Mixture model for EAD using GAMLSS framework

Suttisak Wattanawongwan

University of Southampton

August 30, 2019

Introduction

Figure 1: Global Financial Crisis in 2007-2008. CREDIT: Shutterstock/Randall Vermillion.

In the Literature

In order to model the EAD for credit cards and other forms of revolving credit, the Basel regulation proposes estimating the Credit Conversion Factor (CCF), i.e. the proportion of the undrawn amount that will be drawn at default time.

Even though the CCF models have been widely used in an EAD forecasting, several drawbacks have been inspected, e.g. CCF is highly bimodal and unstable due to the contracting denominator.

Alternatively, more recent work has suggested it may be beneficial to predict the EAD directly, i.e. modeling the balance as a function of a series of risk drivers.

Proposed model

In this work, we propose a novel approach combining two ideas proposed in the literature and test its effectiveness using a large dataset of credit card defaults not previously used in the EAD literature.

Proposed model: Mixture Assumption

We first conjecture that the level of EAD as well as the risk drivers of its mean and dispersion parameters could be substantially different between the borrowers whose Balance has **EVER** touched Limit (i.e. "max out" their cards) prior to default and who do not. We then implement a mixture model conditioning on these two respective scenarios (Leow and Crook, 2016).

Proposed model: Mixture Assumption (Continued)

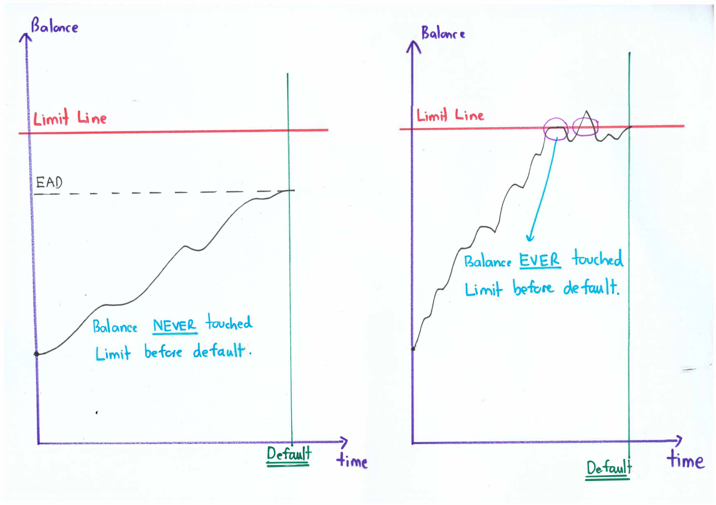

Figure 2: Two different scenarios: (left) NEVER Maxing out card and (right) EVER Maxing out card.

Proposed model: Mixture Assumption (Continued)

We propose a mixture of three models for the estimation of EAD:

$$E(\text{EAD}_i) = P(S_i = 1) \times E(\text{EAD}_i | S_i = 1) + P(S_i = 0) \times E(\text{EAD}_i | S_i = 0).$$

- A model for the estimation of the probability that Balance has **EVER** touched Limit ("maxing out" cards) over the observation period (i.e. between the start of the cohort period and twelve months later), $[P(S_i = 1)]$.
- A model for the estimation of EAD, given Balance has **NEVER** touched Limit $[E(EAD_i | S_i = 0)]$.
- A model for the estimation of EAD, given Balance has **EVER** touched the Limit $[E(EAD_i | S_i = 1)]$.

Proposed model: GAMLSS

The second idea is from Tong et al. (2016) who established a direct EAD model by using a regression GAMLSS framework.

The GAMLSS framework allows the parameters of a response distribution to be modeled as a function of explanatory variables either parametrically or non-parametrically.

GAMLSS model with the response following the distribution D and parameters $\boldsymbol{\theta_k}, k = 1, 2, 3, 4$ could be defined as:

$$\mathbf{Y} \overset{\text{ind}}{\sim} D(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\theta_3}, \boldsymbol{\theta_4})$$

$$g^k(\boldsymbol{\theta_k}) = \mathbf{X}^k\beta^k + s_1^k(x_1^k) + \cdots + s_{J(k)}^k(x_{J(k)}^k),$$

where g is the appropriate link function, $\mathbf{X}$ and x represent predictors, and smooth functions $s()$. The parameters to be estimated are β and $s(\cdot)$.

Hence, different values of predictors contribute to different values of location, scale, and shape parameters.

Proposed model: GAMLSS (Continued)

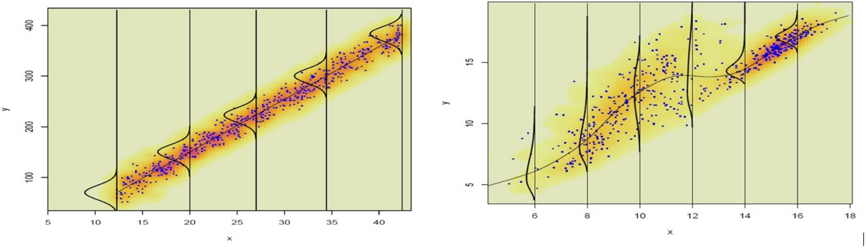

Figure 3: How the fitted conditional distribution of the response y changes for different values of the predictor x in (left) Linear model and (right) GAMLSS.
Source: Stasinopoulos et al. (2017).

Advantages of proposed model

- GAMLSS potentially allows us to produce a much more flexible fitted distribution to EAD dataset. This is important for realized EAD data since its characteristic could be inconsistently right-skewed, heavy tailed, unpredictable, and different from dataset to dataset.

- The ability to model the dispersion of EAD as a function of explanatory variables would be useful for risk controlling. We need to make an EAD estimate more conservative when dealing with a large variability risk of an EAD estimation.

- We can detect if there any differences in risk drivers of an EAD when the balance is likely to touch or not touch the limit before default.

- We can understand how likely borrowers will max out their card, and what covariates are driving it.

Data Description

Extracted from monthly observations of credit card in the Retail section from January 2002 to May 2007 with 74,000 defaulted accounts.

A set of behavioural variables has been collected, for example, **Balance**, **Limit**, and **Time to Default**.

Macroeconomic variables are also considered including Unemployment rate, Interest rate, GDP, and CPI.

Data Description (Continued)

Define an **EAD** as an **outstanding balance at default time** that borrowers owe to banks excluding any interest and additional fees.

Account goes into default state when borrowers:

- could not make the minimum payment for three months or more;
- are declared bankruptcy or charged-off.

Proposed model [Model for $P(S_i = 1)$]

The model for the Limit-touching probability $[P(S_i = 1)]$ is:

$$\log\left(\frac{P(S_i = 1)}{1 - P(S_i = 1)}\right) = \alpha_1 Y_{i,t} + \alpha_2 Z_t + \text{smoothing terms}, \tag{1}$$

- where α_1 and α_2 are unknown vectors of parameters to be estimated and $Y_{i,t}$ are **behavioural variables** for an account i at time t; and Z_t are **macroeconomic variables** at time t.

Proposed model [Models for $E(EAD_i|S_i = 0)$ and $E(EAD_i|S_i = 1)$]

There is a significant number of Zero-EADs in the dataset. We treat them as a special case and model separately from the other non-zero EAD by including the zero probability parameter, Nu [$\nu_i = P(\text{EAD}_i = 0)$].

Proposed model [Models for $E(EAD_i|S_i = 0)$ and $E(EAD_i|S_i = 1)$] (Continued)

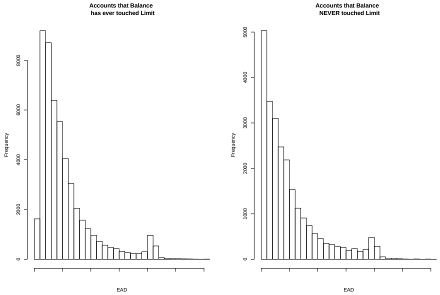

Figure 4: Non-zero EAD histograms for the accounts that Balance has **EVER** touched (Left)/ **NEVER** touched (Right) Limit.

This figure shows histograms of non-zero EAD in two limit-touching scenarios. They imply that we should select a positively skewed distribution. By comparing the performances (AIC/BIC and MAE/RMSE criterion) through validating dataset with full explanatory variables, Gamma distribution is selected for the response EAD.

Proposed model [Models for $E(EAD_i|S_i = 0)$ and $E(EAD_i|S_i = 1)$] (Continued)

Therefore, we assume that an EAD, in both scenarios, follows a mixed discrete-continuous Zero-adjusted Gamma (ZAGA) distribution.

$$\text{ZAGA}(\text{EAD}_i) = \begin{cases} \nu_i & \text{if } (\text{EAD}_i) = 0 \\ (1-\nu_i)\,\text{Gamma}(\text{EAD}_i|\mu_i, \sigma_i) & \text{if } (\text{EAD}_i) > 0, \end{cases} \tag{2}$$

for $0 \leq \text{EAD}_i < \infty$, where $0 < \nu_i < 1$, $\mu_i > 0$, and $\sigma_i > 0$, with

$$\log(\mu_i) = \gamma_1^{\mu} Y_{i,t} + \gamma_2^{\mu} Z_t + \text{smoothing terms};$$

$$\log(\sigma_i) = \gamma_1^{\sigma} Y_{i,t} + \gamma_2^{\sigma} Z_t; \quad \text{logit}(\nu_i) = \gamma_1^{\nu} Y_{i,t} + \gamma_2^{\nu} Z_t,$$

where

$$E(\text{EAD}_i) = (1-\nu_i)\mu_i. \tag{3}$$

Explanatory Variable selection

Stepwise methods with AIC/BIC are performed for selecting the best set of explanatory variables in the Limit-touching probability and EAD models.

The criterion used in $P(S_i = 1)$ model are Pearson goodness of fit statistic from the Hosmer-Lemeshow test (calibration power), AUROC (discrimination power), and residual plots (model adequacy).

The criterion used in EAD model are Pearson correlation (discrimination power), MAE, Normalized MAE, RMSE, Normalized RMSE (calibration power), and residual plots (model adequacy). A normalized version ($\frac{\text{EAD}}{\text{Limit}}$) investigates the performance when the percentage of current limit at default time is of interest.

Results: Risk Drivers for Limit-touching Probability model

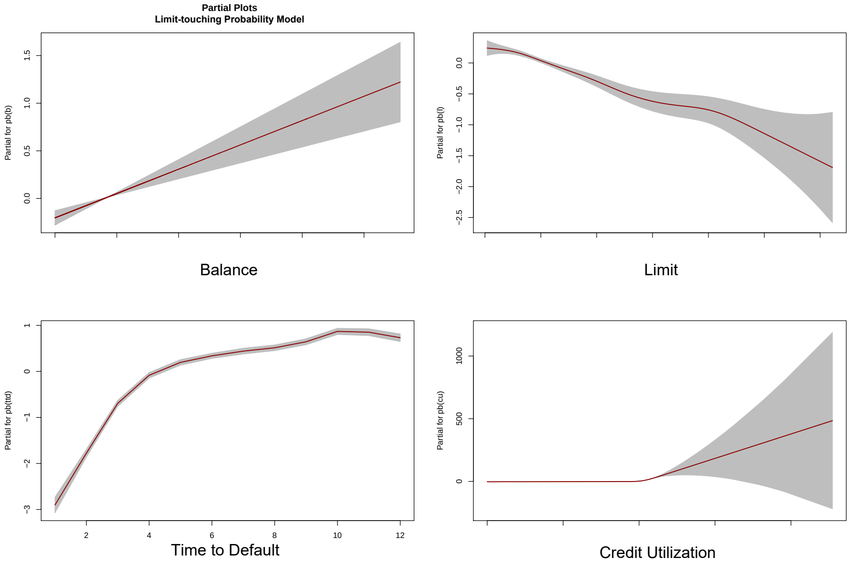

Figure 5: Partial effect plots on logit scale for Limit-touching Probability model

Results: Risk Drivers for EAD Mean [**NEVER** touched]

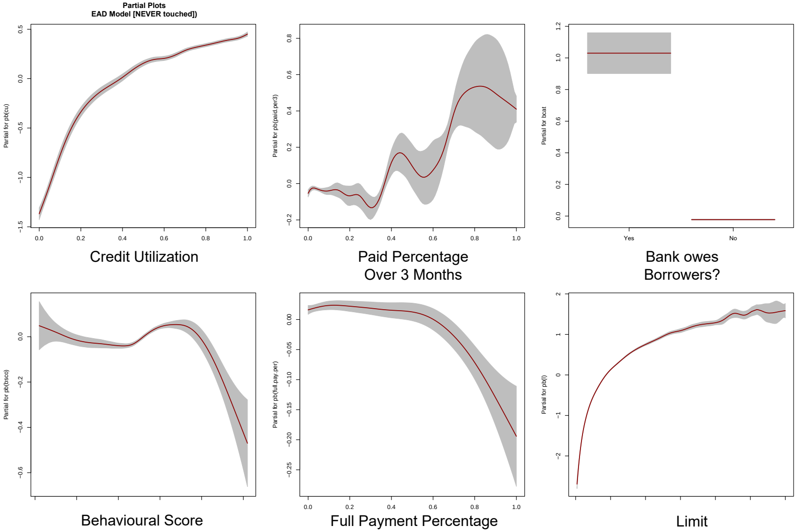

Figure 6: Partial effect plots on log scale for a mean (μ) parameter of EAD of the accounts with a balance **NEVER** touched a limit

Results: Risk Drivers for EAD Mean [**EVER** touched]

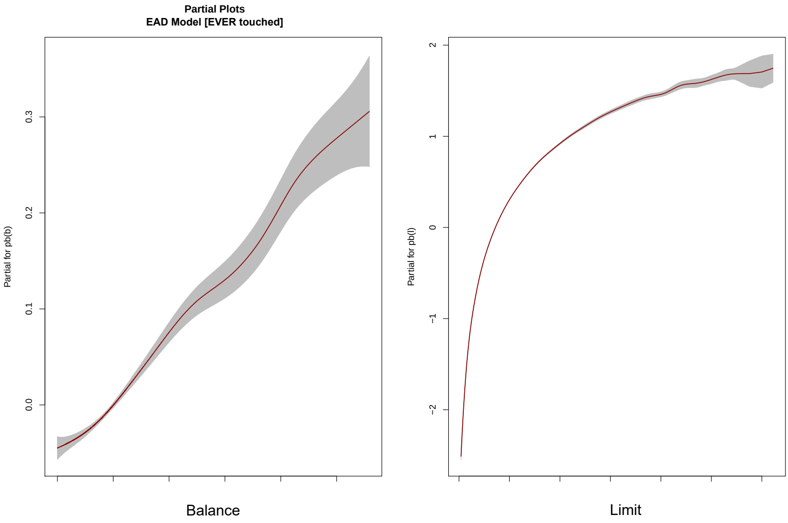

Figure 7: Partial effect plots on log scale for a mean (μ) parameter of EAD of the accounts with a balance **EVER** touched a limit

Results: Risk Drivers for EAD Parameters

Mean						
EAD model: balance **Never** touched limit	**Limit** (+)	**Average paid percentage over 3 months** (+)	**Bank owes to** (+)	Behavioural score (-)	Credit utilization (+)	Full payment percentage (-)
EAD model: balance **Ever** touched limit	**Limit** (+)				**Balance** (+)	

Table 1: A set of strongly significant predictors for EAD mean parameter of EAD model

Sigma									
EAD model: balance **Never** touched limit	**Balance** (-)	**Time to default** (+)	**Limit** (+)	Negative balance (+)	Special behavioural score (+)	Credit utilization (-)	Average paid percentage over 3 months (+)	Number of In.Arrears over 9 months (+)	Age (-)
EAD model: balance **Ever** touched limit	**Balance** (-)	**Time to default** (+)					Absolute change in balance over 9 months (+)		

Table 2: A set of strongly significant predictors for EAD dispersion parameter of EAD model

Nu										
EAD model: balance **Never** touched limit	Interest Rate (-)	Time to default (-)	Limit (+)	Absolute change in balance over 9 months (-)	Credit utilization (-)	Behavioural score (+)	**Average paid percentage over 3 months** (+)	Number of In.Arrears over 3 months (-)	Age (+)	Full payment Percentage (+)
EAD model: balance **Ever** touched limit							**Average paid percentage over 3 months** (+)			

Table 3: A set of strongly significant predictors for EAD Nu parameter of EAD model

Results: Risk Drivers for EAD Parameters

Lastly, by separating two Limit-touching scenarios, we conclude that the number of risk drivers of EAD parameters (mean, dispersion, and zero probability) for those borrowers whose balance are not likely to touch limit are far greater than another ones whose balance tends to hit its limit.

This demonstrates the need of more complicated model when estimating an EAD of borrowers who are not likely to hit the limit.

Benchmark Models

We call our proposed model "GAMLSS.Mix" since we apply a GAMLSS framework with the idea of mixture models. In order to evaluate the predicting performance, we build another three benchmark models.

"GAMLSS" = Unconditional on mixture idea + GAMLSS framework.

"OLS.Mix" = Conditional on mixture idea + OLS and Linear logistic framework.

"OLS" = Unconditional on mixture idea + OLS and Linear logistic framework.

Results: Model Predicting Performances

We compare the discrimination and calibration performance in terms of predictive power (comparing the actual EAD with the predicted $E(EAD_i)$) by using out-of-sample test dataset with the proposed and benchmark models.

	Correlation	RMSE	MAE	Norm.RMSE	Norm.MAE
Gamlss.Mix	0.926	18886.6439	8748.2309	0.2956	0.163
Gamlss	0.925	18569.027	8603.0405	0.2899	0.1598
Ols.Mix	0.9276	19115.6779	10236.3901	0.3685	0.2374
Ols	0.9268	18731.7422	10392.8326	0.3741	0.2475

Table 4: Performance measurements assessed with a test dataset

Conclusion

We construct the statistical model for estimating EAD by implementing GAMLSS framework and the idea of Limit-touching scenarios.

We understand the risk drivers for the mean, dispersion, and zero probability of EAD, and on the likelihood of a balance touching a limit. This provides additional information on EAD rather than focusing only on its mean level.

We realized the difficulty in estimating EAD for borrowers who are not likely to hit the limit.

There is a very important performance gain by using the GAMLSS approach and modeling the non-linearity in the data. However, adding the mixture component does not further improve that performance (although it provides some extra insights).

References

Leow, M. and Crook, J. (2016). A new Mixture model for the estimation of credit card Exposure at Default. *European Journal of Operational Research*, 249(2), pp.487-497.

Tong, E., Mues, C., Brown, I. and Thomas, L. (2016). Exposure at default models with and without the credit conversion factor. *European Journal of Operational Research*, 252(3), pp.910-920.

Thank You!!