'AERB': Developing AIRB PIT-TTC PD Models Using External Ratings[1]

Gaurav Chawla (corresponding author)
Aguais & Associates (AAA) Ltd
14/30 City Business Centre, Winchester, SO23 7TA
Email: gchawla@aguaisandassociates.co.uk

Lawrence R. Forest Jr.
Aguais & Associates (AAA) Ltd
2080 Mackinnon Avenue, Cardiff-by-the-Sea, CA, USA
Email: lforest@aguaisandassociates.co.uk

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

(Received 30th June, 2015; revised 13th August, 2015; accepted 1st September, 2015)

ABSTRACT
The use of credit rating agency ('CRA') ratings in a credit institution's lending practices has been directly criticised since the financial crisis, as these ratings and observed risks have diverged. Some regulators do not allow the use of external ratings as direct inputs to a credit institution's internal PD/ratings model and hence the capital planning process as regulators prefer the use of internal assessments generally. However, regulators allow for the use of CRA's long run average default rates ('DR's) to benchmark an institution's internal PD model output. A recent study conducted by us shows however that indiscriminate use of such benchmarks can introduce important biases in credit lending. Even with these criticisms and new regulatory constraints, one can still make use of the rich CRA data available to create regulatory compliant PD models. In this paper, we propose a class of 'Agency Replication' style models which make use of obligor information and CRA long term DR information. Such models are extremely useful in cases where a credit institution has limited default samples where a purely internal default based model could be potentially erroneous, and where in contrast, agencies have plenty of data supporting development of robust models. In this paper we show how one can use this class of models for modelling portfolios such as Large Corporates, Banks, Insurance, etc. We discuss our experience developing approved, AIRB models augmented by external default data and hence colloquially call them Advanced External Ratings Based ('AERB') models. We show various simplifications of the formulation and show how they can be used in PIT-TTC based credit rating systems.

Keywords: Agency PIT model, Agency Replication model, Point-in-Time (PIT), Through-the-cycle (TTC), Low Default Portfolio (LDP)

[1] 'AERB' is an acronym we started utilizing informally while developing PD models for Regulatory approval under Basel II Waivers. In contrast to AIRB, 'AERB' stands for 'Advanced External Ratings Based' models where Regulators in our recent experience require broader, external, long-run default calibrations to complement internal default data, when internal default data is limited.

1 OVERVIEW

Historically, credit rating agencies (CRA) have provided credit assessments of institutions in a manner which was conducive for development of credit markets. Over time the big three CRAs: S&P, Moody's and Fitch were recognized as 'Nationally Recognized Statistical Rating Organizations' (NRSROs) (see Code of Federal Regulations, 2015). Although some other firms were later recognized as NRSROs, this certainly provided a barrier to entry to the big three and laid the foundation of systemic importance of ratings in credit decision making. Over years, CRA ratings were embedded in a credit institution's lending practices which later received regulatory approval under Basel I and II.

However, the onset of the financial crisis changed this trend. The Financial Crisis Inquiry Report (FCIC, 2011) investigated the rating process of CDOs and criticised the role of CRAs in the financial economic crisis. Over the past few years, CRAs have had billion dollar settlements and have faced additional scrutiny from government agencies like Office of Credit Ratings at SEC in the USA and European Securities and Markets Authority in the EU.

Recently, regulatory authorities were tasked to decouple a credit institution's lending practices from CRA ratings. Almost all regulatory authorities now call for internal assessment of obligors by the credit institution and ask for collection of credit-relevant information, e.g. financial statement information rather than relying on CRA rating. For instance in the UK, the PRA Supervisory Statement SS 11/13 (PRA, 2013) Section 12.31 says that a model with agency rating as the key driver without any additional information does not meet the criteria of a valid IRB model.

Thinking in terms of PD models, this clearly means that CRA ratings as an explanatory factor in a Probability of Default (PD) model is not the best way forward in terms of regulatory approval. Note that the CRA rating is an amalgamation of financial statement data and CRA's judgemental considerations (S&P 2002, Moody's 2008), an introduction of such financial statement data and credit institution's own judgemental factors in an internal model would supersede the need for CRA rating. i.e. a model based on internally collected data would outperform a single CRA rating. Over the past decade, the use of automated financial data feeds from data vendors like S&P CapitalIQ, BvD, etc. have made such data collection easier further reducing the cost of internal assessment based on such data. In a typical AIRB (Advanced Internal Rating based) approved bank, one would typically find automated data feeds for financial data with credit officers only completing the judgemental factors. Development of such an internal PD model with controlled cost and regulatory preference would mean that there is no need for CRA rating to be an explanatory variable in the internal PD model.

Let us have a look at the other use of benchmarking in PD model development. AIRB PD models typically calibrated to internal data with substantially different assumptions in terms of definition of default and modelling techniques can result in variations in modelled PDs for the same entity rated by different credit institutions. Such a variation has led regulatory authorities to benchmark every credit institution's internal PD model output with each other and with CRA long term default rates. For instance in the UK, the Prudential Regulatory Authority (PRA), performs a Hypothetical Portfolio Exercise ('HPE'), where it compares each credit institution's PDs with medians from all reporting institutions and with S&P's default rates (DRs)[2]. Such a benchmarking exercise has its flaws but we would like to point to two big shortcomings:

- the benchmark CRA DRs could be biased in terms of industry, time or by rating, and,

[2] In this document we refer to experienced default rates as DRs. These are 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

- the internal model's PDs could be PIT or TTC or hybrid (mostly TTC in our experience) and a comparison at one instance in time would be invalid.[3]

We have described these shortcomings in our related work on Biased Benchmarks (Forest et al, 2015). But overall, the case for use of CRA ratings for benchmarking is good because of the availability of a rich history of long term default rates covering multiple recessions and a relation to any firm's credit decision making.

With the general invalidity of CRA ratings as direct explanatory variables and the valid use of agency DRs as benchmarks in mind, one can then ask the question of how and when CRA ratings can be used to develop valid AIRB PD models.

In this paper, we introduce a class of models called 'Agency Replication' models. Here, an entity's internal financial assessment and judgmental factors are regressed to corresponding CRA rating's PIT or TTC PDs (but not DRs). We show how we can use this class of models for modelling portfolios such as Large Corporates, Banks, Insurance, etc. For certain portfolios where there is insufficient data, i.e. for Low Default Portfolios (LDP), this is clearly the preferred approach because the use of strictly internal default experience can give erroneous results. We discuss our experience of Advanced Internal Ratings Based (AIRB) approval of these models and hence colloquially call this Advanced External Ratings Based ('AERB') models. We also build these models within our PIT-TTC framework allowing for dual use of PIT and TTC PDs.[4] This dual usage allows direct and consistent support for both regulatory capital objectives requiring TTC model calibrations and stress testing and IFRS9 which require in our opinion PIT model calibrations.

In the following sections we explain the modelling choices and specification, followed up with model diagnostics, benchmarking, monitoring and annual review.

2 MODELLING CHOICES AND MODEL SPECIFICATION

Generally speaking, we can categorise all PD modelling approaches into two:

i. Direct Calibration to default: Here financial and judgemental input factors are used as explanatory variables to explain direct default (binary) events. The most common class of such models is Black-Scholes-Merton style models (Black and Scholes, 1973; Merton, 1974) where proxies of leverage and volatility are used in some form to predict defaults, e.g. Moody's KMV model. Alternatively, reduced form models make use of input data to predict defaults, e.g. Kamakura models. Another example is use of smoothed default rate associated with agency grade as an explanatory variable, together with credit cycle index to explain agency's binary default event. We call this Agency PIT model and have described it in detail in our previous study (Forest et al 2015)

[3] We refer to a model as a PIT PD model if its output is purely Point in Time (assumed or quantified as pure PIT); and refer to a model as a TTC PD model if its output is purely Through the Cycle (assumed or quantified as pure TTC). Alternatively, we call as model a Hybrid model if its output is neither fully PIT nor fully TTC. In our study, we demonstrate that agency ratings in themselves are hybrid indicators of default.

[4] We define PIT PDs as estimates which draw on up-to-date, comprehensive information on the related obligors, and account fully for the future effects of accumulating, systematic and idiosyncratic risk. PIT PDs are 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. 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.

ii. Indirect Calibration to default via grade replication: Here financial and judgemental input factors are used as explanatory variables to explain default rates or default distances associated with CRA ratings. One can think of this as two models linked by agency rating. In our experience, this is a very robust alternative to direct calibration approach in wholesale credit portfolios where internal default data is scarce, but CRAs offer a wealth of long term default data. We call these Agency Replication style models.

The model developer first needs to develop the general development dataset with all relevant obligor information including CRA related information. Such data should be at obligor-year level and Table 1 shows such a typical dataset.

Table 1: Dataset for developing Agency Replication style model

Obligor's explanatory factors at time t				CRA Rating and associated DRs, PDs and DDs					
Liquidity	Profitability	Credit Cycle Index	Agency Rating at time t	CRA Rating Smoothed PDs	Default Distance	CRA Rating TTC PD (segment specific)	Default Distance	CRA Rating PIT PD (segment specific)	Default Distance[5] (DD)
2	10%	+0.5	BBB	0.21%	2.859	0.19%	2.899	0.17%	2.929
2.5	5%	+0.3	BB-	1.46%	2.181	1.13%	2.281	1.00%	2.326
3	15%	+0.4	A-	0.10%	3.098	0.09%	3.110	0.08%	3.156

Over the past decade, we have built Agency Replication style models for several portfolios such as Large Corporate, Banks, Insurance, Broker Dealers, Sovereigns and sub-Sovereign entities and have got AIRB approval from the regulator using our 'AERB' approach. In this discussion however, our comments and analysis have been generalised so that nothing presented is specific or confidential to the approved models at the banks we have recently worked for.

In almost all cases we have had 8-10 factors in PD models. The following table summarizes the availability of data for Agency PIT model, Agency Replication model and an alternative direct to default model. The richness of data is apparent from such a comparison (Table 2) and in our experience Agency Replication style models in such cases are much more robust than an alternative direct to default model.

Table 2: Comparison of data availability for various types of models

Portfolio	Data for alternative Direct to Default model	Data used for Agency PIT model	Data used for Agency Replication model
Large Corporates	~100 internal defaults in ~10 years with lower overall default rate when compared to long term agency DRs	Hundreds of defaults over 30+ years, global coverage, hundreds of defaults for segment specific (Corporate vs Financial Institutions vs Sovereign)	Thousands of entity year data points with financial/judgement al factors and corresponding Agency Ratings leading to
Banks	~30 internal defaults in ~10 years biased against large banks and in favour of small banks when compared to agency default data		
Insurance	<5 internal defaults in ~10 years		

[5] We make use of the term Distance to Default (or Default Distance or DD), in context of Merton style PD models. In Merton model, distance to default is measured as the number of standard deviation of assets that a firm has to lose before assets hit the default point.

Portfolio	Data for alternative Direct to Default model	Data used for Agency PIT model	Data used for Agency Replication model
Broker Dealers	<10 internal defaults in ~10 years	calibrations	robustness of Agency Replication model estimates
Sovereigns	<5 internal defaults in ~10 years with lower overall default rate when compared to long term agency DRs		
Local Authorities, State Governments, Counties, Municipalities, etc	<20 internal defaults in ~10 years and mostly disconnected with Sovereign ratings/PDs.		

We have developed the Agency Replication class of models within the Point-in-Time (PIT) and Through-the-Cycle (TTC) dual PD/ratings approach developed and presented in Aguais et al, 2004, 2007; Forest et al, 2013; Chawla et al 2013. The PIT-TTC framework supports a more detailed analysis of cross-time variations in agency grades by controlling for systematic credit conditions using the PIT-TTC framework's credit cycle indices (CCI). This is extremely important due to the hybrid nature of agency ratings as demonstrated by Forest et al 2015.

In developing an Agency Replication style model, a model developer faces three choices:

i. The explanatory variable, i.e. obligor's input variables (Xs),
ii. The dependent variable, i.e. default rate or default distance (Y)
iii. The link function

With regards to the choice of explanatory variables, one can make use of financial data from external vendors and/or an internal ratings database and judgemental factors from an internal ratings database. Such explanatory variables can be re-defined based on expert's preference or ongoing data availability. Naturally, such factors vary by asset class and one can benchmark the choice with factor selection from other studies in this area. We broadly agree with choice of explanatory variables in other studies such Altman and Rijken, 2004; Kamstra et al 2001; Minardi et al. 2007; Mizen and Tsoukas, 2011 etc.

The key value addition of this study however is the choice of the dependent variable. Most studies on Agency Replication type models make use of ordinal ratings (AAA=1, AA+=2, …) or similar logic and fit an ordered logit regression model.

We believe this is not the best approach because the notches are not equidistant in terms of default behaviour (e.g. there is no meaningful difference in default behaviour in AA+/AA/AA- space). It also completely ignores the variation of CRA default rates and assumes that a rating means the same DR behaviour over time.

We recommend the use of forecasted output of the Agency PIT model, i.e. the Point in Time (PIT) Probability of default (PDs) and associated Default Distance (DD) as the dependent variable for the Agency Replication style models. Table 3 clarifies the input, outputs and forecasts of this way of modelling.

Table 3: Model Specification

Class of Model	Explanatory Factors	Dependent Variable	Model Output	Model Development Dataset	Reference

Direct to default model called Agency PIT model	Smoothed DR corresponding to Agency Rating and Credit Cycle Index only	Agency Binary Default Indicator (0,1)	PIT PD and DD for a particular grade and time. E.g. PIT PD for BBB grade at 1st Jan 2015 = 0.21%	Agency ratings and default dataset covering decades and hundreds of defaults	Forest et al 2015
Agency Replication style model	Obligor's financial and judgemental factors	Model Output of Agency PIT model	Agency Replication PIT and TTC PDs	Co-rated dataset between firm's or data vendor's and agency rated world covering obligor financial/judgemental factors and agency rating related information	This study

The exact model specification for the Agency PIT model is described in our previous study (Forest et al 2015). In summary, the Agency PIT model makes use of a smoothed default rate associated with a CRA rating and credit cycle index as explanatory variables to predict Point in Time (PIT) PD associated with any associated grade. The agency PIT model is required because agency ratings are hybrid and not purely PIT or TTC in nature. Our previous study indicates that CRA ratings are 75%-80% TTC in nature. When we make use of output of the Agency PIT model as a dependent variable for the Agency Replication model we make sure that we are predicting something which is completely PIT as shown in equation (1).

Based on this understanding, we can now specify the Agency Replication model as:

$$\begin{aligned} PIT\_PD_{i,t} &= \Phi\left(-\frac{DD_{i,t} + b \cdot DDGAP_{I(i),R(i),t} + \Delta DDGAP_{I(i),R(i),t}}{\sqrt{1-\rho_{I(i),R(i)}}}\right) \\ PIT\_DD_{i,t} &= \frac{DD_{i,t} + b \cdot DDGAP_{I(i),R(i),t} + \Delta DDGAP_{I(i),R(i),t}}{\sqrt{1-\rho_{I(i),R(i)}}} \qquad (1) \\ DD_{i,t} &= \beta_0 + \beta_k \cdot F_{k\,i,t} \end{aligned}$$

where $PIT\ PD_{i,t}$ is the PIT Probability of Default for the i^{th} entity at time t
$PIT\ DD_{i,t}$ is the PIT Distance to Default for the i^{th} entity at time t
Φ - standard normal cumulative distribution function
$DD_{i,t}$ – company-specific-default distance for i^{th} entity at time t
$DDGAP_{I(i),R(i),t}$ – industry (I) and region (R), credit-cycle index (CCI), at time t. 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.
$\Delta\ DDGAP_{I(i),R(i),t+1}$ –change in industry (I) and region (R) credit cycle index (from *t* to *t+1*)
b – regression coefficient which denotes the degree of TTC-ness of $DD_{i,t}$
ρ – correlation factor related to *DDGAP*
$F_{ki,t}$ – k^{th} factor value for i^{th} entity at time t
β_0 – model intercept
β_k – regression coefficient for k^{th} factor

We estimate the coefficients (β_0, β_k and b) of the Agency Replication model by minimizing the sum of squares the difference between predicted values of PIT DD from Agency Replication model and Agency Direct model:

$$\underbrace{\min}_{\{\beta 0,\beta k,b\}} SSE = \sum_i \left(PIT\_DD_{i,t} - PIT\_DD^A_{i,t}\right)^2 \tag{2}$$

where $PIT\ DD_{i,t}$ is the PIT Distance to Default for the i^th^ entity at time t from Agency Replication model (as described in previous equation)
$PIT\ DD^A_{i,t}$ is the PIT Distance to Default for the i^th^ entity at time t from Agency PIT model (as described in Forest et al 2015)
min function minimises the sum of square errors which are assumed to be normally distributed

In practice, however, we always see b = 1, i.e. the firm's financial and judgemental factors are TTC in nature and credit cycle indices explain all the default behaviour in which case the entire formulation can be simplified into TTC PD, i.e.

$$\underbrace{\min}_{\{\beta 0,\beta k\}} SSE = \sum_i \left(DD_{i,t} - TTC\_DD^A_{i,t}\right)^2 \tag{3}$$

where $DD_{i,t}$ is the Distance to Default for the i^th^ entity at time t from Agency Replication model (as described in previous equation) = TTC $DD_{i,t}$ when b = 1
$TTC\ DD^A_{i,t}$ is the PIT Distance to Default for the i^th^ entity at time t from Agency PIT model (as described in Forest et al 2015)

To simplify this formulation, even further one can make use of an institution's existing model's scores as an explanatory variable and/or average TTC PDs per CRA rating as the dependent variable. This then reduces to creating a link function between the existing scores and the average TTC default distance per CRA rating. The use of simple logistic regression between existing scores and average Agency TTC DDs would mean the model is as simple as estimation of two coefficients. However, such simplification works only in certain assumptions and does not provide a clean mechanism for deriving dual PDs.

$$\begin{aligned} &\underbrace{\min}_{\{a0,a1\}} SSE = \sum_i \left(DD_{i,t} - \underset{t}{average}\,(TTC\_DD^A_{i,t}) \right)^2 \\ &DD_{i,t} = \Phi^{-1}\left(\frac{1}{1+\exp(a_0 + a_1 \cdot S_{i,t})} \right) \end{aligned} \tag{4}$$

where $DD_{i,t}$ is the Distance to Default for the i^th^ entity at time t from Agency Replication model , as a function of existing model's scores, most probably as a logistic regression.

We do not recommend any further simplification of Agency Replication style models as any further simplification leads to errors in calibration and interpretation. Some researchers and model developers do simplify the formulation even further by making use of default rates corresponding to CRA ratings rather than PIT or TTC PDs, as shown in equation (5) below. In doing so, the Agency PIT model formulation is not used at all and one simply makes use of associated default rates behind them.

$$\underbrace{\min}_{\{a0,a1\}} SSE = \sum_i \left(DD_{i,t} - \underset{t}{average} (DR_{i,t}^{A}) \right)^2$$
$$DD_{i,t} = \Phi^{-1}\left(\frac{1}{1+\exp(a_0 + a_1 \cdot S_{i,t})} \right) \quad (5)$$

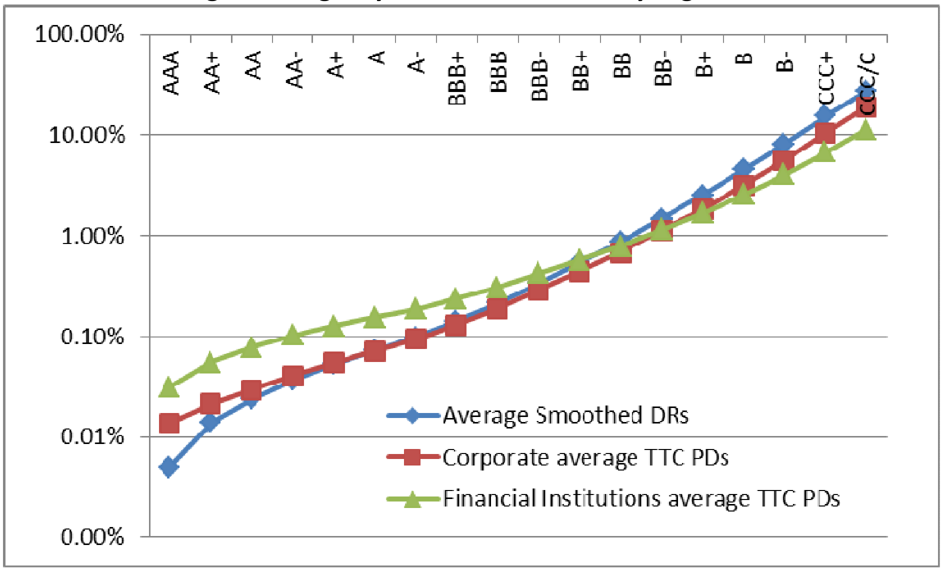
where $DD_{i,t}$ is the Distance to Default for the i[th] entity at time t from Agency Replication model , as a function of existing model's scores, most probably as a logistic regression.

In our experience, this always leads to a bias in estimates. For instance if one builds the model on the past 6 years of CRA default rate history then there is no recession in it. If one builds on past 10-15 years of CRA default history then this is biased against larger banks and in favour of larger corporates when compared to the longer CRA default history. A model making full use of CRA history spanning decades cannot be built because of lack of commensurate long history of financial assessments and judgmental factors as explanatory variables for left hand side of the equation. For this reason we strongly recommend that one makes use of Agency PIT model derived PIT and/or TTC PDs for the development of Agency Replication style models.

The most important decision that goes into development of Agency Replication model is not the choice of input variables or the link function but the choice of an appropriate segment in the Agency Direct model, i.e. the calibration curve appropriate for the portfolio. As we have shown in our previous research (Forest et al 2015), the calibrations are statistically different for corporates vs financial institutions and to a certain degree within financial institutions as well.

In the figure below we show how the choice of Agency PIT model's segment specific TTC PD curve leads to very different PDs when compared to other segments and the average smoothed DRs. For details of derivation of such curves, refer Forest et al (2015).

Figure 1: Agency PIT model TTC PDs by segment



3 MODEL DIAGNOSTICS, BENCHMARKING, MONITORING AND REVIEW

After developing the leading Agency Replication model, the natural next step is to run model diagnostics and conduct sensitivity analysis. Since Agency Replication style models are based on thousands of data points over decades of history they are fairly stable. We have seen R^2 of 80% (r ~ 90%) for most models. The next step is to see the model performance on internal default dataset but only if there are sufficient (at least more than 20 defaults). A challenger internal direct to default model will most of the time outperform the Agency Replication style model when tested on this small sample of default because the direct to default model is built on it. However, direct to default models typically underperform on holdout samples. Hence, we recommend bootstrapping or conducting out of time/out of sample test to compare performance of Agency Replication style model compared to direct to default model.

In certain cases, the internal default dataset could be enough to draw robust conclusions and could be very different compared to Agency long term default rates (e.g. defaults in Banking sector in past decade have been very different compared to the preceding two decades) and hence the Agency Replication model will definitely underperform. If the credit institution strongly believes that the default experience of last decade is relevant and they want to develop models based on such experience then such a situation can be easily remedied by selecting a time segment in the Agency PIT model calibration curve. We have proposed such calibration in our previous study (Forest et al 2015). Alternatively, one could take the output of the Agency Replication model default distances and further calibrating them to internal defaults using a parsimonious formulation, thereby maintaining the relative weights of explanatory variables and rank ordering behaviour of Agency Replication model but adding an internal calibration overlay on top. The equation below shows such an internal calibration add-on done subsequent to Agency Replication step.

$$\underbrace{\max}_{\{a_0, a_t\}} LL = \sum_i d_{i,t} \cdot \ln\left(PIT\_PD_{i,t}^{INT}\right) + \left(1 - d_{i,t}\right) \cdot \ln\left(1 - PIT\_PD_{i,t}^{INT}\right)$$

$$PIT\_PD_{i,t}^{INT} = \Phi\left(-PIT\_DD_{i,t}^{INT}\right)$$
$$PIT\_DD_{i,t}^{INT} = a_0 + a_1 \cdot PIT\_DD_{i,t} \qquad (6)$$

where $PIT\_DD_{i,t}$ is the Point in Time Distance to Default for the i^{th} entity at time t from Agency Replication model (as described in previous equations)
$PIT\_DD_{i,t}^{INT}$ is the 'Internal' Point in Time Distance to Default for the i^{th} entity at time t derived using the Agency Replication model, $PIT\_DD_{i,t}$ by means of a simple linear transformation in DD space using an intercept and slope coefficient.
a_0 = intercept; a_1 = slope coefficient; these are estimated empirically by maximinsing log likelihood based on internal defaults. We test hypothesis of $a_0 = 0$; $a_1 = 1$, to check if internal calibration is any different from that prescribed by pure Agency Replication model
$d_{i,t}$ = default indicator for i^{th} entity at time t

Agency Replication style models are fairly robust and almost nothing changes over a quarter. An annual review might reveal changes in agency rating methodology or evolution in financial statement data (e.g. adoption of IFRS accounting standards) which might necessitate adjustments in the model, however such changes are rare. In our experience of conducing annual reviews of such models we have rarely made any changes to such models.

After model implementation, the best way to check ongoing performance of Agency Replication model is benchmarking of model output with agency ratings. There are likely to be ~100 new ratings per quarter and commensurate external ratings to compare. The Agency Replication model should be compared with the appropriate benchmark that is was built on. Any large deviations between internal and agency ratings should be studied. As a measure of success, one generally thinks that a deviation of 2 or maybe 3 notches is acceptable for individual cases and there should be no significant net overall bias. One typically creates a co-rated matrix between internal ratings mapped to agency rating scale versus actual CRA ratings and expects the diagonal to be dominant with 2 or 3 notch deviation around the diagonal. The typical expectation is shown in Figure 2 where one expects the majority of the population to be along the light grey diagonal with some individual ratings along either side of diagonal in a symmetric way so that the net bias is zero.

Figure 2: Co-rated matrix for Model Benchmarking

	Credit institution's Internal Rating mapped to S&P Rating scale using internal master scale and S&P TTC PDs																
Actual S&P Grade	**AAA**	**AA+**	**AA**	**AA-**	**A+**	**A**	**A-**	**BBB+**	**BBB**	**BBB-**	**BB+**	**BB**	**BB-**	**B+**	**B**	**B-**	**CCC+**	**CCC-C**
AAA																		
AA+																		
AA																		
AA-																		
A+																		
A																		
A-																		
BBB+																		
BBB																		
BBB-																		
BB+																		
BB																		
BB-																		
B+																		
B																		
B-																		
CCC+																		
CCC-C																		

However, in reality, this never happens because of the design of Agency Replication style models. The net bias is designed to be zero, but the perfect diagonal is not possible is design. Instead we often see majority entities along the dark grey diagonal, i.e the internal model overpredicts risk compared to CRA ratings in the highly rated end and underpredicts risk compared to CRA ratings in low rated end. Model developers should not be surprised with this phenomena and should design tolerance limits by rating level, e.g. 2 notches in BBB range and 4 notches in AAA and CCC end.

We present here a simple mathematical proof of why this happens. Consider two rating systems:

$$\begin{aligned} DD_1 &= \mu_1 + \varepsilon_1 \\ DD_2 &= \mu_2 + \varepsilon_2 \\ \varepsilon_1 &\sim N(0, \sigma_1^2) \\ \varepsilon_2 &\sim N(0, \sigma_2^2) \end{aligned} \qquad (7)$$

where DD_1 *and* DD_2 are two Default Distance based rating assessments.
μ is the mean DD and ε is the variation which has a very different interpretation depending if the rating system is PIT or TTC or hybrid.

If the two systems are equally calibrated (for instance in Agency Replication style model by design), then $\mu_1 = \mu_2 = \mu$. Further assume that the two systems are correlated with a parameter ρ, which means we can define the bivariate normal distribution as:

$$\Pr(DD_1 / DD_2 = x) = N\left(\mu + \frac{\sigma_1}{\sigma_2}.\rho.(x - \mu), (1 - \rho^2)\sigma_1^2\right) \qquad (8)$$

A big driver of this distribution is ratio of variances. In case of Agency Replication model, the variance of prediction is smaller when compared to variance in output of Agency PIT model, generally because of the estimation process as a regression equation even with high R^2 ~ 80%. But even if we assume $\sigma_1 = \sigma_2 = \sigma$, the bivariate normal distribution reduces to

$$\begin{aligned} &\Pr(DD_1 / DD_2 = x) = N\left(\mu + \rho.(x - \mu), (1 - \rho^2)\sigma_1^2\right) \\ &\Rightarrow E\left(DD_1 / DD_2 = x\right) = \rho \cdot x + \mu(1 - \rho) \end{aligned} \qquad (9)$$

Clearly,

$$\begin{aligned} E\left(DD_1 / DD_2 = x\right) &= x, when\ \mathrm{x} = \mu \\ E\left(DD_1 / DD_2 = x\right) &> x, when\ \mathrm{x} < \mu \\ E\left(DD_1 / DD_2 = x\right) &< x, when\ \mathrm{x} > \mu \end{aligned} \qquad (10)$$

This simply means that the when compared to agency ratings, the Agency Replication model will over-predict risk at low risk end and under-predict risk at high risk end. Simply speaking the Agency Replication model cannot create as many AAA and CCC-C grades purely based on financial and judgemental data when compared to actual CRA ratings. This effect reduces greatly if CRA judgemental factors are baked into the Agency Replication model and the model is estimated effectively. Since a historical dataset of such CRA judgemental factors is not possible, this would involve assigning sufficiently powerful coefficients to CRA judgemental factors in Agency Replication model, something which regulators would not approve easily.

4 MODEL USAGE

The model as described should preferably be used in a PIT/TTC dual PD/ratings framework. In our opinion, both European and US regulators dislike the use of agency ratings as a primary driver and prefer internal assessments of credit risk. However, the Agency Replication model structure is used to derive the relative and absolute risk assessment due to rich default history. We have developed several models using this approach and have obtained AIRB waivers (post financial crisis); hence this is a tested approach. The resulting models' TTC PDs are used to drive Regulatory Capital Risk Weighted Assets (RWA) calculation and PIT PDs used for benchmarking.

In our experience, credit officers at the higher end of wholesale credit lending view credits in a 'ratings' framework. This may or may not be enshrined in a credit institution's credit policies but is definitely embedded in credit practice and processes. A credit institution's credit lending policy could be based on PIT or TTC ratings – either way the Agency Replication model provides such outputs for credit decision making. In our experience, Agency Replication style models serve as primary models in a credit institution, with Agency PIT based models used for benchmarking of output. This can either be done on a PIT basis or a TTC basis but must be done consistently. For this reason, we highly recommend, developing Agency Replication and Agency PIT style models in a consistent PIT-TTC framework.

5 SUMMARY

The use of credit rating agency (CRA) ratings in a credit institution's lending practices has been severely criticised since the financial crisis. Recent studies also show biases in benchmarks created using CRA long term default history. Despite criticisms and regulatory constraints, CRA rating and default time series is a rich source of data and is a leading choice for building models where internal data is not sufficient to build direct to default models. In this paper, we proposed a class of Agency Replication style models which makes use of the entity's financial and judgemental information as the explanatory variable and Agency PIT model's PIT and/or TTC PD output as the dependent variable.

We have provided a full formulation of such a model built in PIT-TTC framework and motivated the reasoning for developing it within the PIT-TTC framework. Development of an integrated, dual ratings approach is the only way in our experience that multiple regulatory objectives, for capital driven by TTC measures and stress testing and IFRS9 driven by PIT measures – can be consistently and accurately supported.

We then demonstrate some commonly made simplifications to the model based on reasonable assumptions. We show how use of agency default rate information instead of Agency PIT model PIT/TTC PDs as dependent variable can introduce biases. We shared our experience of building models for portfolios such as Large Corporates, Banks and Insurance. Such models received Advanced Internal Ratings Based (AIRB) approval and hence we colloquially call them Advanced External Ratings Based ('AERB') models.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

- Aguais, S. D., Forest, L. R. Jr., Wong, E. Y. L., and Diaz-Ledezma, D. (2004). Point-in time versus through-the-cycle ratings. In *The Basel Handbook: A Guide for Financial Practitioners*, Ong, M. (ed). Risk Books, London.
- Aguais, S. D., Forest, L. R. Jr., King, M., Lennon, M. C., and Lordkipanidze, B. (2007). Designing and implementing a Basel II compliant PIT–TTC ratings framework. In *The Basel Handbook II:A Guide for Financial Practitioners*, Ong, M.(ed).Risk Books, London.
- Altman, E. I., and Rijken, H. A. (2004). How rating agencies achieve rating stability. *Journal of Banking and Finance* **28**, 2679–2714.

- Black, F., and Scholes, M.S. (1973).The pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3), 637–654.
- Code of Federal Regulations (17 CFR 240.17g-1) https://www.law.cornell.edu/cfr/text/17/240.17g-1
- Chawla, G., Forest, L., and Aguais, S. (2013). Deriving point-in-time (PIT) and through the-cycle (TTC) PDs. URL: http://en.wikipedia.org/wiki/Probability_of_default#Throughthe-cycle.28TTC.29_and_Point-in-Time.28PIT.29.
- Financial Crisis Inquiry Commission, "Financial Crisis Inquiry Report", 2011, available online at http://www.gpo.gov/fdsys/pkg/GPO-FCIC/pdf/GPO-FCIC.pdf
- Forest L., Chawla G. and Aguais S. D., (2013) "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, available at http://www.risk.net/journal-of-risk-model-validation/technical-paper/2317266/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
- Forest L., Chawla G. and Aguais S. D., (2015) "Biased Benchmarks", The Journal of Risk Model Validation: Volume 9/Number 2, Summer 2015, available at http://www.risk.net/journal-of-risk-model-validation/technical-paper/2412440/biased-benchmarks
- Kamstra, M., Kennedy, P. and T. K. Suan, *Combining Bond Rating Forecasts Using Logit*, May 2001.
- Minardi, A.M.A.F., Sanvicente, A.Z. and R. Artes, *A Methodology for Estimating Credit Ratings and the Cost of Debt for Business Units and Privately held Companies*, 2007.
- Merton, R. C. (1974). Theory of rational option pricing. *Bell Journal of Economics and Management Science* **4**(1), 141–183.
- Mizen, P. and S. Tsoukas, *Forecasting US Bond Default Ratings Allowing For Previous And Initial State Dependence In An Ordered Probit Model*, August 2011.
- Moody's (2008). Moody's senior ratings algorithm and estimated senior ratings. August.
- Prudential Regulation Authority (PRA), "Internal Ratings Based approaches", Supervisory Statement 11/13, available online at http://www.bankofengland.co.uk/pra/Documents/publications/ss/2013/ss1113.pdf
- Standard & Poor's (2002). Rating performance.