Pioneer Journal of Theoretical and Applied Statistics
Volume 6, Number 1, 2013, Pages 43-51
This paper is available online at http://www.pspchv.com/content_PJTAS.html

EAD MODELING FOR CREDIT CARD

VICTOR GUMBO

Department of Finance
National University of Science and Technology
Zimbabwe

Abstract

The estimation of Basel II/III risk parameters (PD, LGD, EAD, M) is an important task in banking and other credit providers. These parameters are used on one hand as inputs to credit portfolio models, and on the other hand, to compute risk-weighted assets, regulatory, economic and other capital. EAD modeling for the credit card portfolio presents some challenges driven by the characteristics of the portfolio. We seek to demonstrate some practical techniques for the estimation of EAD for Credit Card, mostly based on the work by [1].

1. Introduction

In estimating the Exposure at Default (EAD) for a non-defaulted facility *f*, with an explicit credit limit, two methods are used to link the estimated EAD with the limit [1]:

• In one class, estimates of the EAD are based on a suitable Credit Conversion Factor (CCF) for the total limit of the facility,

Received April 28, 2013

2010 Mathematics Subject Classification: 97K80.

Keywords and phrases: credit card, exposure at default, loan equivalent, credit conversion factor.

$$\text{EAD}(f) = \text{CCF}(f) * \text{Limit}(f). \tag{1}$$

• In the other class, estimates of the EAD are based on another factor, the Loan Equivalent (LEQ), applied to the undrawn part of the limit,

$$\text{EAD}(f) = \text{Current Exposure}(f) + \text{LEQ}(f) * \text{Undrawn Amount}(f). \tag{2}$$

In the above, CCF is defined as the proportional change in the drawn amount at default while LEQ is the percentage of unutilized commitments at default. It is generally understood that if a borrower's creditworthiness deteriorates, the bank can cut the borrower's credit lines and the borrower will seek alternative sources. Simultaneously, if covenants permit, the bank will protect itself from further exposure by cutting off unused commitments. Effectively, the LEQ measures the outcome of the race between the bank and the borrower with regard to the draw-down of unused commitments in adverse circumstances [2].

Given a set of defaulted facilities, there exist several frequently employed approaches (cohort approach, fixed and variable time horizon) to obtain realized conversion factors or other statistics that can be used, in addition with other information, to obtain estimates for the EAD of non-defaulted facilities. All these approaches are based on observations of defaulted facilities at specific reference dates previous to the default date with only the rule for selecting these reference dates differing between approaches. For the model build the cohort approach was chosen. EAD modeling for the credit card portfolio presents some challenges driven by the characteristics of the portfolio. Obligors can either withdraw below or over the prescribed limit, before or after default. This behaviour presents modeling challenges. We seek to present some practical techniques for the estimation of EAD for Credit Card.

2. The Modeling Approaches

This research was based on Credit Card data from one of South Africa's big four banks (name withheld). The observation period for the model build was from January 2006 to December 2008. The following default definition consistent with the Basel reference definition was used:

• accounts which are 90 days in excess (4 cycles),

• accounts in legal or

• accounts written off.

Multiple default events per account across cohorts were considered as individual default events. Where multiple defaults per account occurred within the same cohort, only the first default event was considered.

First, the observation period was divided into 1 year intervals (cohorts). Second, all observed defaulted facilities were grouped into cohorts according to the interval that included their default dates. Third, in order to compute a realized conversion factor (CF) associated with each facility, the starting point of the time interval that contained its default date was used as the reference date. The model was developed using linear regression to predict LEQ and CCF. The following segmentation scheme was applied:

Model 1. Credit card facilities where the drawn exposure was less than the agreed limit at reference date: $(E(rt) < L(t)$ and $L(t) > 0)$.

The motivation for this approach was that one minus percentage usage $(1 - e(rt))$ at reference date has been shown to limit the variability evidenced in realized LEQ factors [1]. Therefore we can conclude that the methodology of associating usage as an explanatory variable in estimating EAD is appropriate.

Observations were obtained from the set of all defaulted overdraft facilities satisfying $E(rt) < L(t)$ and $L(t) > 0$ and where the committed limit of credit is known to the borrower and is given by $L(t)$.

For the structure of the reference data set (RDS): for simplicity, only a basic set of risk drivers was constructed:

$$RDS = \{l(g);\ L(t);\ E(rt);\ E(dt);\ dt;\ rt\},$$

where

$l(g)$ is identifier of facility,

$L(t)$ is the limit at time t,

$E(rt)$ is the exposure at reference time rt,

$E(dt)$ is the exposure at default time dt.

Generally it was noted that the Credit Card data for $E(rt) < L(t)$ was very noisy. Consequently, it was decided to do the following:

1. Remove observations with very low values of $L(t) - E(rt)$ from the RDS because their LEQ values are not informative (in any event the degree of adjustment of the regression line is very low as most of the points, those with $1 - e(rt)$ closer to zero, have little influence on the result of the regression model because of the constraint that there is no independent term).

2. Eliminate from the RDS those anomalous observations with large LEQ factors. It is easy to understand that points associated with large values of $1 - e(rt)$ constitute influential observations and that changes in these points affect the result of the regression and therefore the LEQ estimate. As a result, observations with $LEQ > 1$ were removed from the RDS.

3. Keep the large number of observations with large values of $1 - e(rt)$ which helps ensure the stability of model results.

4. Only data satisfying $0 < eadt < 1$ was considered. Here $eadt = (E(dt) - E(rt))/L(t)$.

5. After data cleansing, a total of 20 874 qualifying observations were split in the ratio of $70 : 30 =$ development sample: hold out sample for Model 1.

Model 2. Facilities where the drawn exposure was greater than or equal to the agreed limit at reference date: $(E(rt) \geq L(t)$ and $L(t) > 0)$.

The scatter plot below of one-minus percentage usage $(1 - e(rt))$ against LEQ_i indicates that usage does not limit the variability evidenced in realized LEQ factors. Therefore, we can conclude that the methodology of associating usage as an explanatory variable in estimating EAD is not appropriate.

Given this and the fact that all accounts are in essence overdrawn at reference date (resulting in negative LEQs), it was instead decided to use the momentum approach whereby a CCF is estimated in place of a LEQ estimate (This approach is expressly allowed for by the FSA).

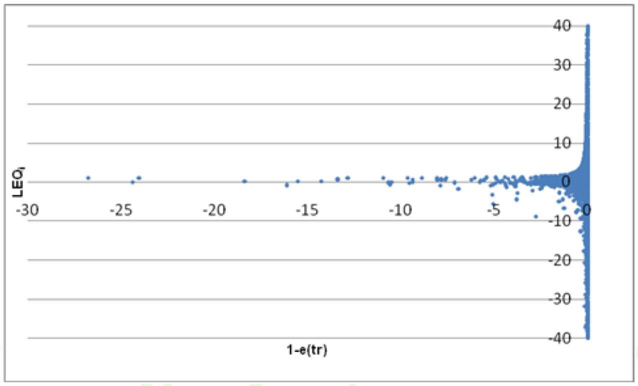

Figure 1. Scatter plot of one-minus percentage usage $(1 - e(rt))$ against LEQ_i.

Observations were obtained from the set of all defaulted credit card facilities satisfying $E(rt) \geq L(t)$ and $L(t) > 0,$ where the product types were all credit card lines with a committed limit of credit that is known to the borrower and is given by $L(t)$.

Similarly, the RDS data set was split 70/30 between development and holdout. It does not include all the internal defaults which took place during the observation period because several filters had been applied previously. For example, it was necessary to identify outliers and influential observations and afterwards to make decisions on which observations had to be removed from the RDS. As a result, the following decision was taken:

• Remove observations with very high limits due to the sparcity of data in this region. This results in a more stable model.

• Eliminate from the RDS those anomalous observations with large $E(dt)$ values, relative to $E(rt),$ thereby adding additional conservatism to the model.

After data cleansing, a total of 18 475 qualifying observations were split in the ratio of $70 : 30 =$ development sample: hold out sample for Model 2.

3. Data Analysis

Model 1. Credit Card $(L(t) > E(rt)$ and $L(t) > 0)$.

In estimating the LEQ factor, it was decided to use the increase of the exposure as a percentage of the observed limit (focus on the percentage increase in usage from *rt* to *dt* and percent usage at the reference date) using the formula

$$ead_i - e(rt) = LEQ[1 - e(rt)]. \tag{3}$$

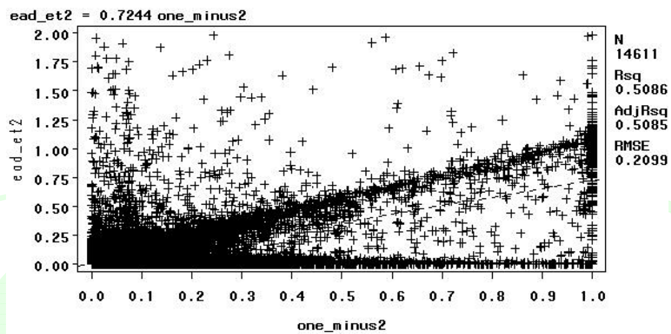

Figure 2. Virgin under-limit fitted to last 3 cohorts.

The resulting regression equation is

$$Ead_i - e(rt) = 0.7244 \times (1 - \text{percentage usage}).$$

Therefore, using this approach, the observable amounts to be explained are $eadi - e(rt) = (EADi - Ei)/Li$ and the explanatory values are $1 - e(rt) = LEQ(RDi) \times (Li - Ei)/Li$. Thus, assuming that the LEQ is constant, this reduces to using regression without a constant whereby the regression estimator represents the slope of the regression line. The resulting regression gives an LEQ of 0.7244 with an *R*-square of 0.5086 and an adjusted *R*-square of 0.5085 and an error of 0.2099.

In this case the model takes the form

$$EAD(f) = \text{Current Exposure}(f) + 0.7244 * \text{Undrawn Amount}(f).$$

Accuracy of CCF estimate

On average, the above LEQ was found to lead to a model over-estimation of 8.60% on the training set and an over-estimation of 10.17% on the validation set. This was deemed too high. This over-estimation is supported by the *R*-square of only 0.5086. However, an LEQ of 0.6000 led to a better estimation of EAD. In this case, the EAD estimator of a facility *f* in normal status based on the fitted model is given by

$$EAD(f) = \text{Current Exposure}(f) + 0.6000 * \text{Undrawn Amount}(f).$$

Tables 1. Training and validation comparisons

Training set - Underlimit

	EAD	Def Bal	Over	% Over
2006	R 8,288,173	R 7,641,972	R 646,201	8.46%
2007	R 92,937,358	R 94,730,851	R –1,793,493	–1.89%
2008	R 86,955,029	R 85,222,447	R 1,732,582	2.03%

Validation set - Underlimit

	EAD	Def Bal	Over	% Over
2006	R 3, 132, 088	R 2, 888, 284	R 243, 804	8.44%
2007	R 39, 941, 740	R 40, 247, 603	R –305, 863	–0.76%
2008	R 35, 369, 761	R 34, 509, 285	R 860, 476	2.49%

Model 2. Credit Card $(E(rt) \geq L(t)$ and $L(t) > 0)$

In estimating the CCF, the limit at time *rt* is regressed onto the EAD at time *dt*. Therefore the following formula is used:

$$EAD(f) = CCF(f) * L(f). \tag{4}$$

Using this approach, the observable amounts to be explained are *EADi* and the explanatory values are *Li*. Thus, assuming that the CCF is constant, this reduces to using regression without a constant whereby the regression estimator represents the slope of the regression line. Figure 3 depicts the fitted regression line.

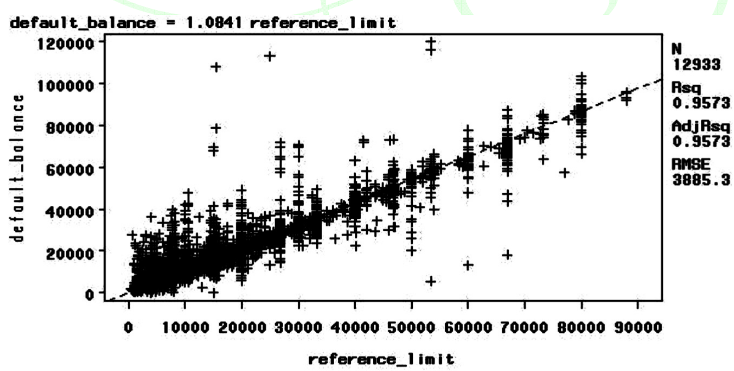

Figure 3. Over-limit development graph.

In this case the model takes the form:

$$EAD(f) = \overline{CCF}(f) \times L(f),$$

$$EAD(f) = 1.0841 \times L(f).$$

Accuracy of CCF estimate

On average, the above CCF was found to lead to a model over-estimation of 1.94% on the training set and an over-estimation of 1.29% on the validation set. The EAD estimator of a facility *f* in normal status based on the fitted model is given by

$$EAD(f) = \max(\text{drawn balance}; 1.0841 * \text{Limit}(f)).$$

Training set - Overlimit

	EAD	Def Bal	Over	% Over
2006	R 19, 215, 879	R 20, 161, 637	R –945, 758	–4.69%
2007	R 89, 042, 189	R 89, 047, 961	R –5, 772	–0.01%
2008	R 68, 706, 092	R 64, 430, 416	R 4, 275, 676	6.64%

Validation set - Overlimit

	EAD	Def Bal	Over	% Over
2006	R 8, 072, 368	R 8, 660, 396	R –588, 028	–6.79%
2007	R 38, 007, 316	R 37, 794, 601	R 212, 715	0.56%
2008	R 29, 358, 787	R 27, 305, 314	R 2, 053, 473	7.52%

4. Conclusion

We have demonstrated two approaches that can be used for Credit Card EAD modeling. There are three weaknesses when using LEQ (Model 1) as the basic input for estimation procedures. First, it is not defined when $L(t) = E(rt)$. This implies that it is not possible to estimate directly $EAD(f)$ based on the value of this statistic for facilities that at the current date exhibit percentage usage $e(rt)$ equal to one. Second, it is not stable when $L(t)$ is close to $E(rt)$. This implies that realized LEQ factors are not very informative when percentage usage is close to one. Finally, it does not take into account changes in the limit over time, this is only one of the causes that justify the existence of realized LEQ factors greater than one.

However, the approach has its benefits too. First, it takes into account drawn and undrawn amounts as explanatory variables in the EAD estimation procedure. This supports the regulatory requirement that the bank must use all relevant and material information in its derivation of EAD estimates.

For Model 2, experience shows that in general, drawn and undrawn limits have strong explanatory power for the EAD. For this reason, this method (with CCF = constant) does not seem to meet the requirement of using all the relevant information (because it does not take into account the drawn and undrawn amounts as explanatory variables in the EAD estimation procedure) for most types of facilities that occur in practice.

However, the method is used by banks to avoid the explicit use of realized negative LEQ factors, or for facilities for which the current usage has no predictive power on EADs. Second, the realized CCF is well defined even when $L(rt) = E(rt)$ and there are no instability issues.

References

[1] G. Moral, EAD Estimates for Facilities with Explicit Limits, Chapter X of the Basel II Risk Parameters, Bernd Engelmann and Robert Rauhmeier, Springer, Berlin, Heidelberg, 2006.

[2] M. Araten and M. Jacobs Jr., Loan equivalents for revolving credits and advised lines, The RMA Journal 2001.