. "Dynamic" denotes a Moody's Rating mapped using cyclically-varying default rate using the procedure explained in Panel 2.

Table 3: Likelihood Values for S&P and KMV Default Predictors

Horizon	Intervals	S&P		KMV	
		Static Mapping	Dynamic Mapping	Binned	Full Resolution
1 year	Jan 90 to Jan 91 to Jan 02 to Jan 03	-963	-935	-884	-831
5 years	Jan 90 to Jan 95 to Jan 98 to Jan 03	-1141	-1161	-1253	-1251

Source: S&P for S&P ratings and defaults and Moody's|KMV for Credit Monitor EDFs

Table 4: Maximum Likelihood Estimates of Optimal Weights on Moody's and KMV Indicators for Default Prediction

Horizon	Moody's		KMV
	Static Mapping	Dynamic Mapping	Full Resolution
1 year	.00	.16	.84
5 years	.46	.00	.54

Source: Moodys|KMV for Moodys ratings and defaults and Credit Monitor EDFs

Table 5: Maximum Likelihood Estimates of Optimal Weights on S&P and KMV Indicators for Default Prediction

Horizon	S&P		KMV
	Static Mapping	Dynamic Mapping	Full Resolution
1 year	.00	.15	.85
5 years	.54	.00	.46

Source: Moodys|KMV for Moodys ratings and defaults and Credit Monitor EDFs

Figure 3: Moody's and KMV One-Year Power Curves

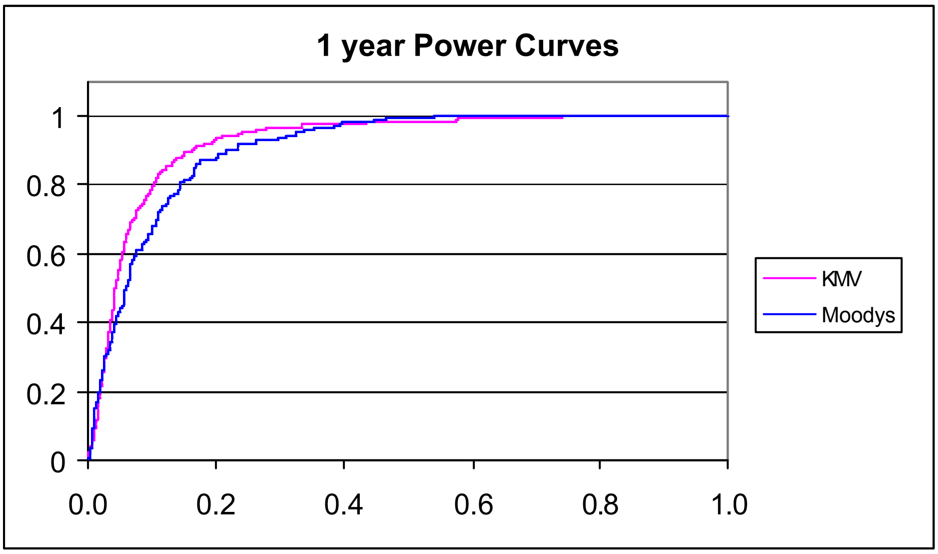


Figure 3: Moody's and KMV Five-Year Power Curves

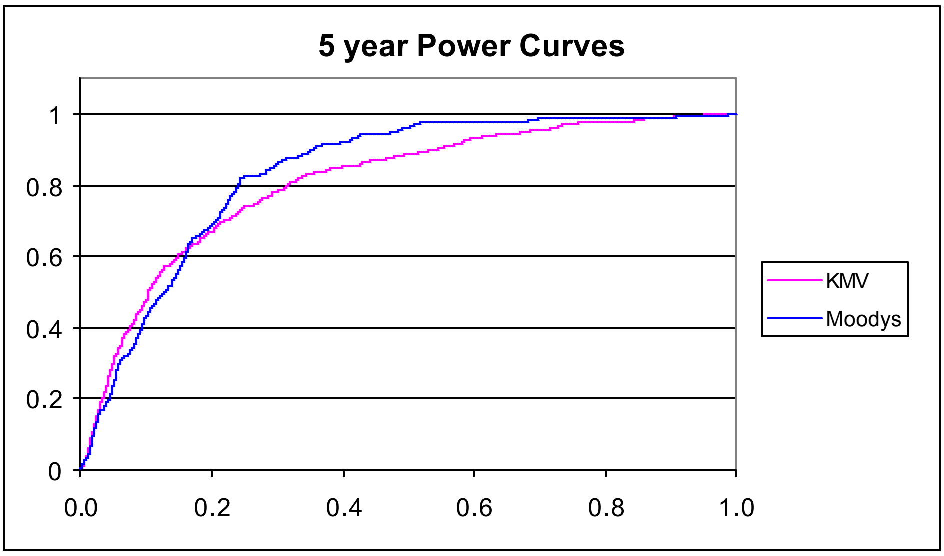


Figure 4: S&P and KMV One-Year Power Curves

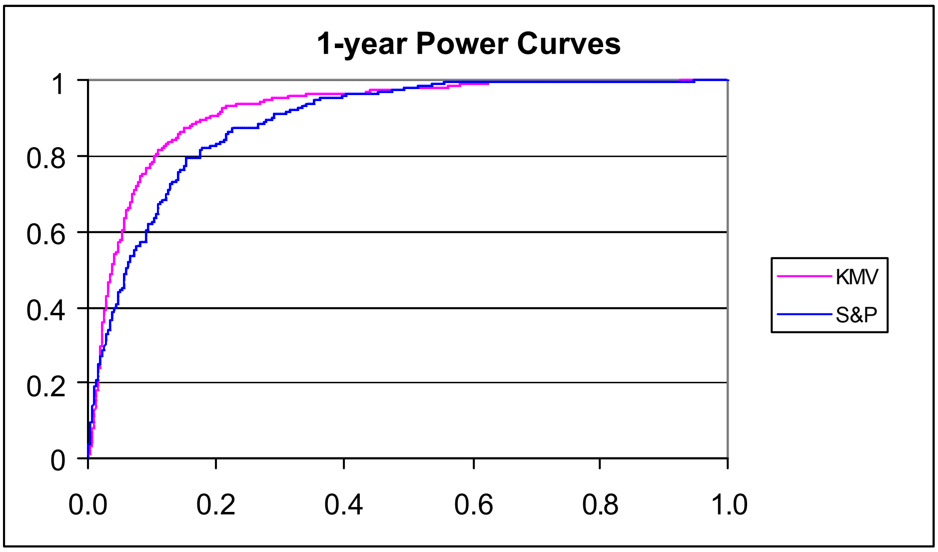
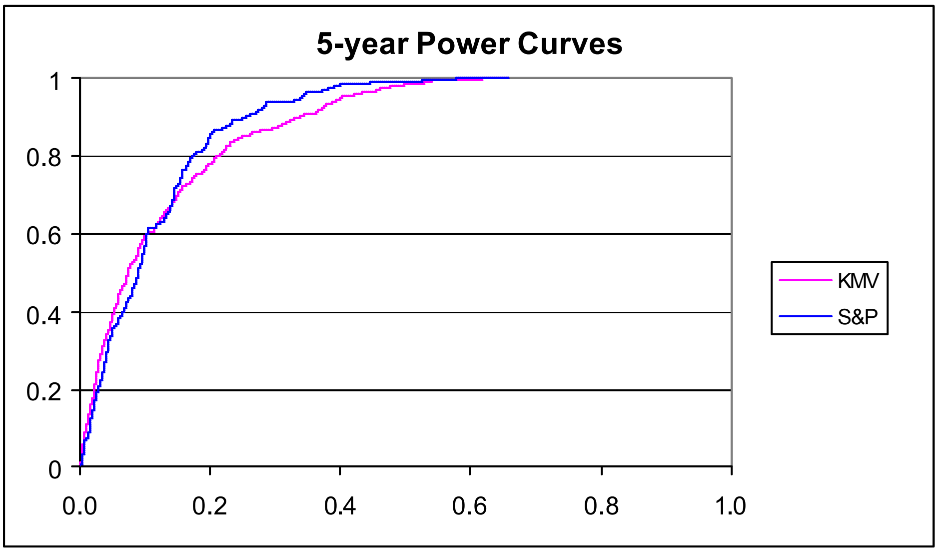


Figure 5: S&P and KMV Five-Year Power Curves



VI. Concluding Comments:

Accurate credit ratings stand out as central to Basel II and to most aspects of managing credit risk. In recent years, analysts have introduced the concepts of PIT and TTC ratings, concepts that purportedly distinguish between some of the existing ratings systems and models.

To make these ideas operational, this chapter defines a PIT rating as one that discriminates among borrowers based on the PD over a 1-year horizon and a TTC rating as one that focuses on a longer horizon of 5 years or more. We observe that the PIT-TTC distinction matters if, as in interest rate modelling, one needs at least two factors to characterize the different term structures of credit risk associated with different borrowers.

This chapter introduces a method for translating agency ratings dynamically to 1-year PDs. The approach applies the CreditMetrics model to median EDFs for each agency grade in estimating latent factors that describe at a point in time whether 1-year PDs are relatively high or low for each agency rating. The apparent TTC focus of agency ratings motivates this dynamic mapping approach.

We also describe statistical tests designed to evaluate the quality of different ratings systems as 1-year (PIT) and 5-year (TTC) default predictors. Initial results corroborate earlier findings that KMV EDFs outperform agency ratings at a 1-year horizon. However, at 5 years, the agency ratings perform as well as KMV 5-year EDFs. This supports the agencies' contentions that their ratings should be viewed as TTC indicators.

Much more analysis lies ahead before one can consider such results as well established. As they stand, however, they suggest that banks need to look ahead to developing separate ratings for short- and longer-term exposures.

References

Aguais, S. D. and D. Rosen, 2001, *Enterprise Credit Risk Using Mark-to-Future*, Algorithmics Publications, pp. 3-15, September.

Altman, E., 1968, "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy", *Journal of Finance*, volume 23.

Basel Committee on Banking Supervision, 2001, The New Basel Capital Accord, Consultative Document, The Internal ratings-Based Approach, Basel Committee on Banking Supervision, January.

Basel Committee on Banking Supervision,1999, Principles for the Management of Credit Risk, Consultative Document, Basel Committee on Banking Supervision, July.

Basel Committee on Banking Supervision, 2000, Range of Practice in Banks' Internal Ratings Systems, Discussion Paper, Basel Committee on Banking Supervision, January.

Belkin, B., S. Suchower and L. Forest, 1998, "The effect of systematic credit risk on loan portfolios and loan pricing", *Credit-Metrics Monitor*, pp.17-28, April.

Belkin, B., S. Suchower and L. Forest, 1998, "A one parameter representation of credit risk and transition matrices", *Credit-Metrics Monitor*, pp.45-56, October.

Crosbie, P., and J. Bohn, 2002, "Modeling Default Risk", Moodys|KMV Working Paper, January.

Friedman, C. and S. Sandow, 2003, "*Model Performance Measures for Expected Utility Maximising Investors*", Standard & Poor's, Risk Solutions Group, Working Paper, January.

Kurbat, M. and I. Korablev, 2002 "*Methodology for Testing the Level of the EDF Credit Measure", Moodys|KMV, Working Paper*, August.

Moody's Investors Services, 1995, Copper Perspective, Moody's Investors Services, Global Credit Research, Special Comment, November.

Moody's Investors Services, 2003, Default & Recovery Rates of Corporate Bond Issuers – A statistical review of Moody's Ratings Performance, 1920-2002, Moody's Investors Services, Special Comment, February.

Ranson, B., 2003, Credit Risk Management, Thomson – Sheshunoff.

Standard & Poor's, 1996, Corporate Ratings Criteria, Standard & Poors Rating Services.

Standard & Poor's, 2003, Ratings Performance 2002, Default, Transition, Recovery and Spreads, Standard & Poor's, February.

Stein, R., 2003, "*Are the Probabilities Right* ?", Moodys|KMV, Technical Report No. 030124, Revised, May.

Treacy, W. and M. Carey, 1998, "*Credit Risk Rating at Large U.S. Banks*", Federal Reserve Bulletin, xx(cc):897-921, November.

Van Deventer, D. and J. Outram, 2002, "*The New Capital Accord and Internal Bank Ratings*", Kamakura Corporation Monograph, May.

[To Typesetter: Please make the following into Panel 1 with the same title. The panel must be inserted in the vicinity where I had marked earlier on the main text.]

Appendix A: Examples of Single- and Two-Factor Credit Models

Consider some of the following credit models for ratings. We start with the single-factor case:

Let $j \in \{1, 2, \ldots, D\}$ represent a *PIT* rating in which *1* indicates the best credit state and *D* the worst, default. Let j_t be the rating value at time *t*. Assume, further, that if we know j_t, then we can determine the rating's *probability distribution* at any future time *t+n* by applying the fixed transition matrix *T* as follows,

$$D(j_{t+n}|j_t) = T^n u(j_t)$$

$$\begin{aligned}
&j_{t+n} \equiv \textit{credit state at time } t+n \\
&D(j_{t+n} \mid j_t) \equiv (p(1_{t+n} \mid j_t), \ldots, p(D_{t+n} \mid j_t))' \\
&p(j_{t+n}|j_t) \equiv \textit{probability of reaching } j_{t+n} \textit{ at time } t+n \textit{ starting from } j_t \textit{ at time } t \\
&u(j_t) \equiv (0,\ldots,0,\underset{\substack{\uparrow \\ j_t^{th}\,slot}}{1},0,\ldots 0)'
\end{aligned} \tag{1}$$

As shown in Eqn (1), if we know j_t, we need no further information to evaluate creditworthiness at any future time *t+n*. Given the assumption of a fixed transition matrix at all times, the relevant distribution of future creditworthiness $D(j_{t+n}|j_t)$ in the case of each possible rating is predicted entirely from knowledge of the current PIT rating j_t. No *TTC* rating can add any useful information. The single PIT rating j_t tells us all we can know about a borrower's future credit status, so we may regard it also as the best *TTC* rating.

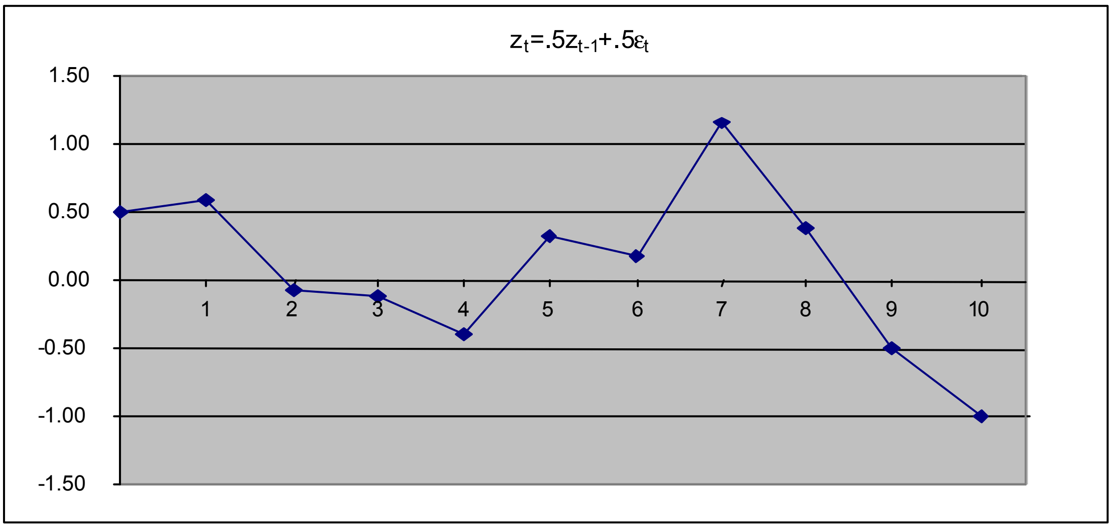
Next consider a two-factor model of credit risk. Suppose that we can explain random changes over time in the transition matrix with a factor *z*. Since the z affects transition rates for all borrowers, it is a systematic factor related to general economic conditions. Thus, $T(z_t)$ represents the transition matrix applicable at time *t* as influenced by the value z_t of the factor at that time. Assume further that the *z* factor evolves according to the mean reverting process depicted below

$$z_{t+1} = bz_t + \sqrt{1-b^2}\,e_t$$

$$\begin{aligned} & 0 < b < 1 \\ & e_t \sim \Phi(0,1) \\ & F(m,s) \equiv normal\ cdf\ with\ mean\ m\ and\ standard\ deviation\ d \end{aligned} \tag{2}$$

Such a process yields time series that can exhibit cycles (see **Figure 6**)

Figure 6: Cyclical Pattern in One Realization of a Mean Reverting Process

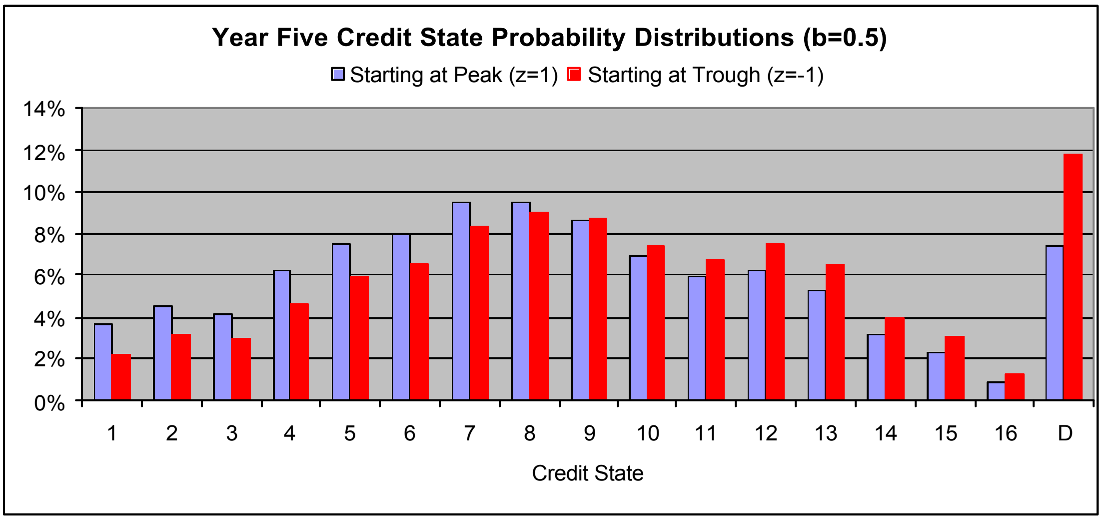


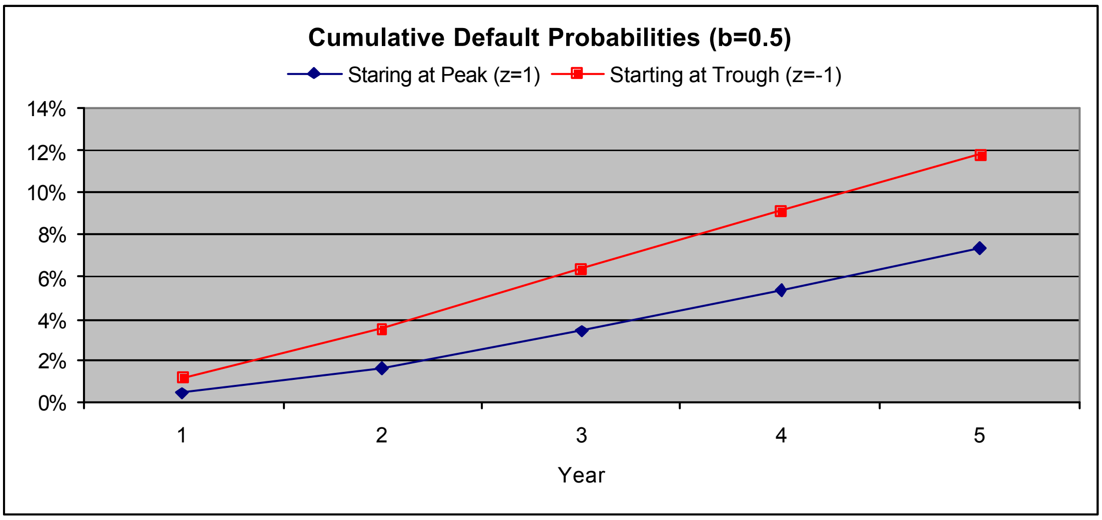
Here, we obtain the probability distributions $D(j_{t+n}|j_t, z_t)$ that describe credit risk at each future time $t + n$ by the integration below

$$D(j_{t+n} \mid j_t, z_t) = \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{n} T(z_{t+i}) f(z_{t+1}, \ldots, z_{t+n} \mid z_t) dz_{t+1} \ldots dz_{t+n} \right] u(j_t) \tag{3}$$

We use Eqn (2) in determining the density f. The formula (3) indicates that the credit state distributions depend jointly on the initial credit state j_t and the starting value of the systematic factor z_t. We observe, however, that if $b = 0$ in Eqn (2), then z_t offers no information useful in predicting z_{t+i} for $i = 1, 2$ and so on, and this case reverts to the single factor model. However, if the z factor displays "persistence" (meaning $b>0$), then we obtain different distributions of future credit states for different initial values of z_t (see **Figure 2**).

Figure 7: Effect of Persistent z Factor on the Credit Risk of a Borrower with Credit-State-10 (BB) Unconditional Transition Probabilities[18]





Source: Calculations based on assumptions described in this panel.

Suppose in the above example we set the PIT rating to the credit state with an unconditional, first-year EDF that most closely matches the first-year EDF of the borrower. If overall credit conditions were strong in our model as indicated by *z=1*, the first-year EDF of the borrower would be about 50 bps, corresponding to a PIT rating of 5 (approximately *A-* in the S&P rating scale). If, on the other hand,

[18] Results derived using Monte Carlo simulations of the *z* process and the CreditMetrics, single-factor, parametric representation of a conditional transition matrix, assuming a value of .1 for the obligor's correlation with general credit conditions. Th e unconditional matrix in the analysis contains estimated, KMV EDF-transitions across 17 EDF bins including default

credit conditions were weak as indicated by $z=-1$, the first-year EDF would be about 120 bps, corresponding to a PIT rating of 10 (*approximately BB/BB-*).

Alternatively, suppose we use the cumulative 5-year EDF in determining a TTC rating. Looking ahead from a good credit environment ($z=1$), we would foresee a 5-year EDF of 740 bps, corresponding to a TTC rating of 9 (*approximately BB+*). Looking forward from a difficult environment ($z=-1$), we would anticipate a 5-year cumulative EDF of 1178 bps, corresponding a TTC rating of 11 (*approximately BB-*).

To see the point of this illustration, consider performing the same mappings in the case of the single factor model in which the transition matrix stays fixed at T. Here, regardless of the state of the economy, we always would map to the borrower's initial credit state (10 is approximately *BB* in the illustration) at every horizon.

This example motivates two conclusions. Given a two-factor credit risk, we need two indicators such as a separate short- (PIT) and long-term (TTC) rating in estimating the credit-risk term structure. Further, if the factor describing credit trends exhibits mean-reversion, then the TTC rating would likely vary less than the PIT one.

In closing this panel, we note that in the case of a two-factor credit risk, we would likely need complementary two-factor models describing the risk-neutral and the real-world processes. Analysts sometimes avoid the complexity of estimating both processes simultaneously, working instead only with the one needed for the purpose at hand. This seems like a dubious practice, since at times the best regarded real-world and risk-neutral analyses imply something implausible about the market's relative weighting of returns in recessions and booms.

[End of Panel 1]

[To Typesetter: Please make the following into Panel 2 with the same title. The panel must be inserted in the vicinity where I had marked earlier on the main text.]

Appendix B: Description of the Approach for Updating Agency Mappings to One-Year PDs

To update the mappings of agency ratings to one-year PDs, we use the CreditMetrics single-factor modelling approach intrinsic to the Basel II proposals and widely adopted by many credit portfolio models. We assume, in effect, that today's PIT default probability for each of an agency's ratings arises from the current realization of a single risk factor measuring the credit cycle. By using only one factor, we impose a uniform, business cycle view across each of an agency's ratings. In this way, we avoid sampling errors that could arise if we were to attempt a more detailed, multi-factor approach based on our limited data.

The approach involves the following steps:

1. Derive the value of a latent risk factor for each of an agency's major risk grades by comparing the current, median, KMV one-year EDF for that grade with its long-run historical average. Specifically, we calibrate the CreditMetrics model for a grade to the long-term average EDF, using a correlation factor that yields factor realizations with a mean close to 0 and a standard deviation near 1. For each month, we back-solve the model for the factor value that generates the current, median KMV EDF for that grade. We use the following formula:

$$Z_{gt} = \frac{F^{-1}(EDF_g) - \sqrt{1-?}\,F^{-1}(EDF_{gt})}{\sqrt{?}}$$

$Z_{gt} \equiv$ *factor value for grade g in period t* $\quad F \equiv$ *standard normal cdf*

$? \equiv$ *correlation coefficient* $\quad EDF_g \equiv$ *long-run average EDF for grade g*

$EDF_{gt} \equiv$ *EDF for grade g in period t*

   For both S&P and Moody's, we derive factor values for each major grade except CCC (Caa-C). In these cases, KMV's capping of EDFs at 20% biases the estimates and make them invalid for this exercise.

2. Obtain a single, monthly credit-cycle series for the agency by averaging the distinct, monthly latent factor values for each of the agency's major grades. We use the following formula:

$$Z_t = \frac{\sum_{g=1}^{N} Z_{gt}}{N}$$

$$Z_t \equiv overall\ average\ credit\text{-}cycle\ factor\ value\ for\ the\ Agency$$
$$N \equiv number\ of\ major\ Agency\ grades\ used\ in\ the\ analysis$$

3. Apply the monthly values of this credit cycle index within the CreditMetrics single-factor model, thereby obtaining *PIT* EDFs consistent with this view for each of the agency's ratings. We use the following formula:

$$PD_{gt} = \Phi\left(\frac{F^{-1}(PD_g) - \sqrt{r}\,Z_t}{\sqrt{1-?}}\right)$$

$PD_{gt} \equiv$ *estimated PD for grade g in period t* $\quad F \equiv$ *standard normal cdf*

$? \equiv$ *correlation coefficient* $\quad PD_g \equiv$ *long-run average PD for grade g*

$Z_t \equiv$ *average factor value for period t*

We use PDs here rather than EDFs to denote one-year default rates. This indicates that we are using historical averages and deriving *PIT* values consistent with an agency's experience. The long-run average PDs for many agency grades differ substantially from the long-run average EDFs for those same grades. Thus, we are implicitly tying the agency mappings to actual agency experience.

We summarize this approach in **Figure 8**.

Figure 8: Using a Single-Factor to Determine *PIT* EDFs