Predicting Loss Given Default: An Extension of Single Factor Model

Xiao Yao, Jonathan Crook, Galina Andreeva

Credit Research Centre, University of Edinburgh Business School

CRC Credit Research Centre

Outline

- Literature Review
- The Models
- The Data
- Empirical Evidences
- Conclusions and Future Research

CRC Credit Research Centre

1. Literature Review

- Fixed Effect Models:
  - Parametric models: generalized linear models, survival models, etc.
  - Non-parametric models: decision trees, neural networks.
- Mixed Effects Models:
  - Factor models

Fixed Effect Models

- Linear Regression:

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

- Fractional Response Regression:

$$E(y_i \mid \mathbf{x}_i) = G(\boldsymbol{\beta}^T \mathbf{x}_i)$$

- Fractional response regression can be regarded as a generalized linear regression model where the dependent variable is bounded between 0 and 1 by imposing a link function.

$$G(\boldsymbol{\beta}^T \mathbf{x}_i) = \exp(\boldsymbol{\beta}^T \mathbf{x}_i) / (1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_i))$$

- Beta Regression:

$$f(y;\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}$$

where the mean and variance are defined as:

$$E(y)=\mu,\quad Var(y)=\frac{\mu(1-\mu)}{1+\phi}$$

Here y is defined in the interval (0, 1)， however, in most practical cases there are massive observations with their LGD equal to zero or one.

- Inflated beta regression:

$$bi_{01}(y;\pi,\psi,\mu,\phi)=\begin{cases}\pi(1-\psi) & if \;\; y=0\\ \pi\psi & if \;\; y=1\\ (1-\pi)f(y;\mu,\phi) & if \;\; y\in(0,1)\end{cases}$$

- We can choose the link function as follows to reparameterize the parameters of above as:

$$g(\mathbf{x}_\mu)=\frac{\exp(\boldsymbol{\beta}_\mu^{\mathrm{T}}\mathbf{x}_\mu)}{1+\exp(\boldsymbol{\beta}_\mu^{\mathrm{T}}\mathbf{x}_\mu)},\;\; h(\mathbf{x}_\phi)=\exp(\boldsymbol{\beta}_\phi^{\mathrm{T}}\mathbf{x}_\phi)$$

CRC Credit Research Centre

Mixed Effects Models:

- Dullman and Trapp, 2004

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \sigma\sqrt{\rho} X_t + \sigma\sqrt{1-\rho}\varepsilon_i$$

where X_t is termed as the systematic risk factor and $\boldsymbol{\varepsilon}$ denotes the idiosyncratic risk factor. They both follow standard normal distribution. Here $\boldsymbol{\sigma}$ is the volatility of recovery rate and $\boldsymbol{\rho}$ is the recovery rate correlation.

- Hamerle *et al*, 2005

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \gamma_1 X_t + \gamma_2 \varepsilon_i$$

CRC Credit Research Centre

The systematic risk factor is regarded as the random effect term and the single factor model is actually a linear mixed effects model.

The research question is:

1) ***Does the inclusion of random effect term make LGD predictions better ?***

2) ***If we can make further improvements by extending the single factor model ?***

2. The Models

- Single Factor Model
- Two Factor Model
- Inflated Beta Mixed Effects Models

CRC Credit Research Centre

1) Single Factor Model

- In Hamerle's work the systematic risk factor is specified with respect to the default years of relevant instruments. The model is given as:

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \gamma_1 X_t + \gamma_2 \varepsilon_i \quad t = 1, \ldots, T$$

- In this study we also investigate the the random effects associated with seniority level and obligor level such as:

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \gamma_1 X_s + \gamma_2 \varepsilon_i \quad s = 1, \ldots, S$$

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \gamma_1 X_k + \gamma_2 \varepsilon_i \quad k = 1, \ldots, K$$

- For any two instruments within the same obligor, the covariance and correlation of their recovery rates can be derived as:

$$Cov(y_i, y_j) = \sigma^2 \rho = \gamma_1^2$$

$$Corr(y_i, y_j) = \rho = \frac{\gamma_1^2}{\gamma_1^2 + \gamma_2^2}$$

- However, during different economic periods the systematic factor may follow different distribution, so we propose a two factor model as follows:

2) A Two Factor Model

$$y_i = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + pY + (1-p)Z + \sigma_3 \varepsilon_i$$

Here the single latent systematic risk factor X is replaced by a mixture distribution of two random factors such as:

$$\begin{cases} Y \square \quad) \quad p \\ Z \square \quad) \quad 1-p \end{cases}$$

- This specification means the systematic risk is expressed by two states with different probabilities.

- For any two instruments of the same group, the covariance and correlation of their recovery rates are given by follows:

$$Cov(y_i, y_j) = p^2\sigma_1^2 + (1-p)^2\sigma_2^2$$

$$Corr(y_i, y_j) = \frac{Cov(y_i, y_j)}{p^2\sigma_1^2 + (1-p)^2\sigma_2^2 + \sigma_3^2}$$

3) Inflated Beta Mixed Effects Model

- We insert the random effects into the inflated beta regression model to define an inflated beta mixed effect model:

$$\mu = g(\mathbf{x}_{\mu} + \gamma u)$$

$$\phi = h(\mathbf{x}_{\phi} + \lambda v)$$

$$g(\mathbf{x}_{\mu}) = \frac{\exp(\mu)}{1+\exp(\mu)}, \quad h(\mathbf{x}_{\phi}) = \exp(\phi)$$

CRC Credit Research Centre

3. The Data

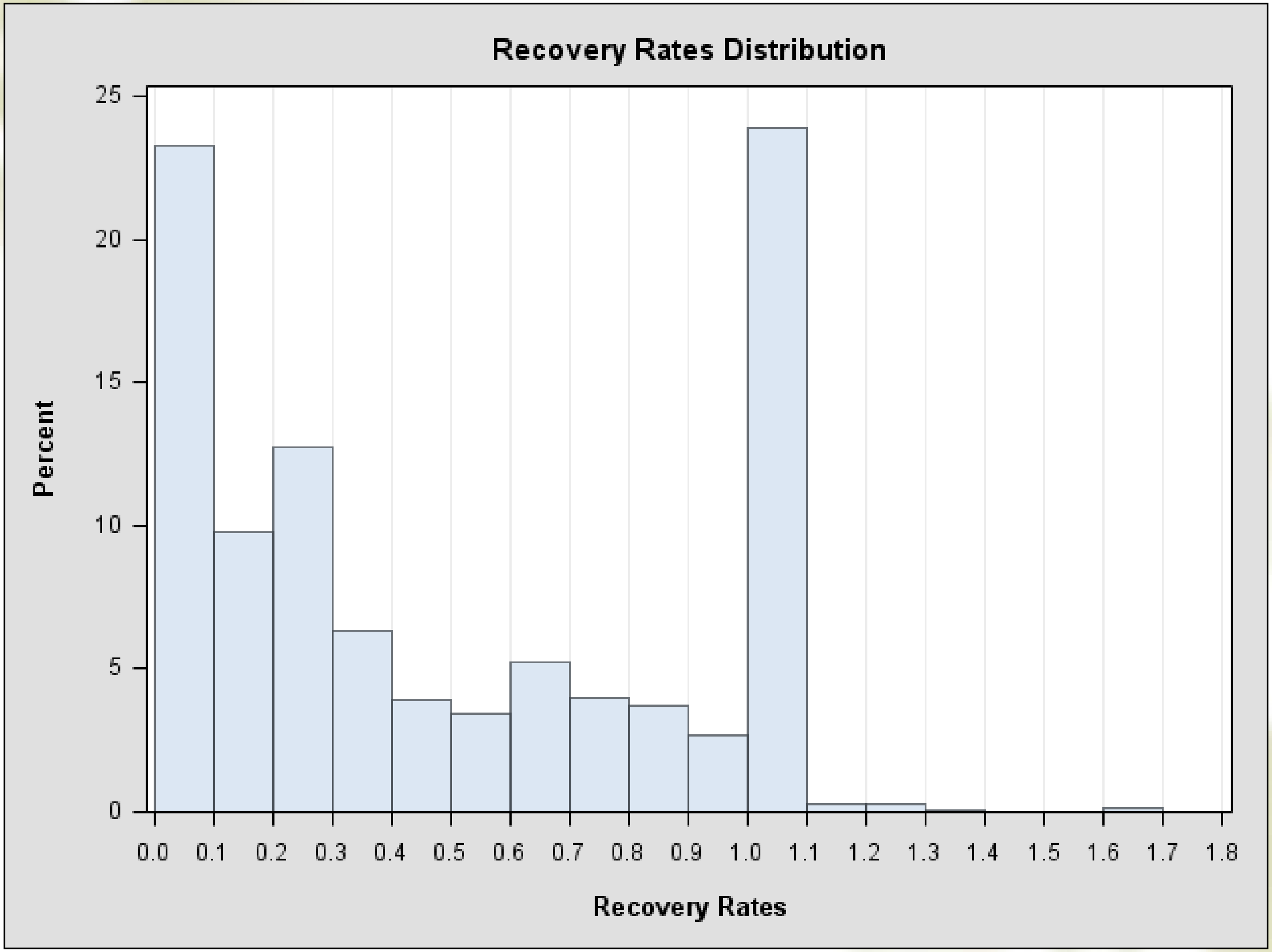
- Recovery information is from Moody's Ultimate Recovery Database (MURD). This database covers the recovery information of more than 3000 instruments range from 1985 until now.
- Accounting information is extracted from Compustat to describe the firm characteristics.
- Macroeconomic variables are obtained from Multiple open resources to capture the economic cyclical trends.

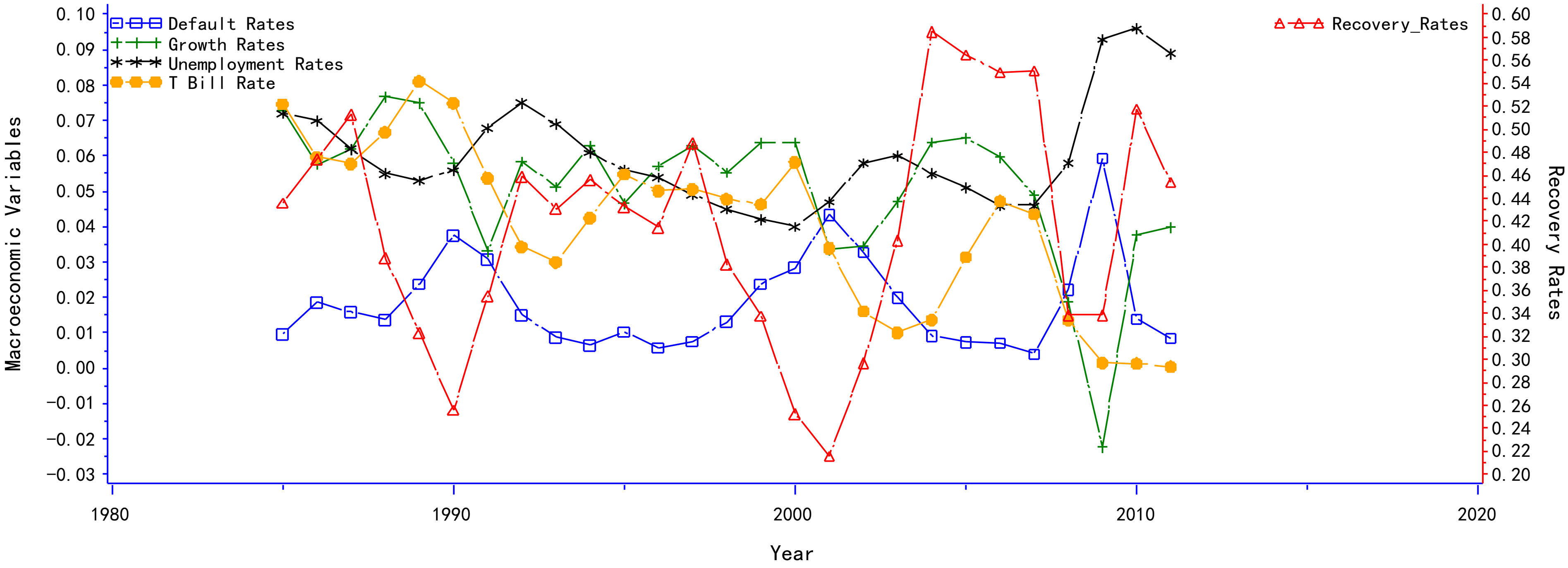
Seniority	No.	Mean	Std	Min	Max
Junior Subordinate	28	0.1628	0.2634	0	1
Senior Secured	332	0.6292	0.3688	0	1.1298
Senior Subordinate	198	0.3150	0.3617	0	1.6978
Senior Unsecured	681	0.5100	0.3813	0	1.0499
Subordinate	174	0.3217	0.3743	0	1.3691

CRC Credit Research Centre

Variables	Mean	Median	Std	Min	Max
Debt Characteristics					
Collateral Rank	2.1677	2	0.9277	1	6
Percent Above	0.3230	0.2812	0.2894	0	1
Log(Original Amount)	18.4256	18.7762	1.6264	11.5129	22.1063
Firm Characteristics					
Log(Total Asset)	7.6011	7.3594	1.8685	2.8233	11.5513
EBITDA	680.8151	55.6090	1920.55	-2439.94	10489
Leverage	0.9945	0.8843	0.4575	0.2893	4.8787
Debt Ratio	19.0926	0.4639	518.8207	0.0436	19455.2
Book Value per Share	1687.89	1.6962	46726.03	-875083	255000
Asset Tangibility	0.3310	0.1344	0.5004	0.0000	5.1922
Quick Ratio	0.7196	0.5565	0.6549	0.0124	5.9174
Macroeconomic Variables					
Growth Rate	5.1017	5.8080	1.6635	-2.2237	7.6852
Three Months T-Bill Rate	4.2160	4.36	1.9791	0.05	8.11
Aggregated Default Rates	2.4307	2.3770	1.2968	0.3980	5.9340
Unemployment Rate	5.1183	4.7	0.9521	4	9.6

CRC Credit Research Centre





Default Rates
Growth Rates
Unemployment Rates
T Bill Rate
Recovery_Rates
Macroeconomic Variables
Recovery Rates
Year
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
0.60
0.58
0.56
0.54
0.52
0.50
0.48
0.46
0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
0.28
0.26
0.24
0.22
0.20
1980
1990
2000
2010
2020

4. Empirical Evidences

- Out-of-sample Predictions
- Implications on Portfolio Loss Distribution

- Performance Metrics: R-square (R^2), root mean squared errors (RMSE) and mean absolute errors (MAE).

$$R^2 = 1 - \frac{\sum_i (r_i - \hat{r}_i)^2}{\sum_i (r_i - \bar{r})^2}, \quad \bar{r} = \frac{1}{N}\sum_i r_i$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_i (r_i - \hat{r}_i)^2}$$

$$\text{MAE} = \frac{1}{n}\sum_i | r_i - \hat{r}_i |$$

- For the mixed effect models, we examine the random effects specified on three levels: obligor, seniority and default year.

- We use stratified sampling method to generate training and testing set.
- The strata are defined at obligor, seniority and default year levels to be consistent with random effect levels.
- Make sure any instrument in the testing set should have a corresponding instrument of the same group in the training set.

	Training Set		Testing Set	
Strata	**Obligor**	**Instrument**	**Obligor**	**Instrument**
Obligor	**398**	**1037**	**144**	**376**
Seniority	**352**	**991**	**196**	**422**
Default year	**356**	**1002**	**197**	**411**

CRC Credit Research Centre

Methods	Sampling Strata	Train			Test		
		R^2	RMSE	MAE	R^2	RMSE	MAE
Single Factor Model	obligor	0.8942	0.1256	0.0866	0.8667	0.1473	0.0981
	seniority	0.3617	0.3119	0.2548	0.2855	0.3320	0.2744
	default year	0.4130	0.3010	0.2452	0.3444	0.3134	0.2568
Two Factor Model	obligor	0.8943	0.1270	0.0867	0.8727	0.1391	0.0953
	seniority	0.3494	0.3149	0.2600	0.2923	0.3304	0.2737
	default year	0.3922	0.3037	0.2487	0.4104	0.3036	0.2459
Inflated Beta Mixed Effects	obligor	0.5994	0.2443	0.1859	0.6110	0.2519	0.1967
	seniority	0.3309	0.3206	0.2681	0.3722	0.3088	0.2646
	default year	0.3726	0.3086	0.2568	0.4064	0.3046	0.2516

CRC Credit Research Centre

THE UNIVERSITY OF EDINBURGH

Methods	Sampling Strata	Train			Test		
		R^2	RMSE	MAE	R^2	RMSE	MAE
Linear Regression	obligor	0.2918	0.3249	0.2725	0.4456	0.3004	0.2524
	seniority	0.3553	0.3141	0.2600	0.2884	0.3293	0.2714
	default year	0.3525	0.3162	0.2613	0.2992	0.3240	0.2670
Fractional Response Regression	obligor	0.3083	0.3211	0.2669	0.4636	0.2956	0.2469
	seniority	0.3789	0.3083	0.2490	0.3184	0.3223	0.2620
	default year	0.3749	0.3106	0.2521	0.3249	0.3181	0.2578
Inflated Beta Regression	obligor	0.3014	0.3225	0.2743	0.4390	0.3030	0.2620
	seniority	0.3312	0.3205	0.2678	0.3701	0.3093	0.2650
	default year	0.3629	0.3136	0.2614	0.3053	0.3227	0.2694

- Both single and two factor models with random effects at the obligor level give remarkable better performances than the other methods. In general the mixed effects models outperform the related fixed effect models

- Two factor models present better out-of-sample predictions compared with single factor model.

- Among the fixed effects regression models, fractional response regression shows slightly better performances compared with linear regression and inflated beta regression models.

Implications to the credit portfolio risk management

- Generating Loss Distribution: Monte Carlo simulation
  - Fit the two factor model with the obligor level random effect
  - Simulate the macroeconomic variables
  - Obtain the simulated recovery rate and LGD_i
  - Generate default indicator d_i by the related default rate of the instrument default year.
  - Set EAD=1, then calculate the loss rate at the m-iteration as

$$L_m = \frac{1}{N}\sum_{i=1}^{N} d_i LGD_i$$

  - Repeat the above procedure for M times (e.g. M=100000)

- Comparison of Loss Distributions of AIRB and FIRB.

  In FIRB:

  - the PD is estimate as the default rate of the corresponding default year.
  - LGD is specified following the Basel II, where
    - Senior unsecured bond: 0.45
    - Subordinated bond: 0.75
    - Senior secured bond: average LGD of senior secured bonds in our sample, which is 0.3708
- Portfolio is studied at both aggregated and segmented levels

Credit Research Centre

A description of LGD at aggregated and segmented levels portfolios

Levels	No.	Mean	Std
Senior Secured	332	0.3708	0.3688
Senior Unsecured	681	0.4900	0.3813
Subordinate	400	0.6927	0.3628
Aggregated	1413	0.5194	0.3915

- VaR (Value at Risk) and ES (Expected Shortfall)

$$\text{VaR}_q(L) = \min\{l \mid P(L > l) \le 1-q\}$$

$$\text{ES}_q = E(L \mid L > \text{VaR}_q)$$

where l is the smallest value such that the probability that the loss rate L exceeds it is at most 1-q.

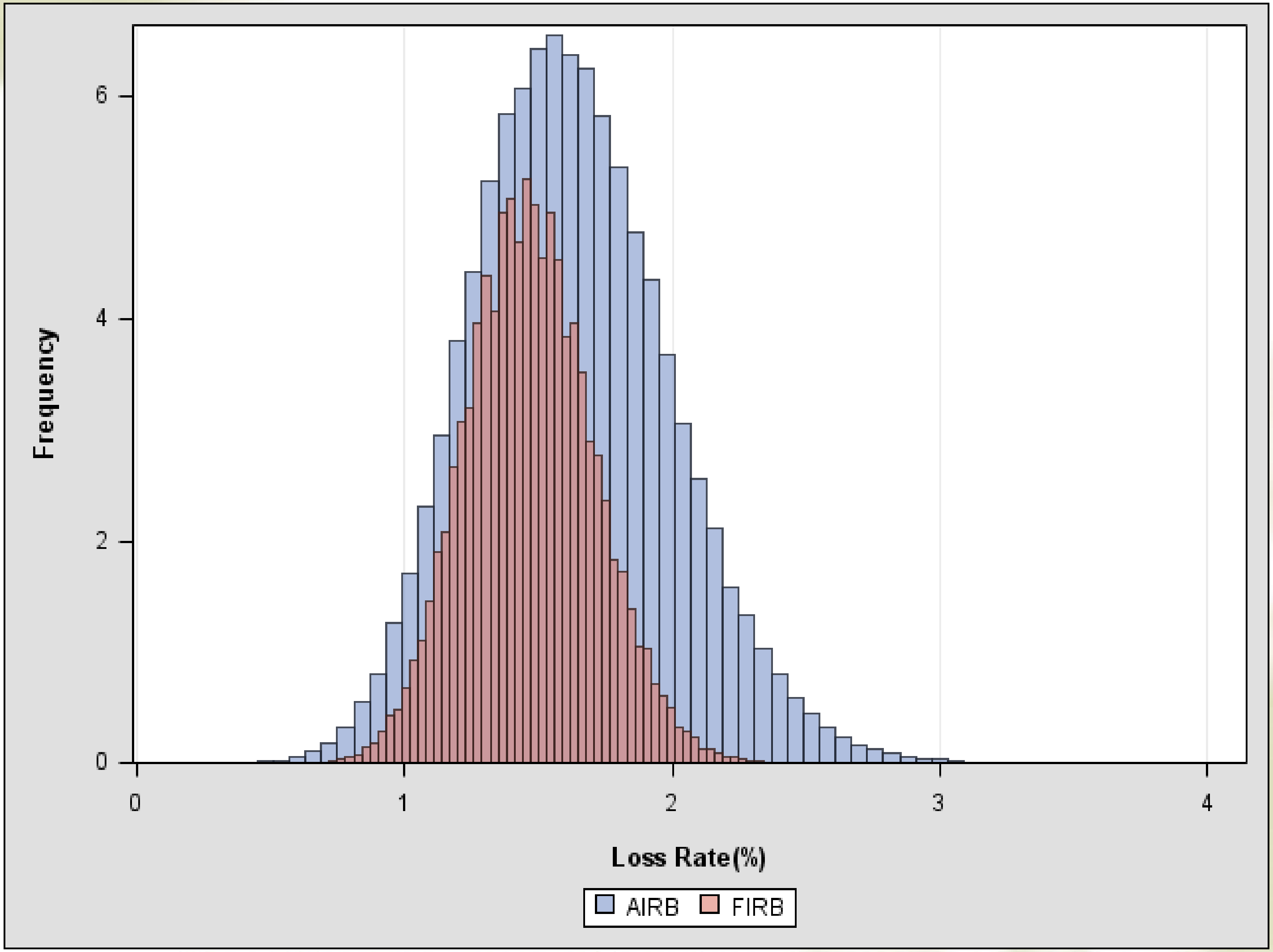
- EC (Economic Capital): the distance between the Expected Loss (EL) and ES.

$$\text{VaR}_q(L) = \hat{L}_q$$

$$\text{ES}_q = \frac{1}{N_q}\sum_{j=1}^{M} L_j I\left[L_j > \hat{L}_q\right]$$

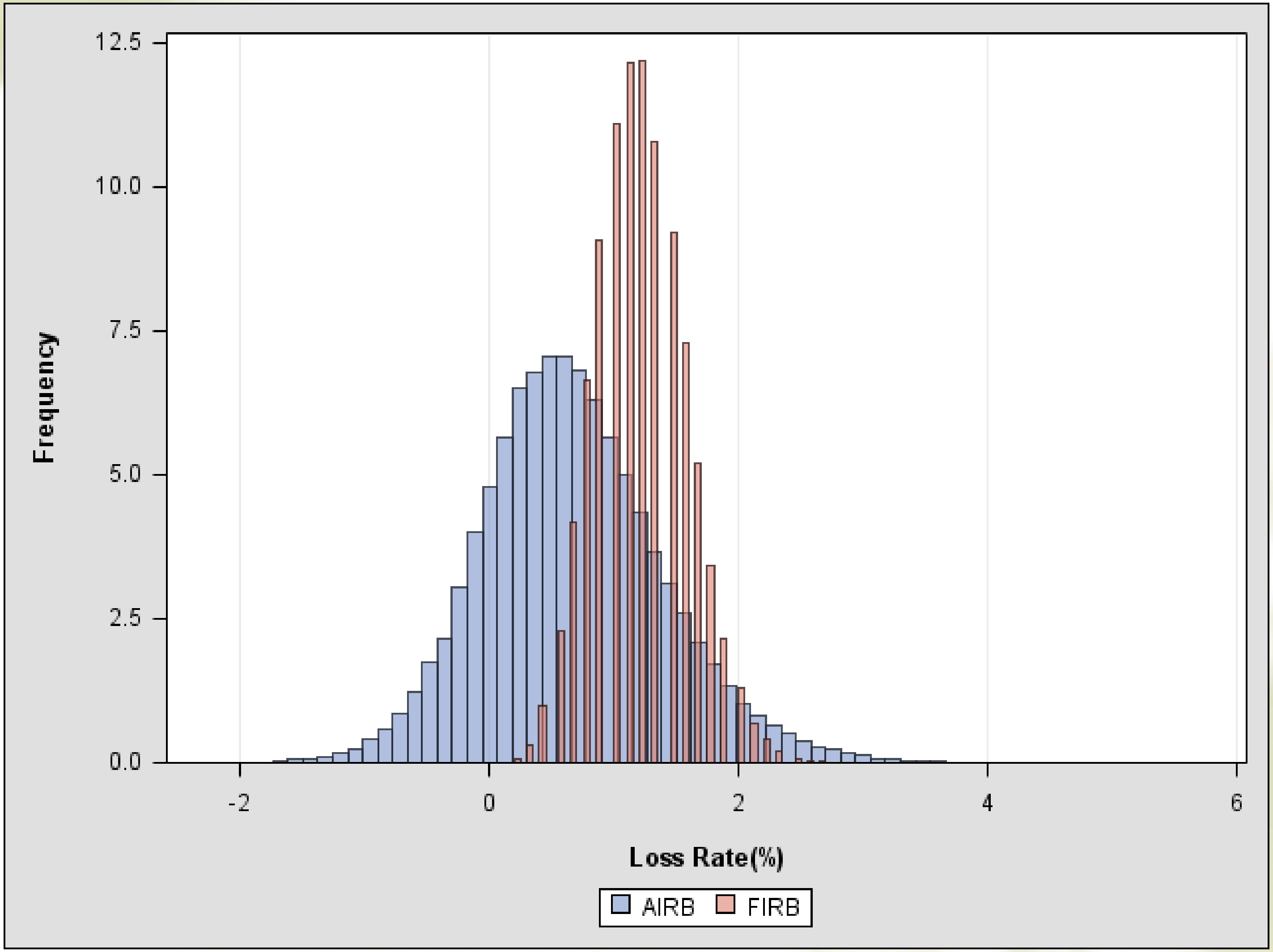
	VaR		ES		EL	EC	
q	0.05	0.01	0.05	0.01		0.05	0.01
				Senior Secured Bond			
FIRB	1.8986	2.1220	2.0235	2.2213	1.2178	0.8057	1.0035
AIRB	1.9394	2.6510	2.3749	3.0620	0.6572	1.7177	2.4049
				Senior Unsecured Bond			
FIRB	1.7841	1.9824	1.8899	2.0682	1.2785	0.6114	0.7897
AIRB	2.4941	2.9151	2.7550	3.1371	1.6485	1.1065	1.4886
				Subordinated Bond			
FIRB	4.5000	5.0625	4.8000	5.2972	3.2432	1.5568	2.0540
AIRB	3.7827	4.5048	4.2260	4.9013	2.3482	1.8778	2.5531
				Aggregated Portfolio			
FIRB	1.8698	2.0497	1.9788	2.1390	1.4691	0.5097	0.6699
AIRB	2.2618	2.5647	2.4498	2.7293	1.6215	0.8283	1.1078

A. Aggregated Portfolio



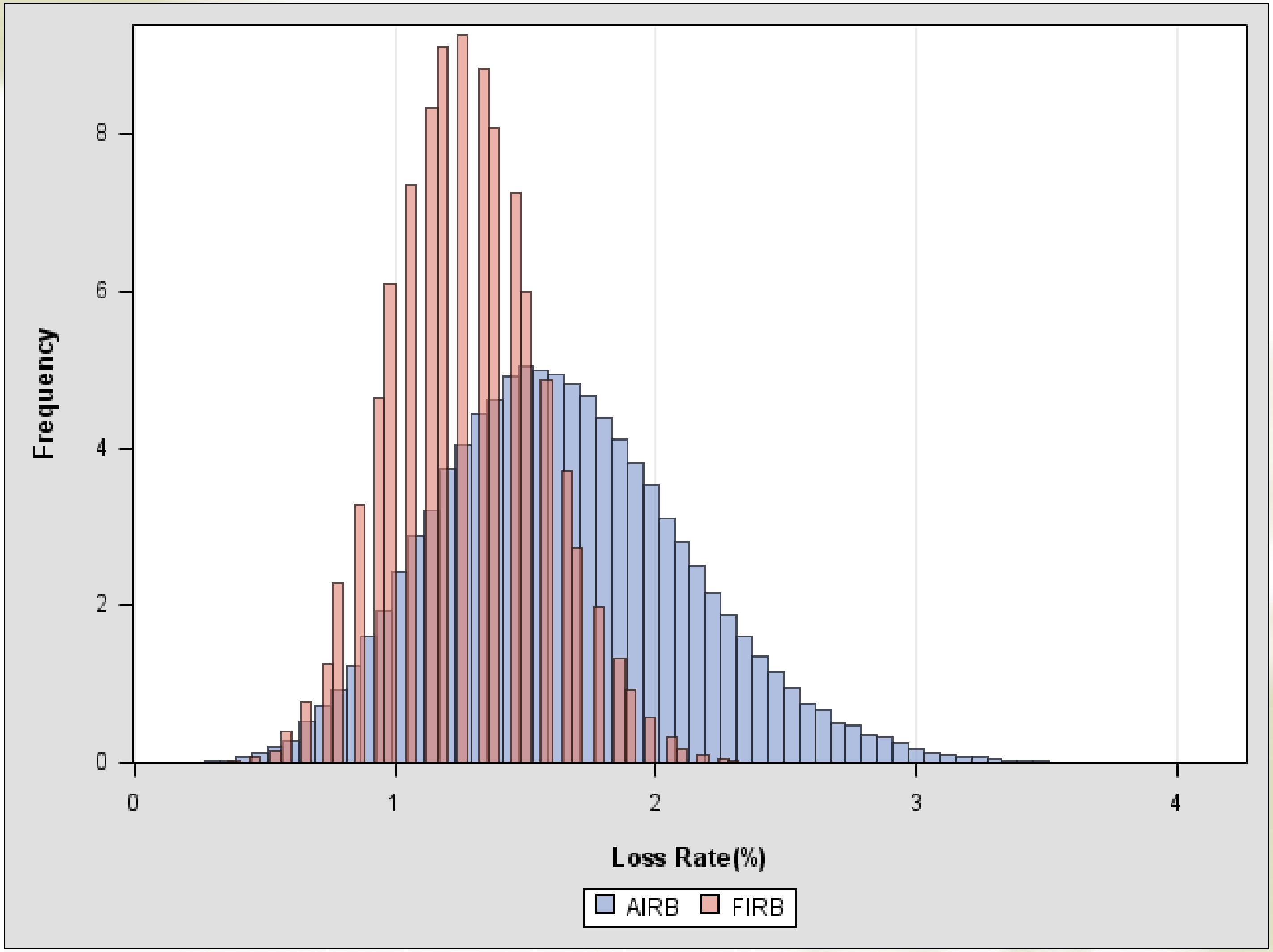
CRC Credit Research Centre

B. Senior Secured Bonds

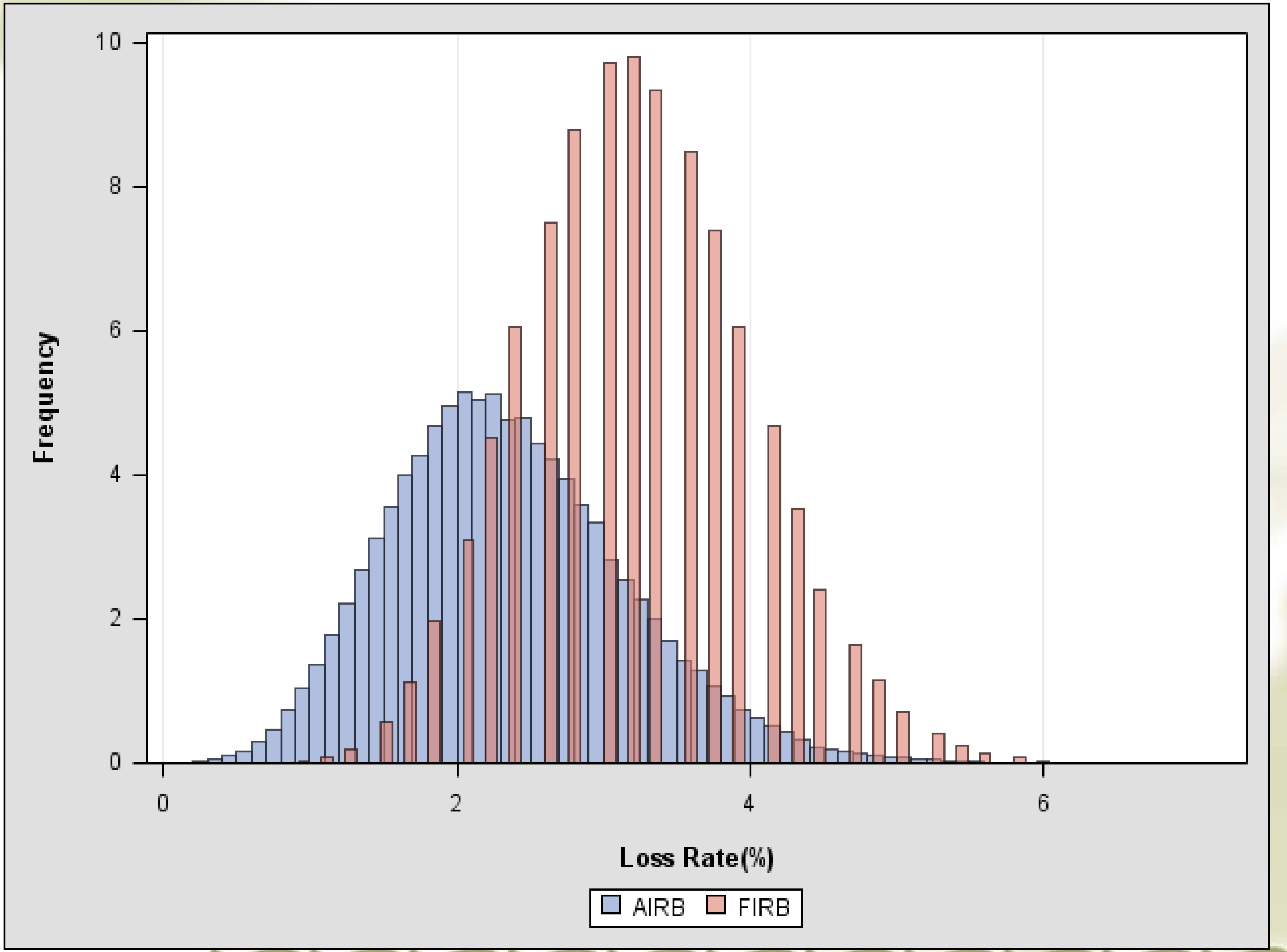


CRC Credit Research Centre

C. Senior Unsecured Bonds



D. Subordinated Bonds



CRC Credit Research Centre

- AIRB has a a much right skewed fatter tail loss rate distribution compared with the FIRB at the aggregated level.

- For both senior secured and unsecured bonds, the FIRB shows a higher VaR and ES than AIRB although FIRB gives a higher EL than AIRB on the senior bonds. The EC calculated based on AIRB is higher than FIRB for both segments.

- On subordinated bonds the FIRB approach obtains a higher VaR, ES and EL. But the EC calculated at both 0.05 and 0.01 confidence levels by AIRB is still higher than the FIRB.

- In summary, under the AIRB approach bank should allocate more buffer capital to hedge the unexpected loss of credit portfolios at both aggregated and segmented levels.

5. Conclusions and Further Research

- We investigate applying random effect at obligor, seniority and default year levels, and find that the obligor level linear mixed effects models give the best predictions.

- We propose a two factor model to better explain the unobservable heterogeneity and it shows better out-of-sample predictive abilities in corporate bonds LGD modeling.

- We generalize the inflated beta regression by inserting random effects and it presents better predictions compared with inflated beta regression, but not comparable to the linear mixed effects models.

- We examine the portfolio loss distribution at both aggregated and segmented levels. The simulation results indicate that the FIRB given in Basel II may underestimate the LGD of senior unsecured bonds and overestimate the subordinated bonds, and underestimate portfolio risk at the aggregated level.

CRC Credit Research Centre

Future Research

- Investigate the correlation between PD and LGD by joint model
- Study the retail loans' LGD
- Other distribution to simulate LGD

Thank you !

Questions please