Developing, Validating, and Monitoring a PD Model for Foundation IRB Approach

Nguyen, Van Phuong

2 October 2019

Developing, Validating, and Monitoring a PD Model for Foundation IRB Approach

Phuong Van Nguyen[*]

Abstract

Assuming that a given bank wants to comply with the Basel Accord requirements, in particular the Foundation IRB approach. Accordingly, it has to develop a PD model to predict the probability of default of its borrower within one year. Hence, this paper aims to present a simply empirical procedure for developing, validating, and monitoring a PD model.

Keywords: Basel Accord II, Loss Distribution, Expected Loss, Unexpected Loss, Risk Components, VaR, CAR, Model Validation and Monitoring, AUROC, CAP, PSI.

JEL Classification: G0, G10, G11

---

[*]Corresponding email: `phuong.nguyen@economics.uni-kiel.de`

1   Introduction

The most striking feature of the standardised approach is easy and simple for the bank to compute its Risk-Weighted Asset (RWA). However, its fundamental drawback is as follows. The bank will suffer from the problem of high capital burdens. But why? This is because there is the least differentiation between safer and risker assets when a bank applies the standardised approach. Therefore, from the perspective of the President and the Board of Management of a bank, they might not be happy. To overcome the problem of high capital burdens, the Basel Committee on Banking Supervision, or BCBS for short, has been encouraging the banking institutions to use either the Foundation or Advanced IRB approaches (see BCBS (2005a,b, 2006a,b)). In this project, we aim to comply with the foundation IRB one. Indeed, we present a simple empirical procedure for developing, estimating, and monitoring a PD model.

On the other hand, it is worth noting that the probability of default (PD) is one of four key risk components/inputs for the Risk Weight Function. For further understanding of this mathematical function, one might refer to the document about an explanatory note on the Basel-II IRB Risk Weight Function (see BCBS (2005a)) or our attempt to deriving this function (see Nguyen (2019)). Thus, the PD parameter must be computed in both the foundation and advanced approaches.

The rest of this paper is organized as follows. Section 2 introduces our dataset. Section 3 show the model estimation. Section 4 presents the procefure for model validation, whereas Section 5 monitors the model performance over time. Finally, section 6 gives conclusions.

2   Data Preparation

2.1   Data Collection

We collect data on credit card default of client from the UCI Machine Learning Repository[1]. A brief introduction to this dataset is as follows. The target feature is the credit card default of clients. Indeed, it is a binary variable. The 23 remaining attributes are explanatory variables related to age, sex, marriage, and other financial indicators. Table 1 presents the descriptive statistics of this raw data.

---

[1] One might refer to the link: `https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients`

Table 1: Descriptive statistics

	count	mean	std	min	25%	50%	75%	max
LIMIT_BAL	30000.0	167484.32	129747.66	10000.0	50000.00	140000.0	240000.00	1000000.0
SEX	30000.0	1.60	0.49	1.0	1.00	2.0	2.00	2.0
EDUCATION	30000.0	1.84	0.74	1.0	1.00	2.0	2.00	4.0
MARRIAGE	30000.0	1.56	0.52	1.0	1.00	2.0	2.00	3.0
AGE	30000.0	35.49	9.22	21.0	28.00	34.0	41.00	79.0
PAY_0	30000.0	-0.02	1.12	-2.0	-1.00	0.0	0.00	8.0
PAY_2	30000.0	-0.13	1.20	-2.0	-1.00	0.0	0.00	8.0
PAY_3	30000.0	-0.17	1.20	-2.0	-1.00	0.0	0.00	8.0
PAY_4	30000.0	-0.22	1.17	-2.0	-1.00	0.0	0.00	8.0
PAY_5	30000.0	-0.27	1.13	-2.0	-1.00	0.0	0.00	8.0
PAY_6	30000.0	-0.29	1.15	-2.0	-1.00	0.0	0.00	8.0
BILL_AMT1	30000.0	51223.33	73635.86	-165580.0	3558.75	22381.5	67091.00	964511.0
BILL_AMT2	30000.0	49179.08	71173.77	-69777.0	2984.75	21200.0	64006.25	983931.0
BILL_AMT3	30000.0	47013.15	69349.39	-157264.0	2666.25	20088.5	60164.75	1664089.0
BILL_AMT4	30000.0	43262.95	64332.86	-170000.0	2326.75	19052.0	54506.00	891586.0
BILL_AMT5	30000.0	40311.40	60797.16	-81334.0	1763.00	18104.5	50190.50	927171.0
BILL_AMT6	30000.0	38871.76	59554.11	-339603.0	1256.00	17071.0	49198.25	961664.0
PAY_AMT1	30000.0	5663.58	16563.28	0.0	1000.00	2100.0	5006.00	873552.0
PAY_AMT2	30000.0	5921.16	23040.87	0.0	833.00	2009.0	5000.00	1684259.0
PAY_AMT3	30000.0	5225.68	17606.96	0.0	390.00	1800.0	4505.00	896040.0
PAY_AMT4	30000.0	4826.08	15666.16	0.0	296.00	1500.0	4013.25	621000.0
PAY_AMT5	30000.0	4799.39	15278.31	0.0	252.50	1500.0	4031.50	426529.0
PAY_AMT6	30000.0	5215.50	17777.47	0.0	117.75	1500.0	4000.00	528666.0
default.payment.next.month	30000.0	0.22	0.42	0.0	0.00	0.0	0.00	1.0

2.2 Data cleaning and engineering

As mentioned before, data cleaning and validation are time-consuming. In this project, due to constraint space, we briefly introduce our task as follows.

- For cleaning data, we solve abnormal values in three categorical variables including education, marriage.

- Because of the different measurement units, we standardize the scale of the financial variables.

- We apply the Synthetic Minority Over-sampling Technique (SMOTE) algorithm to solve the highly imbalanced data. It is worth noting that the SMOTE is applied to the training dataset.

- For variable selection, we use several algorithms, such as the Pearson Correlation Matrix, Principle Component Analysis (PCA), even the Random Forest algorithm. For the first attempt, we use all 24 attributes. However, in practice, the variable selection would be strictly conducted. Otherwise, the performance of a Techcombank PD model can be reduced.

3 The PD Model estimation

3.1 The PD Model selection

Given that our data validation in compliance with the Basel Accord II, the next step is to choose an algorithm, which is best fits our validated dataset. According to the Basel Accord II, the banks are free to choose their own PD model. There are many algorithms. For example, one can refer to a wide range of the PD models in the document by Engelmann and Rauhmeier (2006). However, for simplicity, in this project, we compare the performance of four different algorithms, such as Logistic, Support Vector Classifier (SVC), K Nearest Neighbor (KNN), and Ridge Classifier. To reduce the problem of overfitting, we use the k-Fold Cross-Validation for each PD model. Indeed, we use 10 folds. On the other hand, according to the Basel Accord II (see (BCBS, 2005b)), we use the indicator of the Area Under Receiver Operating Characteristic (AUROC) curve to evaluate the performance of these four different PD models. The result is shown in Figure 1 below.

Figure 1: The Performances of Four PD Models
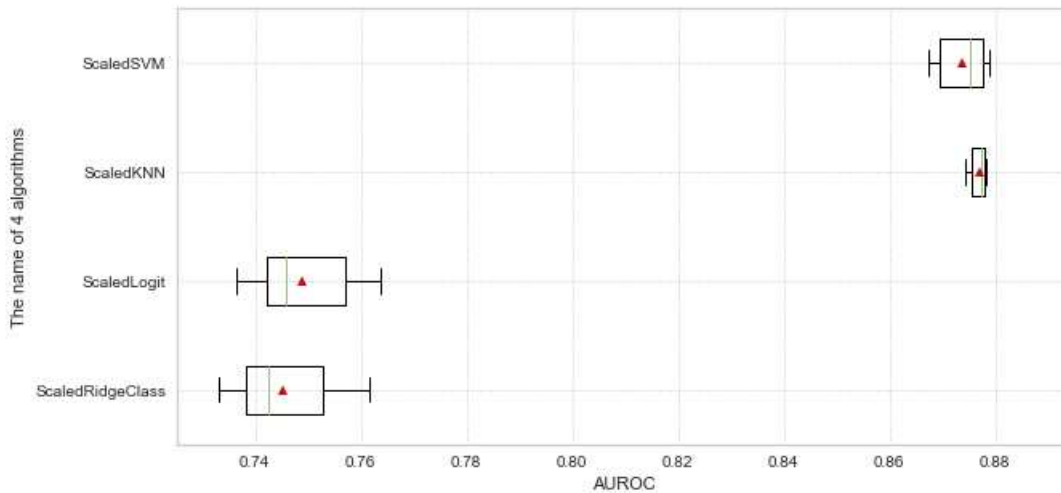
Based on Figure 1, we prefer the PD model using the KNN algorithm because of the most stable and highest AUROC. However, we choose the PD model using the Logistic algorithm as a potential PD one. This is because it seems that the commercial banks in Vietnam prefer this algorithm. Indeed, we do not really know the reason behind, even though we know a number of factors that can reject the application of this simple algorithm. For example, in this project, one can see clearly that the performance of the PD model using the Logistic algorithm is almost worst.

3.2 Estimating a PD model with the Logistic algorithm

It is worth noting that we split our dataset into two parts: training data and test data. Moreover, our splitting procedure is conducted randomly. In this section, we use the training data with 24000 observations to estimate the PD model. Afterward, the test data with 6000

observations are used to validate and monitor the performance of our estimated PD model. The estimation result using the maximum likelihood (ML) method is shown in Table 2 below. From the perspective of the Frequentist econometricians, the aspects concerning the P-value of each coefficient, R-squared, the assumption of a normal distribution, etc are typically examined. To what extent, these examinations are still important. However, one with the Baysian technique typically does not treat them as an important task. Thus, we skip these examinations.

Table 2: The estimation result

|	coef	std err	z	P> |z|	[0.025	0.975]
LIMIT_BAL	-0.0000	0.0000	-5.0104	0.0000	-0.0000	-0.0000
SEX	-0.1995	0.0298	-6.7012	0.0000	-0.2578	-0.1411
EDUCATION	-0.1273	0.0235	-5.4231	0.0000	-0.1733	-0.0813
MARRIAGE	-0.2664	0.0256	-10.4166	0.0000	-0.3165	-0.2163
AGE	-0.0004	0.0015	-0.2869	0.7742	-0.0034	0.0025
PAY_0	0.5859	0.0198	29.5385	0.0000	0.5470	0.6248
PAY_2	0.0813	0.0225	3.6105	0.0003	0.0372	0.1255
PAY_3	0.0703	0.0254	2.7698	0.0056	0.0206	0.1200
PAY_4	0.0362	0.0279	1.2979	0.1943	-0.0185	0.0909
PAY_5	0.0314	0.0298	1.0543	0.2918	-0.0270	0.0899
PAY_6	0.0070	0.0247	0.2833	0.7769	-0.0415	0.0555
BILL_AMT1	-0.0000	0.0000	-4.6866	0.0000	-0.0000	-0.0000
BILL_AMT2	0.0000	0.0000	1.1007	0.2710	-0.0000	0.0000
BILL_AMT3	0.0000	0.0000	2.1135	0.0346	0.0000	0.0000
BILL_AMT4	0.0000	0.0000	0.2639	0.7918	-0.0000	0.0000
BILL_AMT5	-0.0000	0.0000	-0.8891	0.3739	-0.0000	0.0000
BILL_AMT6	0.0000	0.0000	0.7890	0.4301	-0.0000	0.0000
PAY_AMT1	-0.0000	0.0000	-5.5754	0.0000	-0.0000	-0.0000
PAY_AMT2	-0.0000	0.0000	-4.8721	0.0000	-0.0000	-0.0000
PAY_AMT3	-0.0000	0.0000	-0.8671	0.3859	-0.0000	0.0000
PAY_AMT4	-0.0000	0.0000	-1.5444	0.1225	-0.0000	0.0000
PAY_AMT5	-0.0000	0.0000	-2.1420	0.0322	-0.0000	-0.0000
PAY_AMT6	-0.0000	0.0000	-1.4585	0.1447	-0.0000	0.0000

4 Model Validation

Many aspects of an estimated model should be examined and validated. For example, One can refer to the practical approach to validating a PD model for mortgage loans at a commercial bank in the Netherlands (see Medema et al. (2009)) or the German banking system (see Engelmann et al. (2003)). In addition, one can refer to the very detailed guide to the validation of Internal Rating System in the Basel Accord II (see BCBS (2005b)). However, for simplicity, this simple project only focuses on the field concerning the discriminate power of our estimated PD model.

According to the document concerning studies on validation of Internal Rating system published by the Basel Committee on Banking Supervision (see BCBS (2005b)), seven criteria are introduced to evaluate the discrimal power of a PD model. Basically, these criteria are linked with each other both statistically and mathematically. Thus, we do not attempt to compute all these seven criteria. Instead, we compute two fundamental indicators, such as AUROC and CAP or GINI. However, first, we compute the confusion matrix. This is because it is a baseline criterion. On the other hand, it is also worth noting that we use the test data with 6000 observations for the model validation.

4.1 Confusion Matrix

Figure 2 presents the confusion matrix. This matrix should be the first criterion since one can compute other important indicators, such as recall and precision. It is worth noting that the total of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) is identical to the number of observations concerning the dependent variable in our test dataset, here is 6000 observations. In general, the best model should produce no Type Error I and II. However, it is difficult to have such kind of perfect model. Moreover, one should understand the inverse trade-off between these two types of errors. From a banking perspective, we mainly focus on reducing Type Error II. This is because a small Type Error II implies a high accuracy of predicted PD, which is one of the key risk components in the IRB system (see BCBS (2005a,b, 2006a,b)). Based on these analysis, one can argue that the performance of our estimated PD model is bad because of a high number of both FN and FP. On the other hand, based on this confusion matrix, it is easy to compute other important indicators including Precision, Recall, and f1-score. However, this simple project we skip it.

Figure 2: The Confusion Matrix
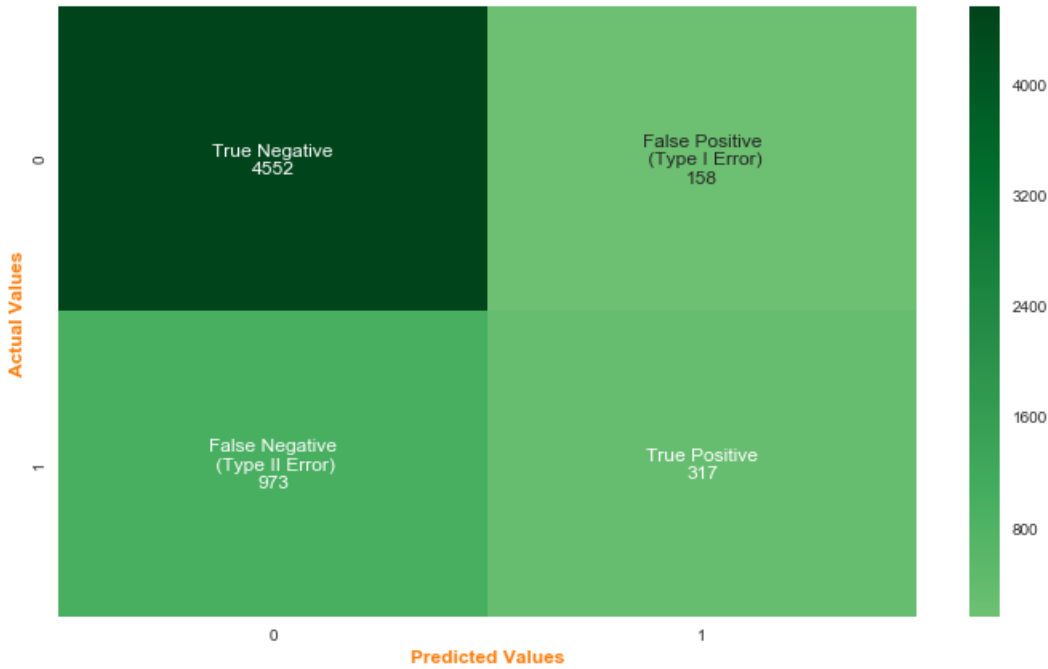
4.2 The Area Under Receiver Operating Characteristic (AUROC) curve

Figure 3 shows the AUROC. One of the fundamental criteria, which is recommended in the Basel Accord II (see BCBS (2005b)), AUROC is used to assess the discriminal power of the PD model. Accordingly, a higher AUROC is, a better PD model is. In our case, the AUROC of around 71 % implies relatively good performance. Thus, our PD model should be further developed or substituted by other algorithms.

Figure 3: The AUROC
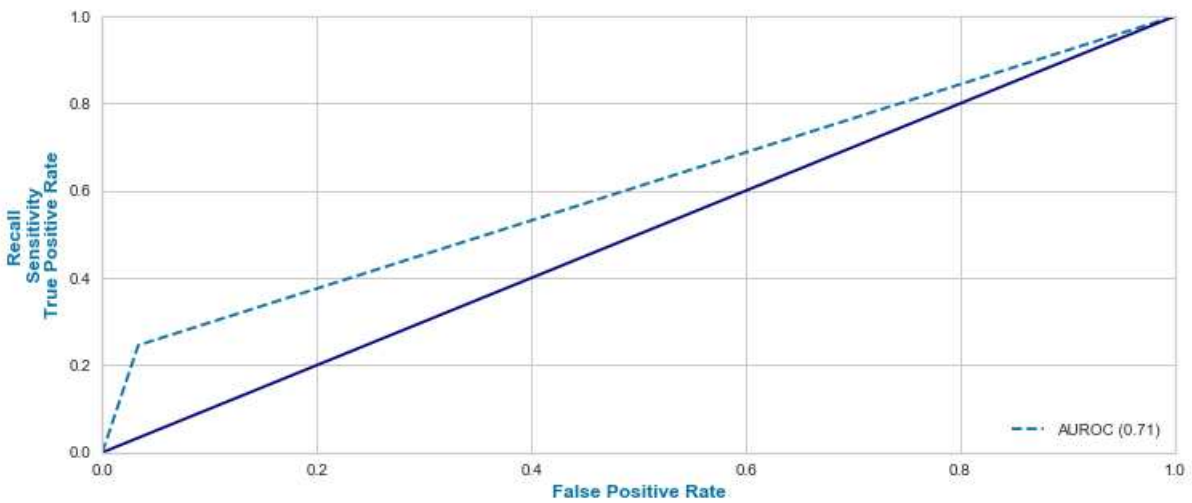
4.3 The Cumulative Accuracy Profile (CAP), GINI curve

For further understanding, one can refer to the definition of the Cumulative Accuracy Profile (CAP) on (BCBS, 2005b). We briefly introduce this criterion as follows.

CAP is also known as the GINI curve, Power curve, or Lorenz curve. Indeed, it is a visual tool whose graph can be easily drawn if two representative samples of scores for defaulted and non-defaulted borrowers are available.

The CAP can be used to evaluate a model by comparing the curve to the perfect CAP in which the maximum number of defaulted borrowers is achieved directly and to the random CAP in which the defaulted borrowers are distributed equally. A good model will have a CAP between the perfect CAP and the random CAP with a better model tending to the perfect CAP.

The accuracy ratio (AR) is defined as the ratio of the area between the model CAP and the random CAP and the area between the perfect CAP and the random CAP. For a successful model, the AR has values between zero and one, with a higher value for a stronger model.

Figure 4 shows the CAP of our estimated PD model. Accordingly, the area of CAP is almost identical to the one of the perfect model. However, this finding is not reasonable since we expect that the CAP should be significantly lower than the one of the perfect model. Furthermore, we found a bug in our computer programming, which should be fixed soon.

Figure 4: The Cumulative Accuracy Profile
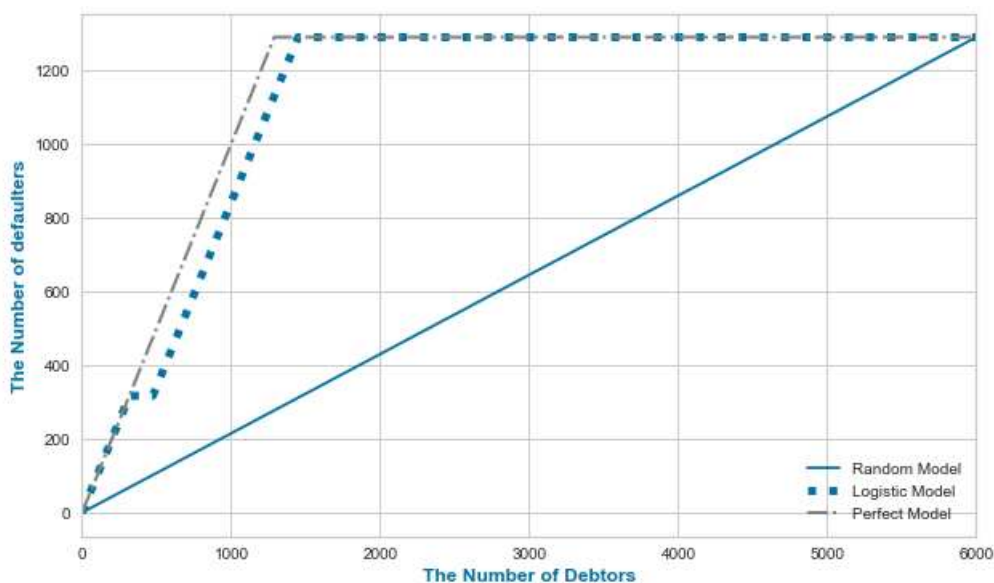
## 5 Model Monitoring: our model performs stably over time?

Given that our estimated PD model has passed all examinations and validations conducted by both our bank and the State Bank of Vietnam. It implies we can use our estimated Logistic model to compute the PD value. Thus, by using the Risk Weight function (see BCBS (2005a)) and the other three risk components, RWA can be computed for each asset class. So, is this project ended?

No, there still exists another issue as follows. We estimated our PD model based on the historical data. Afterward, we apply it in a new economic situation. The input of model is still identical to two periods. However, the economic situation would be changed. Thus, a question is that is the performance of our estimated PD model stable over time?

To make sure the stability of the performance of our model is always maintained, the model monitoring is really necessary every three months, for instance. To this end, we record the performance of our model. Afterward, we compare it with the past. The Population Stable Index (PSI) is used to give us information about whether we should continue to use our model.

Figure 5 below shows the distributions of the predicted PD in two different periods of time. The blue area donates the training period, whereas the light green one presents the out-of-sample part. Graphically, one might see that these two curves overlay on each other. It implies the performance of our model seems to be stable over time.

Figure 5: The Distribution of the Forecasted Values of PD
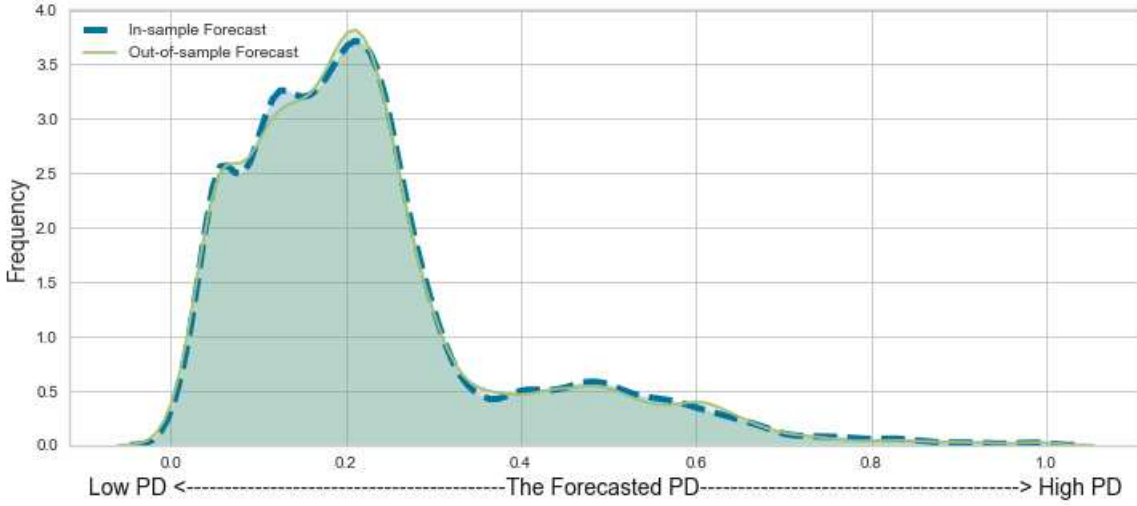
In terms of the statistical aspect, the model monitoring by using PSI is briefly described as follows. We sort all PD according to ascending order. Afterward, PD is allocated into the different interval, here is defined as the bucket. The number and percentage of PD are then computed in each bucket. Finally, PSI in each bucket is computed as the following formula.

$$PSI = \sum_{i=i}^{n} \left[\left(\frac{c_i}{V} - \frac{b_i}{B} \right) * \log \left(\frac{c_i}{V} : \frac{b_i}{B} \right) \right] \tag{5.0.1}$$

where c_i is the count of validation records in the i-th score band. V donates the total count of the validation records. b_i the count of baseline records in the i-th score band and B is the total count of baseline records

How to use PSI?

The total of PSI in each bucket is used to examine whether we continue to use our model. Indeed, we do as follows. PSI < 0.1 - No change. You can continue using existing model. PSI >= 0.1 but less than 0.2 - Slight change is required. PSI >=0.2 - Significant change is required.

Ideally, you should not use this model any more.

The results of PSI are shown in Figure 6 and Table 3. Accordingly, the total PSI is 0.019, which implies that we can continue to use our estimated LD model.

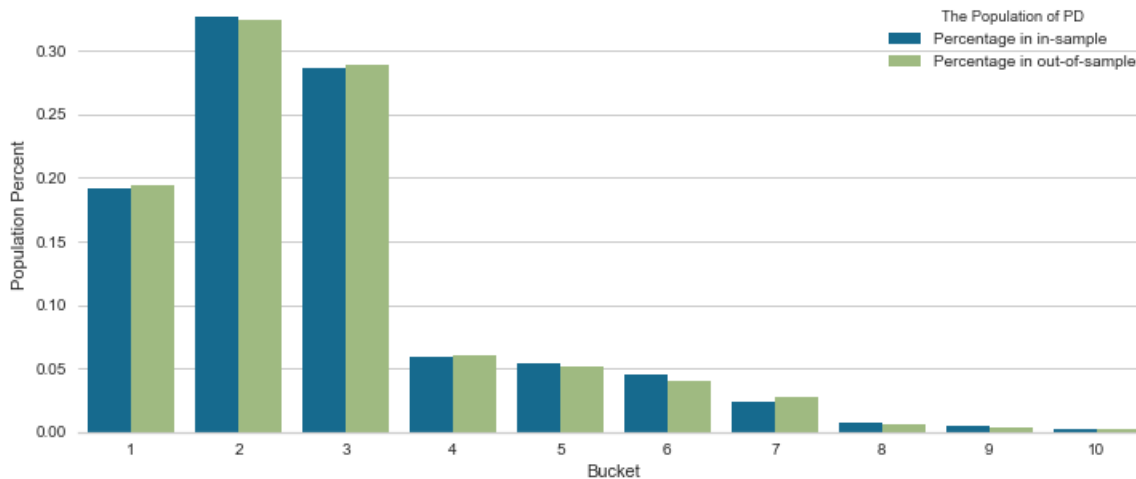Figure 6: The Distribution of the Forecasted Values of PD in Buckets

Table 3: The results of PSI computation

	Bucket	Breakpoint Value	counts in in-sample	counts in out-of-sample	Percentage in in-sample	Percentage in out-of-sample	PSI
0	1	0.099516	4589	1162	0.191208	0.193667	0.000031
1	2	0.199032	7844	1949	0.326833	0.324833	0.000012
2	3	0.298547	6877	1734	0.286542	0.289000	0.000021
3	4	0.398063	1410	360	0.058750	0.060000	0.000026
4	5	0.497579	1295	311	0.053958	0.051833	0.000085
5	6	0.597095	1077	240	0.044875	0.040000	0.000561
6	7	0.696611	570	166	0.023750	0.027667	0.000598
7	8	0.796127	182	39	0.007583	0.006500	0.000167
8	9	0.895642	105	23	0.004375	0.003833	0.000072
9	10	0.995158	51	16	0.002125	0.002667	0.000123

6 Conclusions

In this paper, we present the fundamental procedure for developing, validating, and monitoring a PD model within the Foundation IRB framework. To strengthen the practical application, however, the aspects including stability of parameters, calibration, bootstrap should be examined.

References

BCBS (2005a). An Explanatory Note on the Basel II IRB Risk Weight Functions. *Bank for International Settlements Papers*.

BCBS (2005b). Studies on the Validation of Internal Rating Systems (revised). *Bank for International Settlements, Working Paper No. 14*.

BCBS (2006a). International Convergence of Capital Measurement and Capital Standards: A Revised Framework. *Bank for International Settlements Papers.*

BCBS (2006b). International Convergence of Capital Measurement and Capital Standards: A Revised Framework Comprehensive Version. *Bank for International Settlements Papers.*

Engelmann, B., Hayden, E., and Tasche, D. (2003). Measuring the Discriminative Power of Rating Systems. *Deutsche Bundesbank Discussion paper Series 2: Banking and Financial Supervision No 01/2003.*

Engelmann, B. and Rauhmeier, R. (2006). *The Basel II Risk Parameters: Estimation, Validation, and Stress Testing.*

Medema, L., Koning, R. H., and Lensink, R. (2009). A practical approach to validating a PD model. *Journal of Banking Finance*, 33(4):701–708.

Nguyen, P. V. (2019). An attempt to derive the risk weight function for the bank. *Munich Personal RePEc Archive (MPRA)Paper No. 100585.*