Biased Benchmarks

Lawrence R. Forest Jr
Senior Consultant to PricewaterhouseCoopers
2080 Mackinnon Avenue, Cardiff-by-the-Sea, CA, USA
Email: lforestjr@aol.com

Gaurav Chawla (corresponding author)
Risk Rating Modelling Leader, GE Capital
201 Talgarth Road, Hammersmith, London, W6 8BJ, UK
Email: Gaurav.Chawla@gatech.edu

Scott D. Aguais
Managing Director, Aguais & Associates Ltd.
20-22 Wenlock Road, London, N1 7GU, UK
Email: saguais@aguaisandassociates.co.uk

The views presented in this article are those of the authors and have not been endorsed by their past or current employers.

Abstract: Regulators and credit analysts have used long run average, default rates (DRs) from the S&P and Moody's default studies and EDFs from the MKMV Public Firm Model as benchmarks for evaluating the accuracy of an institution's PD models. But recent evidence indicates that these benchmarks have over the last 11 years, been exaggerating default risk for non-financial, corporate entities (Corps). For Corps, over the cyclically neutral period from the start of 2003 through 2013, the average one year, realised DRs of almost every S&P or Moody's, alpha-numeric grade is well below the average DRs experienced before 2003. Expressed in terms of grades, it appears that both S&P and Moody's over the past 11 years have been grading Corps more harshly than earlier by about one alpha-numeric notch in the speculative-grade range and by about two in the investment-grade range. For financial institutions (FIs), recent over-estimation of default risk occurs only in the sub-investment grades. Reflecting catastrophic failures of some highly rated institutions during 2008-09, the DRs in the low-risk grades equivalent to S&P A+ or better have been moderately higher than before 2003. We find patterns similar to these with Moody's KMV (MKMV) EDFs, except that for FIs the over-estimation is more pervasive than with S&P and Moody's grades. The sources of this time inconsistency bias remains unclear. It could be due to unidentified improvements in risk management (especially in Corps) or due to the growing asymmetry in the attitudes of regulators and others toward under- and over-estimation of risk. The evidence presented here raises concerns that lending institutions applying these benchmarks may be unduly restricting corporate lending.

Keywords: Agency Ratings, Probability of Default (PD), Point-in-Time (PIT), Through-the-cycle (TTC), MKMV EDFs, Benchmarking, Hypothetical Portfolio Exercise (HPE), credit cycle index, credit policy, risk weighted assets (RWA)

Overview

Under Basel II Advanced Internal Rating Based (AIRB) approach, banks have the option of using internal Probability of Default (PD) models in determining risk weighted assets (RWA) in a manner which is more risk sensitive than the simpler, standardized approach. Within the Basel II framework, to be approved for use in determining RWA, advanced PD models must pass muster from both internal bank reviewers and the regulators. These two levels of review have inspired greater rigor in model development and closer adherence to regulatory guidelines. Under this more rigorous development and review process, newly approved AIRB PD models have typically been calibrated to internal credit data involving potentially different conventions for defining default, exposure, and loss and varying margins of conservatism. Such variation in modelling choices leads to variations in model PDs for same obligors and transaction risks, when comparing model output of different banks.

It is this concern over the general validity of models developed using limited, internal credit data under Basel II that has motivated increased usage of benchmarks. One sees this, for example, in the UK, Prudential Regulatory Authority's (PRA's), recurring, Hypothetical Portfolio Exercise ('HPE'). In the HPE, the PRA compares each bank's credit risk parameters with medians from all reporting banks. Further, based on a selection of S&P rated entries, the PRA compares each bank's median PDs for each alpha-numeric grade with the 1981-to-date, long-run-average default rate (DR) for each grade.

This growing application of benchmarks involves the potential danger that the benchmarks themselves may also be inaccurate. In particular, reconciliation with the medians from other banks produces consensus and not necessarily the most accurate representation of risk. Further, the grades from S&P and other major Ratings Agencies involve bespoke, highly judgmental methods. Consequently, one does not have the discipline of quantitative, default models to enforce consistency over time and across asset classes. Moreover, due to the bespoke, judgmental nature of those methods, the Rating Agencies can't restate past grades to reflect current methods that accumulate improvements. This raises a concern with 'time inconsistency' and this paper finds evidence of that.

This paper focuses primarily on assessing the 'time inconsistency' of S&P and Moody's grades for non-financial corporate (Corps) and for financial institutions (FIs) using an agency rating based default model, which is based on Point-in-Time (PIT) and Through-the-Cycle (TTC) dual ratings approach developed and presented in Aguais et al, 2004, 2007; Forest et al, 2013; Chawla et al 2013. This PIT-TTC framework supports a more detailed analysis of cross-time variations in key credit benchmarks because 'time inconsistency' of agency grades can be proved after controlling for systematic credit conditions using the PIT-TTC framework's credit cycle indices (CCI). We compare the relationship between DRs and grades over 2003-13 with that evident in earlier years. We find evidence of temporal shifts that are statistically significant and cause the long-run-average, DRs per grade to exaggerate default risk in recent years. We present these findings in Sections 1 and 2, with more details of the agency rating based default model in Appendix A. Important technical terms and acronyms used in this document is presented in Appendix B.

We also present evidence developed by MKMV that indicates that the MKMV Public Firm EDFs exhibit an upward bias in recent years. This finding arose from an unsuccessful effort to use the EDFs as evidence that the S&P and Moody's long-run-average benchmarks were biased. MKMV, however, is currently rolling out a new, Version 9, Public Firm model that reportedly reduces or eliminates the upward bias in its Version 8 model.

1. Downward Shift in Corporate Default Risk within Most Grades

For Corps over the years 2003-13, the average one-year, realised DRs of almost every alpha-numeric, S&P or Moody's grade sits well below the average experienced in earlier years (**Figure 1** and **Figure 2**). Aside from the extremely low-risk grades that have, since 1980, experienced no defaults within a subsequent calendar year, we see only one exception (out of 14 grades) to this pattern -- the highest risk grade (CCC/C, Caa2/C).

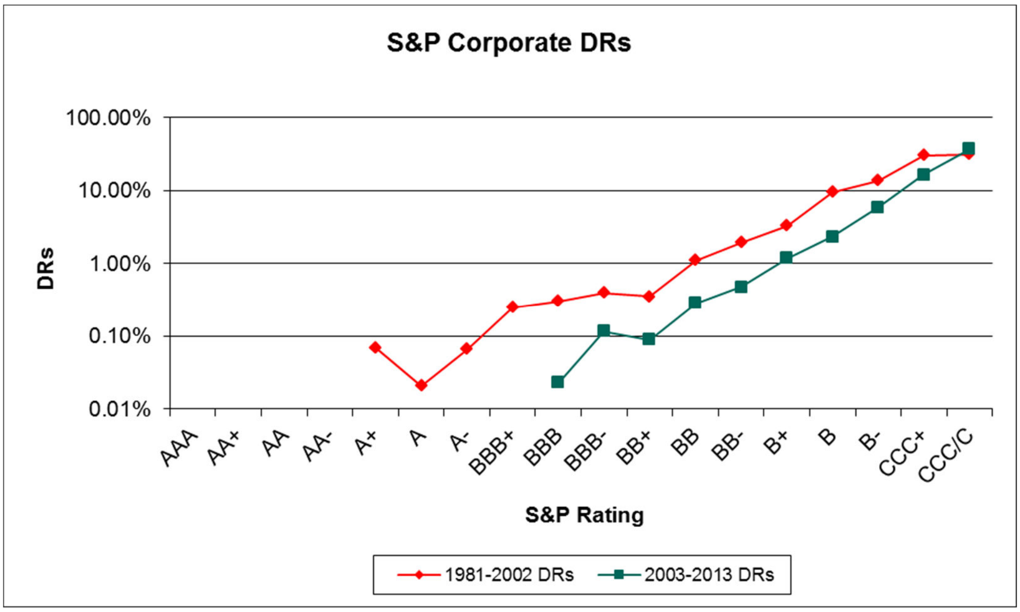

Figure 1: S&P Average DRs by Grade -- 1981-2002 and 2003-2013

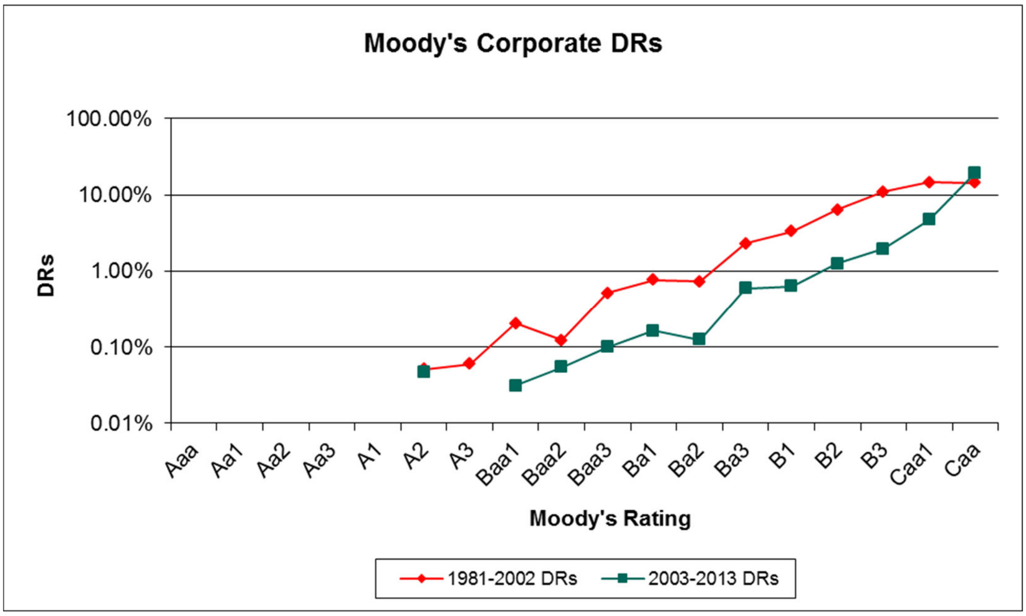

Figure 2: Moody's Average DRs by Grade: 1983-2002 and 2003-2013

To assess the statistical significance of these apparent shifts, we apply a (Probit) PD model that allows for possible shifts since 2003 in the curve that expresses the relationship between each, S&P and Moody's grade and the long-run-average Probit, default distance (DD) associated with that grade. We assume that the curve applicable to a sector (S&P Corp, S&P FI, Moody's Corp, or Moody's FI) arises by translating and rotating a smoothed, base curve that reflects the totality of S&P and Moody's default experience over 1981-2013 for Corps and FIs combined.[1] We specify that a revised curve applicable starting in 2003 may occur through a further translation or rotation (or both) of the curve that applies to 1991-2002. This translation/rotation specification holds down the number of free parameters. To guard against excluded-variable, specification error, the model also includes credit-cycle indexes (CCIs quantified as DDGAPs) that derive from median MKMV EDFs for each of 20 industry and 14 regional groupings.[2] We estimate the model (depicted below as Equation (1)) by maximum likelihood, applied separately to the S&P and Moody's default samples since 1990. We limit estimation to 1990-2013, since the EDF data available to us start in 1990.[3]

$$PD_{t+1} = \Phi\left(- \frac{DD_{i,t} + b_{S(i)} DDGAP_{I(i),R(i),t} + \Delta DDGAP_{I(i),R(i),t+1}}{\sqrt{1-\rho_{I(i),R(i)}}} \right)$$

$$DD_{i,t} = a_{0,S(i)} + s_{0,S(i)} d_{03\&} + (a_{1,S(i)} + s_{1,S(i)} d_{03\&}) DD_{g(i,t)}$$

$$\begin{aligned}
&PD_{i,t+1} = PD\ of\ the\ i^{th}\ entity\ over\ the\ year\ ending\ at\ time\ t+1 \\
&\Phi = standard-normal,\ cumulative\ distribution\ fun\ ction\ (CDF) \\
&DD_{i,t} = DD\ implied\ by\ the\ i^{th}\ entity's\ grade\ at\ time\ t \\
&DDGAP_{I(i),R(i),t} = credit-cycle\ index\ for\ the\ i^{th}\ entity's\ primary,\ industry-region \\
&\Delta DDGAP_{I(i),R(i),t} = one-year\ change\ in\ the\ credit-cycle\ index \\
&\rho_{I(i),R(i)} = correlation\ factor\ for\ the\ i^{th}\ entity's\ industry-region \\
&S(i) = i^{th} entity's\ sector\ (Corp\ or\ FI) \\
&d_{03\&} = time\ dummy\ with\ a\ value\ of\ 1\ for\ t \geq 2003\ and\ 0\ otherwise \\
&DD_{g(i,t)} = base-curve\ DD\ implied\ by\ the\ i^{th}\ entity's\ grade\ g\ at\ time\ t
\end{aligned} \qquad \text{Equation (1)}$$

---

[1] The base curve arises from a three-step process of averaging the long-run DRs for each matching S&P and Moody's grade, transforming those average DRs into DDs by applying the inverse-normal CDF, changing the sign of the result from negative to positive, and smoothing and enforcing a monotone relationship in the resulting curve of DDs per grade. To mitigate effects of sampling variation, we build the base curve using the largest available sample that combines S&P and Moody's experience with both Corps and FIs.

[2] The industry or region, credit-cycle index for a particular month arises from that month's median MKMV EDF in the industry or region less the long-run average of such medians, with the result expressed in DD units. Then, for each permissible industry-region combination, we get a combined index as a weighted average of the separate, industry and region indexes. We estimate the weights so that changes in the combined index best explains past changes in the DDs of the related companies. For more details on development and use of these indices see Aguais et al, 2004, 2007; Forest et al, 2013; Chawla et al 2013.

[3] The availability of MKMV EDF data from 1990 onwards does not limit the conclusions of this study. We have conducted this entire analysis using MKMV's research (non-production) dataset on EDFs going back to 1970s and found that the conclusions presented here still hold true.

We see below that for both S&P and Moody's, the data indicate that, over the period since 2003, the DD curve sits significantly below the curve applicable to earlier times, with this change occurring mainly as a slope adjustment. The borderline significant, negative intercept adjustment accounts for there being no more than a small shift for the lowest quality grade. But combining this intercept shift with the positive slope adjustment, we get higher DDs over the rest of the grade range, with the gap rising as default risk declines. We reject at a 99% confidence level the null hypothesis that the benchmarks drawn from default experience prior to 2003 still apply.

Table 1: PD Model Estimates for S&P Rated and for Moody's Rated Non-Financial Corps

		S&P Model			Moody's Model		
Variable	Coefficient	Estimate	Std Error	t-stat*	Estimate	Std Error	t-stat*
Constant	**a_0**	-0.39	0.06	-6.77	0.13	0.04	3.06
DD_g	**a_1**	1.10	0.03	3.33	0.90	0.02	-5.00
d_{03}	**s_0**	-0.14	0.09	-1.59	-0.11	0.07	-1.58
d_{03}	**s_1**	0.24	0.05	4.73	0.29	0.05	6.16
DDGAP **	**b**	*0.87*	*0.01*	*-13.0*	*0.80*	*0.01*	*-20.0*

* Reported t-stats are for individual null hypothesis of $a_0 = 0$; $a_1=1$; $s_0=0$; $s_1=0$. Rejection of null hypothesis would mean that default data supports the single base curve.
** The DDGAP coefficient varies by region. We show the result for global non-financial corporates. The coefficient and standard error in this case come from a preliminary, instrumental-variable regression of industry-region credit-cycle indexes on a noisier index based on a smaller sample of agency-graded companies only. The resulting instrument enters the final equation with coefficient of one. A null hypothesis of b=0 (ratings are fully PIT) is overwhelmingly rejected. t-stat is presented for null hypothesis of b=1 (ratings are fully TTC) which is also rejected.

2. Flattening of the Curve for FIs

For FIs the recent experience suggests that the grade-PD relationship has flattened, reducing the gap between PDs of the best and worst grades (**Figure 3** and **Figure 4**). Here, small samples may play a role, but the statistical results still reject at conventional confidence levels the null hypothesis that the DD curve consistent with data prior to 2003 still applies (**Table 2**).

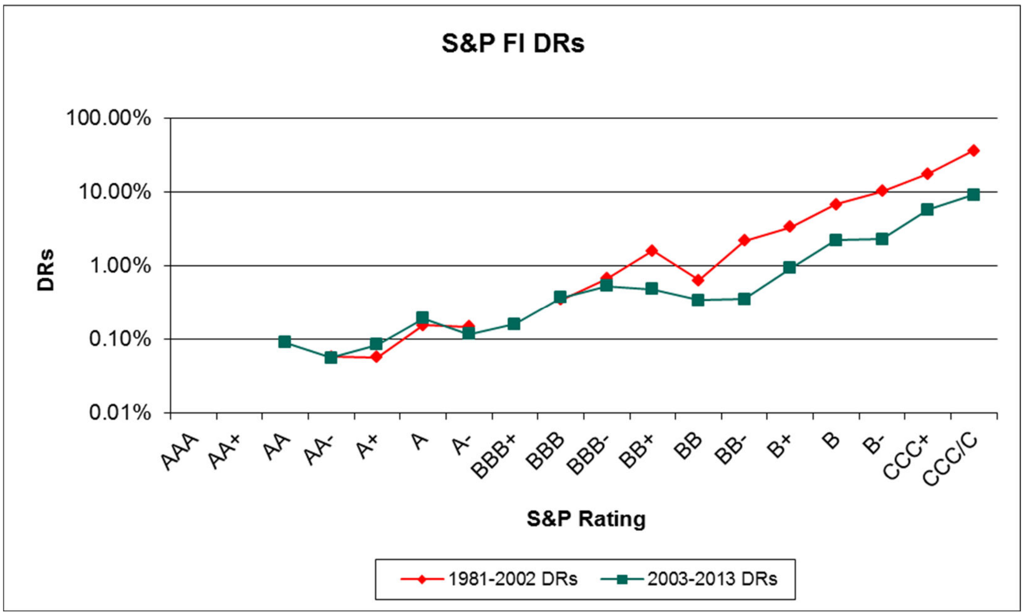

Figure 3: S&P Average DRs by Grade -- 1981-2002 and 2003-2013

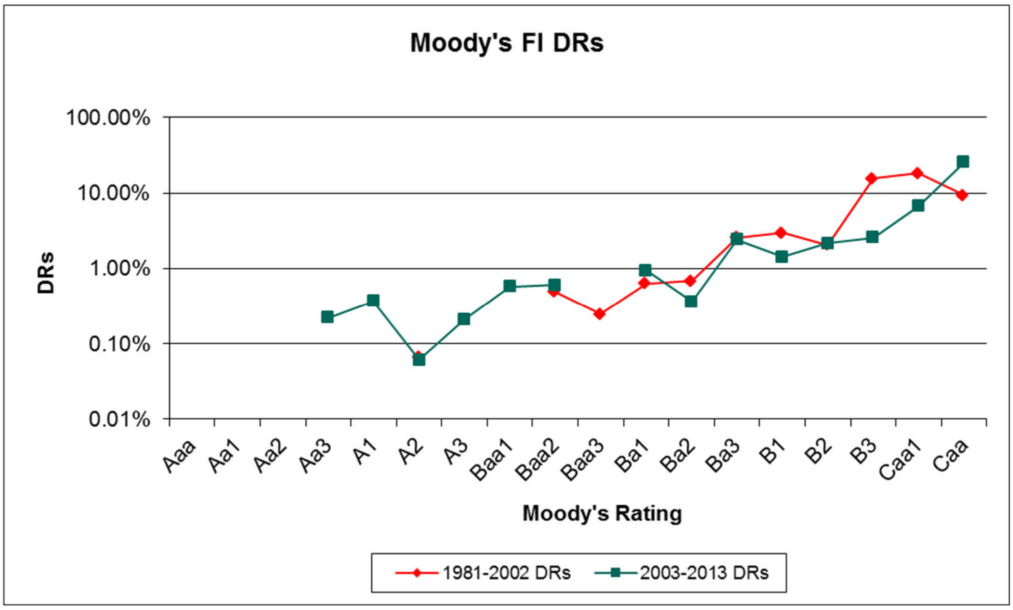

Figure 4: Moody's Average DRs by Grade -- 1983-2002 and 2003-2013

Table 2: PD Model Estimates for S&P Rated and for Moody's Financial Institutions

Variable	Coefficient	S&P Model			Moody's Model		
		Estimate	Std Error	t-stat	Estimate	Std Error	t-stat
Constant	a_0	-0.15	0.13	-1.10	-0.10	0.16	-0.66
DD_g	a_1	0.97	0.06	-0.50	1.06	0.08	0.75
d_{03}	s_0	1.00	0.18	5.42	0.78	0.22	3.53
d_{03}	s_1	-0.30	0.08	-3.92	-0.37	0.10	-3.70
DDGAP **	b	*0.81*	*0.01*	*-19.0*	*0.97*	*0.02*	*-1.50*

[*] Reported t-stats are for individual null hypothesis of $a_0 = 0$; $a_1=1$; $s_0=0$; $s_1=0$. Rejection of null hypothesis would mean that default data supports the single base curve.

Variable	Coefficient	S&P Model			Moody's Model		
		Estimate	Std Error	t-stat	Estimate	Std Error	t-stat
** The coefficient and standard error in this case come from a preliminary regression of industry-region credit-cycle indexes on a noisier index based on agency-graded companies only. The resulting instrument enters the formula above with coefficient of one. A null hypothesis of b=0 (ratings are fully PIT) is overwhelmingly rejected. t-stat is presented for null hypothesis of b=1 (ratings are fully TTC) which is also rejected.							

3. Similar Patterns in MKMV Public Firm EDF model

Since MKMV EDFs provide current PIT (rather than so-called TTC) measures of default risk, one anticipates that they would serve as benchmarks that remain relevant at each point in time, without the need, as with S&P and Moody's experience by grade, to neutralise credit-cycle effects by averaging over a series of years. But all of MKMV's recent validation documents reveal that EDFs from its Public Firm EDF 8.0 Model have been exaggerating the DRs that it calculates from its default sample. The patterns seem similar to those evident in the S&P and Moody's data. For North American Corporates, we see evidence of substantial over-estimation of default risk almost everywhere from the low- to the high-risk end of the spectrum (see Figure 6 of Crossen et al. (2011)).

Lately, MKMV has also published validation studies for Europe (including UK) and Asia-Pacific corporate segments. For both regional segments, the model performs not as well when compared to the North American segment, in terms of rank ordering and level calibration, and again consistently over predicts defaults when compared to historical default rates.

For the European corporate segment, we see model over-prediction as experienced default rates in 2001-2010 are somewhere close to the 25th percentile and always less than model predicted average EDFs (see Figure 5 of Crossen and Zhang, 2011a). For the Asian corporate segment, we see massive model over-prediction as experienced default rates in 2001-2010 are somewhere close to the 10th percentile (see left panel of Figure 5 of Crossen and Zhang, 2011b).

For Banks, we see evidence of a flattening of the risk curve (Munves et al. 2010 and see over-estimation occurring at the bottom of the investment-grade range. This is evident from the contrasting view of Figure 2a of Munves et al (2010) which depicts model predicted EDFs and observed DRs aligning very well for 1996-2006 period compared with Figure 2b of Munves et al (2010) which depicts model predicted EDFs lower than observed DRs in low risk end but model predicted EDFs higher than observed DRs in high risk end.

4. Explaining the Bias

The above evidence alerts us that some, conventional benchmarks used to assess PD model accuracy appear to have been biased up for several years. But until we gain some understanding of the sources of this bias, we can't be confident that the corrections that we would make based on recent data will remain accurate for long. So far we have considered two possible hypotheses:

- unidentified improvements in risk management within larger Corps and smaller FIs, or
- growing asymmetry in the attitudes of creditors and regulators with respect to under- and over-estimation of risk.

In light of recent criticisms subsequent to the crisis, one can easily understand credit analysts, especially those at the Ratings Agencies, and regulators including at least a subliminal, upward bias in their risk assessments. But one has difficulty imagining the delivery of such flawed information as an optimal risk-management arrangement. If creditors and analysts treat upward biased, credit-risk measures as correct, then this will likely lead to undue restraints on corporate lending and exaggerated concerns over the safety and soundness of larger banks when compared to smaller counterparts.

At this stage, we have not found any econometric evidence for testing either of the two proposed hypotheses. So we have only heuristic arguments motivating these possibilities.

The first, ‘unidentified improvements hypothesis’ arises from the observation that risk-management technology has clearly improved, but this is not something easily gauged from credit information including financial-statement data. One might consider this circumstance as similar to that involved in measuring productivity advances as a residual. Further, this view seems consistent with Duffie et al (2009) finding that frailty factors (unmeasured systematic features) affect default risk.

The second ‘asymmetric hypothesis’ arises from our own experience observing the behavior of credit officers and regulators. For example, we have seen that credit-officer, over-rides of model-produced, credit grades are disproportionately downward (in the direction of higher risk). Thus, it is not hard to imagine this tendency ratcheting up over time. Further, in work on statistical default models combining objective and judgmental inputs, we have observed that actual default rates and the objective measures tend to be trendless over long periods and the judgmental inputs as well as the related, S&P and Moody’s grades imply downward trending creditworthiness.

5. Summary

Default data over the past 11 years indicate that S&P and Moody’s in their grading and MKMV in its Public Firm EDFs have been over-stating default risk for most Corps and all but the lowest risk FIs. The source of this bias remains unclear, but growing asymmetry in the attitudes of regulators and others toward under- and over-estimation of risk may play a role. This raises the possibility that banks might be unduly restricting corporate lending.

References

Agency Ratings related

1. Altman, E. I. and H. A. Rijken (2004). “How Rating Agencies Achieve Rating Stability” Journal of Banking & Finance vol. 28, pp 2679-2714.
2. Carey, M. and M. Hrycay (2001). “Parameterizing Credit Risk Models with Rating Data” Journal of Banking and Finance vol 25, pp197-270.
3. Hamilton, D. (2005). "Moody's Senior Ratings Algorithm & Estimated Senior Ratings," Moody's Global Credit Research, July 2005.
4. Löffler, G. (2004). “An Anatomy of Rating through the Cycle”, Journal of Banking & Finance vol. 28, pp 695-720.
5. Moody’s, Maintaining Consistent Corporate Ratings Over Time, Aug 2008.

6. Moody's, Moody's Senior Ratings Algorithm & Estimated Senior Ratings Consistent Corporate Ratings Over Time, Aug 2008
7. Moody's Default & Recovery Rates of Corporate Bond Issuers, Jan 2004.
8. Standard & Poor's Rating Performance, 2002.
9. Loffer, "Can rating agencies look through the cycle?", Review of Quantitative Finance and Accounting, May 2013, Volume 40, Issue 4, pp 623-646

Point in Time and Through the Cycle framework related

10. Aguais, S. D., Lawrence R. Forest, Jr., Elaine Y. L. Wong, and Diana Diaz-Ledezma, 2004, "Point-in-Time versus Through-the-Cycle Ratings", in M. Ong (ed), The Basel Handbook: A Guide for Financial Practitioners (London: Risk Books).
11. Aguais, S.D., Lawrence R. Forest Jr, Martin King, Marie Claire Lennon, and Brola Lordkipanidze, 2007, "Designing and Implementing a Basel II Compliant PIT–TTC Ratings Framework", in M. Ong (ed), The Basel Handbook II: A Guide for Financial Practitioners (London: Risk Books). Available at http://mpra.ub.uni-muenchen.de/6902/1/aguais_et_al_basel_handbook2_jan07.pdf
12. Carlehed M. and Petrov A "A methodology for point-in-time–through-the-cycle probability of default decomposition in risk classification systems", The Journal of Risk Model Validation: Volume 6/Number 3, Fall 2012, (3-25)
13. Forest L., Chawla G. and Aguais S. D., "Comment in response to 'A methodology for point-in-time–through-the-cycle probability of default decomposition in risk classification systems' by M. Carlehed and A. Petrov", The Journal of Risk Model Validation: Volume 7/Number 4, Winter 2013/14
14. Chawla G., Forest L. and Aguais S, "Deriving Point-in-Time(PIT) and Through-the-cycle(TTC) PDs", key authors to Wikipedia article http://en.wikipedia.org/wiki/Probability_of_default#Through-the-cycle.28TTC.29_and_Point-in-Time.28PIT.29, last edited in 2013

MKMV EDF related

15. Arora, N., Bohn J., and I. Korablev, "Power and Level Validation of the EDF™ Credit Measure in the U.S. Market." Moody's KMV, 2005.
16. Crosbie, P. and J. Bohn, "Modeling Default Risk." Moody's KMV, Revised December 2003.
17. Crossen, C., Qu, S. and X. Zhang, "Validating the Public Firm EDF Model for North American Corporate Firms", Moody's Analytics, May 2011.
18. Crossen, C. and X. Zhang, "Validating the Public EDF Model for European Corporate Firms", Moody's KMV, October 2011a.
19. Crossen, C. and X. Zhang, "Validating the Public EDF Model for Asian-Pacific Corporate Firms", Moody's KMV, October 2011b.
20. Dwyer, D. and S. Qu, "EDF™ 8.0 Model Enhancements." Moody's KMV, January 2007.
21. Dwyer, D. and I. Korablev, "Power and Level Validation of Moody's KMV EDF™ Credit Measures in North America, Europe, and Asia." Moody's KMV, September 2007.

22. Hamilton D., Sun Z. and M. Ding, "Through the Cycle EDF Credit Measures", Moody's KMV, August 2011.

23. Korablev, I. and S. Qu, "Validating the Public EDF Model Performance During the Recent Credit Crisis," Moody's KMV, June 2009.

24. Kurbat, M. and I. Korablev, "Methodology for Testing the Level of the EDF Credit Measure." Moody's KMV, August 2002.

25. Munves, D., Smith, A. and D. Hamilton, "Banks and their EDF Measures Now and Through the Credit Crisis: Too High, Too Low, or Just About Right?", Moody's Analytics, 2010.

26. MKMV, 2013, "Public Firm EDF™ – Product Update", email circulated to clients in February 2013

Other

27. Black, F. and M. S. Scholes (1973). 'The Pricing of Options and Corporate Liabilities', Journal of Political Economy, 81 (3), 637-654.

28. Merton, R. C. (1974). "Theory of Rational Option Pricing." Bell Journal of Economics and Management Science 4.

29. Duffie, D., Eckner, A., Horel, G. and Saita, L. (2009), Frailty Correlated Default. The Journal of Finance, 64: 2089–2123.

Appendix A

This appendix provides further details on the default model that we use above in conducting significance tests. This model arises in three steps as described next.

Step One: We start by deriving a provisional, *base* curve of long-run average DDs for each of 18, S&P and related, Moody's, alpha-numeric grades. The 18 S&P/Moody's grades include AAA/Aaa, AA+/Aa1, AA/Aa2, AA-/Aa3, A+/A1, A/A2, A-/A3, BBB+/Baa1, BBB/Baa2, BBB-/Baa3, BB+/Ba1, BB/Ba2, BB-/Ba3, B+/B1, B/B2, B-/B3, CCC+/Caa1, <=CCC/<=Caa). To accomplish this, we

- calculate, for each matching, S&P and Moody's grade, the combined, S&P and Moody's DR over 1981-2013 for non-financial corporate entities and financial-institutions combined,
- convert these DRs by grade to DDs by applying the negative of the inverse-normal, CDF (i.e. $DD_g = -\Phi^{-1}(DR_g)$), and
- fit a smooth, parametric curve to the DDs implied by the realised DRs and in doing so enforce continuity in the curve and a monotonic relationship in which DDs rise as the grades improve.

We consolidate all of the S&P and Moody's default experience so as to minimize sampling variation in the tabulated DRs. After that, we fit a smooth, monotonic curve to those DRs. This curve optimizes an objective function that penalizes both second derivatives and deviations from the observed DRs. In the further steps below we allow the base, grade-to-long-run-average-PD curves to vary across major obligor classes (Corps, FIs), time intervals (<2003, ≥2003), and (for Corps) regions (North America, EU and UK, APAC, LatAm). However, to keep the approach simple and less demanding of the data, we assume that each, more detailed curve arises from translation and rotation of the one, smooth, provisional curve obtained in this step.

In this work, we use S&P Long term issue ratings and Moody's Long term Senior Unsecured Obligation ratings. We have used ratings outlook as an additional input in the model and found it to be insignificant in explaining DRs. The methods used by rating agencies are proprietary and the agencies differ in their approaches to credit assessment. For more details refer to the S&P and Moody's websites.

The dependent variable in a PD model is the binary default status (i.e. 1 = default, 0 = non-default). Thus, to understand what the model is predicting, one must pay close attention to the definition of default. According to S&P CreditPro, a default is recorded upon the first occurrence of a payment default on any financial obligation subject to a bona-fide commercial dispute. An exception occurs when an interest payment missed on the due date is made within the 30-day grace period. Distressed exchanges are also considered defaults whenever the debt holders are coerced into accepting substitute instruments with lower coupons, longer maturities, or any other, diminished financial terms. Bankruptcy filings also are usually accepted as definitive indicators of default.

For Moody's, the Default Risk Service (DRS) uses essentially the same definition for default as other Moody's risk management products. According to Moody's DRS, a default includes three types of credit events:

- Missed or delayed disbursement of interest and/or principal, including delayed payments made within a grace period;
- Bankruptcy, administration, legal receivership, or other legal blocks (perhaps by regulators) to the timely payment of interest and/or principal; or

- Distressed exchange occurs in which the issuer offers debt holders a new security or package of securities that amount to a diminished financial obligation (including the exchange of debt for preferred or common stock, or debt with a reduced coupon, lower par value, lesser seniority, or longer maturity) or (ii) the exchange had the apparent purpose of helping the borrower avoid default.

In general, we accept S&P or Moody's decisions on the occurrence of default. There is a notable exception, however. S&P has not identified as defaulters several financial institutions that, during the recent financial crisis defaulted on subordinate but not on senior debt. We've reclassified those cases as defaults.

Step Two: We derive, from MKMV-Public-Firm EDFs, credit-cycle indexes (DDGAPs) for selected, industry-region combinations. To create these indexes, we

- compute, for each of 20 industries and each of 14 regional groupings, times series of median EDFs,
- translate the median EDFs to median DDs by applying the negative of the inverse-normal CDF,
- form weighted averages of the median DDs for each admissible, industry-region pair using, in the case of each industry, the weights that produce the industry-region composites that best explain changes in the DDs of the companies within the industry, and
- express each, monthly, industry-region DD as a deviation (DDGAP) from a long-run, normal value calculated from long-run averages of the industry and region, median EDFs.

These indexes are latent factors. This means that they arise from summarizing the default experience that we are trying to explain with the aid of these factors. More precisely, the credit-cycle factors summarise not exactly the default experience itself but rather the Credit Edge model's estimate of that experience. We view the model estimates as instruments that depict the underlying experience more accurately than the industry-region, realized DRs, which are subject to substantial sampling error.

The use of the inverse-normal function in extracting spot estimates of DD from one-year EDFs/PDs works so long as the related, default model assumes that credit conditions evolve as a generic, random walk. In this case, the EDFs/PDs exhibit a one-to-one relationship to the spot DD. Otherwise, if, for example, the model assumed that credit conditions mean revert, the DD inferred in this manner would no longer constitute a spot estimate. Instead it would amount to an average over the coming year and that average would vary depending on the state of credit conditions at the start of the one-year horizon. The forthcoming MKMV Public Firm model will incorporate mean reversion and so, in the future, the approach for deriving DDGAPs from that model will change.

Step Three: We estimate a default model of the form displayed in equation (1). This model uses an applicable, base curve to infer a preliminary DD from an S&P or Moody's grade. That DD will amount mostly to a relative-risk (TTC) measure, but it will also incorporate the (minor) share of cyclical fluctuations picked up by S&P or Moody's ratings. It then combines that DD with the DDGAP for the company's primary industry and region. The grade-implied DD plus the current DDGAP, weighted by the proportion of the cycle not recognised by ratings, yields an estimate of the company's, current (PIT) DD. That spot DD plus an estimate of the change in the DDGAP over the coming year provides an estimate of the expected value of the DD over the year. That estimate entered into a standard-normal CDF accounting for unpredictable, DD variations yields an estimate of the one-year PD.

In calibrating this model to S&P or Moody's default experience, we obtain point estimates and standard errors for all of the parameters including those that measure the differences between the grade-to-PD curves that best explain default experience over 1991-2002 and the ones that best explain it over 2003-2013. One sees from the results reported in the text that the estimated changes in base curves are highly significant.

To further establish the validity of the results reported above, we have fitted a nested set of models listed below as Model 1, Model 2 and Model 3 and conducted hypotheses tests comparing each successive pair of models. In each case, the test rejects the maintained hypotheses that the less detailed model is the valid one. This sequence of tests leads us to the models listed below, which is the one used in testing the significance of the curve shifts after 2002. Similarly, we continued to add regional dimension to Corporate sector and found some statistical significance.

To judge our models, we make use of two statistical tests: the conventional t-test for each parameter estimate and the Likelihood Ratio (LR) test statistic for validity of models which use of additional explanatory variables. The LR statistic is mathematically defined as twice the difference of the Log Likelihoods of the two models in question. The statistic is chi-squared distributed with N degrees of freedom where N is the number of additional coefficients in used to explain the increase in likelihood.

Model 1: The same, overall, base curve applies to all rated entities, both Corps and FIs, in all regions over all time periods.

Model 2: This model assumes that there are potentially two base curves, one for Corps and one for FIs, and that each curve applies globally over all time period.

Model 3: This model assumes that, for both Corps and FIs, the base curve applicable over 2003-2013 is potentially different from the one applicable over 1991-2002.

The generalised nested formulation, i.e. Model 3 is different when compared to Equation (1) and its econometric formulation is presented in Equation (2) below. The estimation results from different model formulations for S&P default data is presented in Table 3. We see that S&P default data supports our conclusion that there are two different curves for Corporates and Financial Institutions and their default behaviour is different before and after 2003. Table 4 shows the estimation results from different model formulations for Moody's default data which also supports the overall hypothesis of different curves by asset class and time, however we note the statistical results are somewhat weaker when compared to S&P default data.

$$PD_{t+1} = \Phi\left(-\frac{DD_{i,t} + b_{S(i)} DDGAP_{I(i),R(i),t} + \Delta DDGAP_{I(i),R(i),t+1}}{\sqrt{1-\rho_{I(i),R(i)}}} \right)$$

$$\begin{aligned} DD_{i,t} &= a_0 + a_1 \cdot DD_{g(i,t)} + (a_{0,Corp} + a_{1,Corp} \cdot DD_{g(i,t)}) \cdot Sec_i + (s_0 + s_1 \cdot DD_{g(i,t)}) \cdot d_{03\&} \\ &+ (s_{0,Corp} + s_{1,Corp} \cdot DD_{g(i,t)}) \cdot d_{03\&} \cdot Sec_i \end{aligned} \qquad \text{Equation (2)}$$

$Sec_i = \textit{entity's sector dummy with a value of 1 for Corp and 0 otherwise}$

$d_{03\&} = \textit{time dummy with a value of 1 for } t \geq 2003 \textit{ and 0 otherwise}$

Table 3: S&P Estimation Results for Corps and FIs

	Model 1			Model 2			Model 3		
Coefficient	**Estimate**	**Std Err**	**t-stat**	**Estimate**	**Std Err**	**t-stat**	**Estimate**	**Std Err**	**t-stat**
a_0	-0.3182	0.0377	-8.44	0.5047	0.1028	4.91	-0.2636	0.1610	-1.63
a_1	1.1312	0.0205	6.42	0.7766	0.0424	-5.27	1.0025	0.0680	0.03
$a_{0,Corp}$				-0.9822	0.1111	-8.84	-0.2544	0.1728	-1.47
$a_{1, Corp}$				0.4432	0.0487	9.10	0.1538	0.0754	2.03
s_0							1.2597	0.2151	5.85
s_1							-0.3776	0.0885	-4.26
$s_{0, Corp}$							-1.3604	0.2321	-5.86
$s_{1, Corp}$							0.6251	0.1018	6.13
b	0.81	0.01	-19	0.87	0.011	-11.8	0.87	0.011	-11.8
b_F				0.73	0.016	-16.8	0.73	0.016	-16.8
Log Likelihood	-5191.245			-5149.440			-5054.56		
Likelihood Ratio				83.611			234.551		
Degrees of freedom used	3			6			10		
Null Hypothesis	a0=0, a1=1, b=1			a_0=0, $a_{0,Corp}$ =0, a_1=1, $a_{1,Corp}$ =0, b=1, b_F =1			a_0=0, $a_{0,Corp}$ =0, a_1=1, $a_{1,Corp}$ =0, b=1, b_F =1, s_0=0, s_1=0, $s_{0,Corp}$=0, $s_{1,Corp}$=0		
p-value of LR test				0.0000%			0.0000%		

Table 4: Moody's Estimation Results for Corps and FIs

	Model 1			Model 2			Model 3		
Coefficient	**Estimate**	**Std Err**	**t stat**	**Estimate**	**Std Err**	**t stat**	**Estimate**	**Std Err**	**t stat**
a_0	0.3363	0.0275	12.23	0.6558	0.0992	6.61	0.6400	0.1526	4.19
a_1	0.867	0.0168	-7.91	0.7044	0.0413	-7.15	0.7477	0.07	-3.60
$a_{0, Corp}$				-0.3665	0.1035	-3.53	-0.2996	0.1582	-1.89
$a_{1, Corp}$				0.1983	0.0456	4.34	0.0554	0.0743	0.74
s_0							0.007	0.2017	0.03
s_1							-0.0561	0.0879	-0.63
$s_{0, Corp}$							-0.2739	0.2112	-1.29
$s_{1, Corp}$							0.4256	0.0986	4.31
b	0.85	0.013	-11.5	0.80	0.01	-20.0	0.80	0.01	-20.0
b_F				0.986	0.018	-0.77	0.986	0.018	-0.77
Log Likelihood	-5664.4			-5627.2			-5583.4		
Likelihood Ratio				74.4			162		
Degrees of freedom used	3			6			10		
Null Hypothesis	a0=0, a1=1, b=1			a_0=0, $a_{0,Corp}$ =0, a_1=1, $a_{1,Corp}$ =0, b=1, b_F =1			a_0=0, $a_{0,Corp}$ =0, a_1=1, $a_{1,Corp}$ =0, b=1, b_F =1, s_0=0, s_1=0, $s_{0,Corp}$=0, $s_{1,Corp}$=0		
p-value of LR test				0.0000%			0.0000%		

Appendix B

This appendix lists important technical terms and acronyms used in this document.

DRs	Default Rates. In this document this refers to annual default rates for S&P or Moody's based on cohorts as of 1st Jan. The ideal objective of any PD model should be to predict temporal and cross sectional variation in DRs as closely as possible
DD	Distance to Default (or Default Distance), mostly used in context of Merton style PD models.
CCI	Credit cycle index
DDGAP	DDGAP is a quantification of credit condition using PIT-TTC dual ratings approach. It measures how far an industry or region credit conditions are from its long run average.
PD	Probability of Default
PIT	Point in Time
TTC	Through the Cycle
PIT PD	PIT PDs draw on up-to-date, comprehensive information on the related obligors, account fully for the future effects of accumulating, systematic and idiosyncratic risk, and supposed to track closely the temporal fluctuations in default rates (DRs) of large portfolios. We define the PIT PD as the unconditional expectation of an entity's probability of default
TTC PD	We define TTC PD as the conditional expectation of an entity's probability of default assuming that credit conditions are close to long term average.
PIT model	A PD model whose output is purely Point in Time (assumed or quantified as pure PIT)
TTC model	A PD model whose output is purely Through the Cycle (assumed or quantified as pure TTC)
Hybrid model	A PD model whose output is neither purely PIT nor purely TTC (assumed or quantified). In our study, we demonstrate that agency ratings in themselves are hybrid indicators of default.
EDF	Expected Default Frequency, is PDs produced by Moody's KMV Public Firm model
AIRB	Advanced Internal Ratings Based Approach, which requires own estimates of PDs
RWA	Risk Weighted Assets where own estimate of PDs is a key component and any bias would lead to over or under capitalization
PRA	Prudential Regulation Authority
HPE	Hypothetical Portfolio Exercise