Indian Statistical Institute

STATISTICAL INDIAN INSTITUTE

भिन्नेष्वैक्यस्य दर्शनम्

UNITY IN DIVERSITY

Project Internship Report

Submitted by: - **Harshad Kumar**

Roll Number : - **QR1919**

Organisation: **Credit Suisse Business Analytics, India**

CREDIT SUISSE

Prof. Nandini Das

____

ACKNOWLEDGEMENT

Any type of work cannot be completed without giving the gratitude to the people whose priceless co-operation and guidance make the work possible to complete. It is my radiant sentiment to place on record my best regards and deepest sense of gratitude to my supervisor *Prof. Nandini Das, SQC & OR Unit, Indian Statistical Institute, Kolkata*, for his careful and precious guidance which is extremely valuable for this project.

I would also like to thank my department for providing me the opportunity to do internship at Credit Suisse Business Analytics India. These six months have been a great learning experience for me, moreover I have got the actual taste of data analysis in the practical field of finance.

I am highly indebted to *Mr. Tarun Kumar, VP, Credit Suisse* and *Omkar Chitale, Director, Credit Suisse* for their immense co-operation, support and guidance to bring this project at the present state.

I am also thankful to all my colleagues at Credit Suisse, my friends and my professors for their support and help.

Table of Contents

*The contents are hyperlinked (use Ctrl+ click to access the content)

1 ABOUT CREDIT SUISSE

Credit Suisse is a global wealth manager, investment bank and financial services company founded and based in Switzerland. Headquartered in Zürich, it maintains offices in all major financial centers around the world and is one of the nine global "Bulge Bracket" banks providing services in investment banking, private banking, asset management, and shared services. Credit Suisse is known for its strict bank–client confidentiality and banking secrecy practices.

Credit Suisse was founded in 1856 to fund the development of Switzerland's rail system. It issued loans that helped create Switzerland's electrical grid and the European rail system. In the 1900s, it began shifting to retail banking in response to the elevation of the middle class and competition from fellow Swiss banks UBS and Julius Bär. Credit Suisse partnered with First Boston in 1978. After a large failed loan put First Boston under financial stress, Credit Suisse bought a controlling share of the bank in 1988. From 1990 to 2000, the company made a series of acquisitions dramatically increasing their market share via the purchases of Winterthur Group, Swiss Volksbank, Swiss American Securities Inc. (SASI) and Bank Leu, among others.

Credit Suisse has two divisions, Private Banking & Wealth Management and Investment Banking. A Shared Services department provides support functions like risk management, legal, IT and marketing to all areas. Operations are divided into four regions: Switzerland, Europe, the Middle East and Africa, the Americas and the Asian Pacific. Credit Suisse Private Banking has wealth management, corporate and institutional businesses. Credit Suisse Investment Banking handles securities, investment research, trading, prime brokerage and capital procurement. Credit Suisse Asset Management sells investment classes, alternative investments, real-estate, equities, fixed income products and other financial products.

1.1 Financial Products

Credit Suisse endorses a strategy called bancassurance of trying to be a single company that offers every common financial services product. The investment bank is intended for companies and wealthy individuals with more than 50,000 euro. Credit Suisse developed the Credit Risk + model of risk assessment in loans, which is focused exclusively on the chance of default based on the exogenous Poisson method. As of 2002 about 20 percent of Credit Suisse's revenue was from its insurance business it gained through the 1997 acquisition of Winterthur.

The investment bank's insurance products are primarily popular in the domestic market and include auto, fire, property, life, disability, pension and retirement products among others. Historically 20–40 percent of the bank's revenue has been from private banking services, one of its higher profit-margin divisions.

Credit Suisse produces one of the six hedge funds following European stock indices that are used to evaluate the performance of the markets. The investment bank also has a 30 percent ownership in hedge fund investment firm York Capital Management.

York sells hedge funds independently to its own clients, while Credit Suisse also offers them to private banking clients. Credit Suisse manages the financial instruments of the Dow Jones Credit Suisse long/short equity index (originally called Credit Suisse/Tremont Hedge Fund Indexes).

Project 1 Title: - Model Development (EAD Models)

1 INTRODUCTION

The division in which I have completed my internship is QAT, which stands for Quantitative Analysis and Technology. QAT comprises of various teams starting from IDEAL (Internal Data Repository of CS), TSRD, Data Teams, Modelling Teams like (FWST, AIRB), ERC Teams, and also Validation Teams like MRM.

I have worked in AIRB (Advanced Internal Rating Based Models)-Modelling teams.

In the next part, with the help of definitions/concepts related to financial models, I have tried to present insight about the role of the team and its functions.

1.1 AIRB (Advanced Internal Rating Based Models)

Under the Basel II guidelines (which are recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision), banks are allowed to use their own estimated risk parameters for the purpose of calculating regulatory capital. This is known as the Advanced Internal Ratings-based (A-IRB) approach to capital requirements for credit risk.

1.2 Credit Risk

A credit risk is risk of default on a debt that may arise from a borrower failing to make required payments. In the first resort, the risk is that of the lender and includes **lost principal and interest, disruption to cash flows, and increased collection costs**. The loss may be complete or partial. In an efficient market, higher levels of credit risk will be associated with higher borrowing costs. Because of this, measures of borrowing costs such as yield spreads can be used to infer credit risk levels based on assessments by market participants.

1.3 Credit Risk Management by holding capital

Starting with a small question: **-**

Why holding capital amount is necessary for Bank? Here is the answer, to cover the credit risk exposure, Bank need to show enough capital/cash to bear the loss from this risk exposure. For more understanding think it as best and worst case scenario. The worst case one could imagine would be that banks lose their entire credit portfolio in a given year. This event, though, is highly unlikely, and holding capital against it would be economically inefficient. Banks have an incentive to minimise the capital they hold, because reducing capital frees up economic resources that can be directed to profitable investments. On the other hand, the less capital a bank holds, the greater is the likelihood that it will not be able to meet its own debt obligations, i.e. that losses in a given year will not be covered by profit plus available capital, and that the bank will become insolvent. **Thus, AIRB team must carefully balance the risks and rewards of holding capital.**

Under A-IRB banks are supposed to use their own quantitative models to estimate PD (probability of default), EAD (exposure at default), LGD (loss given default) and other parameters required for calculating the RWA (risk-weighted asset). Then total required capital is calculated as a fixed percentage of the estimated RWA. The detailed explanation about various AIRB approaches and used approach is showed in approach section.

2 OBJECTIVE

To ensure carefully and statistically monitored models to estimate Capital Requirement following objective needs to be executed which are regulated by respectively national supervisors: -

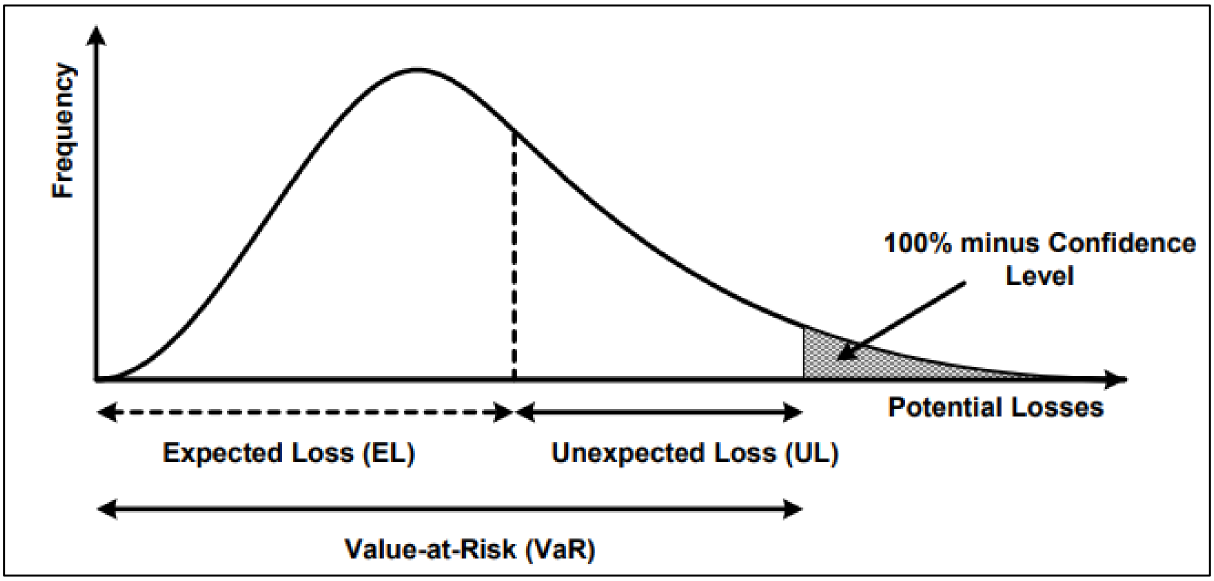
Need to develop models to estimate/calibrate the three risk parameters on which the capital requirement depends: -

i. PD (Probability of Default): - A multinominal model
ii. LGD (Loss given Default): - A regressive model
iii. EAD (Exposure at default): - Credit Conversion Approach is used (explained later stage)

Some more details about these models are explained in Approach Section. My project objective includes development of model for EAD estimate using R programming language and Bank internal data sets.

3 APPROACH

The AIRB approach adopted for estimating capital requirement for the bank is focused on the frequency of bank insolvencies arising from credit losses that supervisors are willing to accept. By means of a stochastic credit portfolio model, it is possible to estimate the amount of loss which will be exceeded with a small, pre-defined probability. This probability can be considered the probability of bank insolvency. Capital is set to ensure that unexpected losses will exceed this level of capital with only this very low, fixed probability. This approach to setting capital is illustrated in below figure



The curve in above figure describes the likelihood of losses of a certain magnitude. The area under the entire curve is equal to 100% (i.e. it is the graph of a probability density). The curve shows that small losses around or slightly below the Expected Loss occur more frequently than large losses. The likelihood that losses will exceed the sum of Expected Loss (EL) and Unexpected Loss (UL) - i.e. the likelihood that a bank will not be able to meet its own credit obligations by its profits and capital - equals the hatched area under the right-hand side of the curve. 100% minus this likelihood is called the confidence level and the corresponding threshold is called Value-at-Risk (VaR) at this confidence level. If capital is set according to the gap between origin and VaR then the likelihood that the bank will remain solvent over a one-year horizon is equal to the confidence level. Under Basel II, capital is set to maintain a supervisory fixed confidence level.

So far, the Expected Loss has been regarded from a top-down perspective, i.e. from a portfolio view. It can also be viewed bottom-up, namely from its components. The Expected Loss of a portfolio is assumed to equal the proportion of obligors that might default within a given time frame (1 year in the Basel context), multiplied by the outstanding exposure at default, and once more multiplied by the loss given default rate (i.e. the percentage of exposure that will not be recovered by sale of collateral etc.). Of course, banks will not know in advance the exact number of defaults in a given year, nor the exact amount outstanding nor the actual loss rate; **these factors are random variables**. But banks can estimate average or expected figures. These three factors are termed as risk parameters as mentioned in the introduction section. These are

- Probability of Default (PD) per rating grade, which gives the average percentage of obligors that default in this rating grade in the course of one year
- exposure at default (EAD), which gives an estimate of the amount outstanding (drawn amounts plus likely future drawdowns of yet undrawn lines) in case the borrower defaults
- loss given default (LGD), which gives the percentage of exposure the bank might lose in case the borrower defaults. These losses are usually shown as a percentage of EAD, and depend, amongst others, on the type and amount of collateral as well as the type of borrower and the expected proceeds from the work-out of the assets.

The Expected Loss (in currency amounts) can then be written as

$$EL = PD * LGD * EAD$$

Note, this model is portfolio invariant, i.e. the capital required for any given loan should only depend on the risk of that loan and must not depend on the portfolio it is added to. This characteristic has been deemed vital in order to make the new framework applicable to a wider range of countries and institutions.

EAD Model

1. Definition of EAD and modelling framework

EAD for a non-defaulted facility is an estimate of the gross exposure upon default of the obligor. It is viewed as a random variable with distribution depending on the drawn amount, $D(t_0)$ at the reference date, t_0 and limit C of the facility which is assumed to persist throughout the observation period. Note, that C like the credit limit which we have for Credit card transaction.

Hence, EAD = $EAD(D(t_0), C)$ where drawn amount, $D(t_0)$ at the reference date is known to us.

The modelling framework allows for the possibility that the expected EAD depends on the distant to default $\Delta t = d - t_0$, where d is the default date. Specifically, Δt itself is modelled as a random variable and hence the conditional random variable $EAD^c = (EAD|\Delta t)$ may be correlated with Δt. However, in the CCF modelling approach to estimate EAD, we can easily ignore this correlation from our model.

2. CCF Approach to estimate EAD

EADs are calculated following the Credit Conversion Factor (CCF) approach. Under this approach, a scalar CCF is used to convert an undrawn but committed amount into a loan equivalent. Specifically, EAD is modelled for each facility as the sum of the drawn exposure $D(t_0)$ at reference date t_0 plus a percentage (CCF) of the undrawn portion of the commitment.

$$\mathbf{EAD(D(t_0), C) = D(t_0) + (C - D(t_0)) * CCF \qquad \text{---------} \ (A)}$$

Now, CCF is a random variable whose distribution does not depend on $D(t_0)$ or C. The CCF represents the percentage of the undrawn portion of a commitment that is expected to be drawn by the time the counterparty defaults. Importantly, the CCF estimate does not intent to capture movements in exposure and commitments with time.

However, for capital purposes a threshold value of CCF is assigned to all the facilities of product type. This threshold value is based on an appropriate estimate of the expected EAD.

The CCF estimate is obtained using historical information on realized CCFs. This type of calculation requires information on exposures for defaulted counterparties both at default and at a given date prior to default (reference date). When these are known, the realized CCF for a given facility (i) is given by

$$\boldsymbol{CCF}\mathbf{i}(\mathbf{k}) = \frac{\mathbf{Di}(\mathbf{d}) - \mathbf{Di}(\mathbf{d}-\mathbf{k})}{\boldsymbol{Ci}(\boldsymbol{d}-\boldsymbol{k}) - \boldsymbol{Di}(\boldsymbol{d}-\boldsymbol{k})}$$

Where subscript i represents the facility index and D(d) is the drawn amount at default date(d) which is numerically equivalent to the exposure at default (EAD). Now, k corresponds to the time horizon (e.g. 12 months/1 year). C(d-k) is the commitment of the facility at k horizon time before default. In simple words, realized CCF is just the percentage of amount drawn from available commitment (i.e. undrawn amount, also called as headroom) at reference date (k horizon time before default) which is inline with our definition of CCF.

I have computed CCF using two methodologies **a) Fixed time horizon (FTH) and b) Variable time horizon (VTH).**

Under the FTH approach, CCF estimates are obtained by averaging the realized CCF over the 'n' different facilities while for each facility realized CCF is computed for a fixed time horizon(k)

$$\boldsymbol{CCF} = \frac{\sum_{i=1}^{n} \boldsymbol{CCFi}}{\boldsymbol{n}}$$

Under the VTH approach, CCF estimates are obtained averaging over two variables(i,k). Here, realized CCFs are computed for 'm' various time horizons (k_1, k_2, k_3, k_4,……,k_m)

$$\boldsymbol{CCF} = \frac{\sum_{k=k1}^{km} \sum_{i=1}^{n} \boldsymbol{CCFi}(\boldsymbol{k})}{\boldsymbol{m} * \boldsymbol{n}}$$

While applying the above concept, some more assumptions need to be made.
Stepwise CCF computation explanation: -

Step 1: The first step in the calculation of the Base CCF consists on defining a minimum undrawn amount (or headroom) at the reference date. The definition of a minimum over the undrawn amounts is driven by two considerations. On one hand, the inclusion of cases with small undrawn amounts is not consistent with the use of the CCF parameter when calculating EAD in the Investment Bank division (e.g. it is expected that IB deals are associated with either significant undrawn amounts or loans fully drawn). Therefore, for a given horizon, cases with undrawn amounts threshold of USD 10,000 are removed. Thus, these cases are treated as equivalent to fully prior to default for which it is not meaningful to calculate a realised CCF.

Step 2: The second step is related to those cases for which exposures decrease or do not change in running to default. In the absence of **accrued interest (AI),** these cases yield a negative respectively. These cases are 'floored° to zero if after adding accrued interests the resulting negative. In case it is increased then the maximum value of drawn exposure is considered. Both of these adjustments are conservative in nature. Note, accrued interest (AI) refers to the amount of interest that has been incurred, as of a specific date, on a loan or other financial obligation but has not yet been paid out.

Step 3: The third step deals with those cases for which commitments change between the reference date and the default date but for which exposures increase in the run to default. In case of decrease in commitments, there is no change in CCF calculation. When commitments increase between reference and default dates, an assumption about how these increases are assigned between the credit line and the adjusted one. For this, it is assumed that any increase exposure comes proportionately from the available headroom at the observation date and the increase commitment until default. Hence the denominator is accordingly adjusted.

4 ANALYSIS

The listed approach is practically applied using R programming language. The development sample is based on information from IB Default & Loss (IDEAL) database which contains internal information about the exposure and commitments from defaulted CS clients. IDEAL database provides default, exposure and loss data for the erstwhile CS Investment Banking portfolio. All counterparties and exposure types that are subject to the CS RWA calculations are in scope for the compilation of this database. This generally refers to exposures being booked in the Banking Book. However, for comprehensiveness and in order not to restrict potential use in model development, the relevant data for all counterparties that meet the Basel default criteria is sourced.

Following data cleaning steps has been performed: -

1. **Remove facilities with missing facility type or branch information or any other critical information**: Such facilities need to be excluded. The number of facilities under this filter is negligible.
2. **Consistency between exposure and commitment tables:** The facilities with information but no commitment information is deleted. Their deletion, does not have any impact in the calculations as all of them do not have commitments one year prior or post the default date. More important are the number of facilities with commitment information which do not have any information about exposure i.e. these cases could represent facilities which were not used.
3. **Removing facilities associated with non-defaulted counterparties.**
4. **Removing pre-2002 defaults**: These defaults are not considered as IDEAL systematically compiles information about exposures and commitments for post-2001 defaults.
5. **Removing records with time-span inconsistencies**: It refers to the inconsistency in "valid from" and "valid uptill" dates in the data points. Record with such inconsistency must be filtered.
6. **Remove records with change in effective date:** Exposure and/or commitments for facilities with changes in effective date are currently not captured in IDEAL.
7. **Keep records for only relevant time window:** Only transactions valid 12 months prior to default are captured. The time – window is defined by regulatory guidelines.
8. **Remove records with facility information only after default:** These counterparties with facility information only appearing after default are not considered in CCF calculations.
9. **Remove facilities with missing FX rate:** Facilities with at least one period of missing exchange rate (FX) information are excluded from the sample.

After sourcing the Data, CCF are calculated using R programming language following the discussed approach for each bank's product type like Regular loans, Term Loan, Syndicated Loans and Contingent Liabilities.

5 RESULT

The estimated CCFs are used to estimate the EAD as per the equation (A). Similarly, after getting results from the LGD/PD models from their respective model owners we used to get estimated values of all the three-risk parameter defined in AIRB approach. These estimated risk parameters under AIRB approach are used to estimate the Capital requirement, which Bank holds for its Credit Risk exposure.

6 CONCLUSION AND RECOMMENDATION

The CCFs estimates from the present model are quite stable and conservative in nature and model is also approved from the regulators. I want to share a recommendation to the team for EAD model development. As already discussed, presently I have developed EAD model indirectly using CCF approach. The credit conversion factor (CCF), the proportion of the current undrawn amount that will be drawn down at time of default, is used to calculate the EAD and poses modelling challenges with the distribution bounded between zero and one. I recommended them to use alternative EAD models which ignore the CCF formulation and target the EAD distribution directly. There are various publications on these alternatives, one of them proposes a mixture model with the zero-adjusted gamma distribution and existing CCF approach.

7 FOLLOW UP

These alternative EAD model approaches can be implemented and after a discussion with the regulator approach can be approved for next year analysis.

Project 2 Title: - Annual Model Monitoring

1 INTRODUCTION

We need to annually perform monitoring of AIRB models to ensure that the models are fit for the regulatory capital estimation purposes. For this I used to perform backtesting analysis, in which I test whether the realized PDs, LGDs and EADs are greater than the estimated one using various hypothesis testing described in the approach section. In addition to the backtesting, we also perform Discriminatory power analysis. Discriminatory power analysis is a goodness of fit approach which evaluates how much models are able to discriminate between facilities with high and low values of estimates, statistical concept is explained in approach section. Monitoring also involves one more exercise to report the present materiality figures (the bank exposures) at the various portfolio level. In other words, to quantify the bank business at various segments (like Corporates, Hedge Funds, Mutual Funds) which are termed as portfolios.

2 OBJECTIVE

To perform the monitoring analysis for LGD/EAD models which are included in the Annual MPM (Model Performing Monitoring) Report which is submitted to the concerned regulators (FINMA/ PRA) for evaluation.

3 APPROACH

We perform the monitoring of AIRB models as per the European Central Bank (ECB) guidelines. I have monitored two models: - LGD and EAD models, complete statistical approach is described below

3.1 Monitoring EAD model (based on CCF estimates)

As already discussed in EAD model development, EAD computation is based on CCF estimate. Hence, tests are performed on CCF estimates.

- **CCF Backtesting using a t-test**

The one-sample t-test for paired observations compares the estimated CCF with the realised CCF under the null hypothesis that the estimated CCF is greater than the true one (one-sided hypothesis test), assuming independent observations. Under the null hypothesis, the test statistic is asymptotically Student-t distributed with (R − 1) degrees of freedom, where R is the number of facilities (back-testing).

The data basis for the t-test consists of all facilities that have defaulted during the relevant observation period. Facilities that are affected by outlier treatment, such as floors, but are included in the internal validation sample, form part of the relevant data basis.

Calculate the t-test statistic as follows:

$$\mathrm{T} = \sqrt{R} \cdot \frac{\sum_{i=1}^{R}(CCF_i^R - CCF_i^E)}{\sqrt{s_{CCF}^2}},$$

$$s_{CCF}^2 = \frac{\sum_{i=1}^{R}\left((CCF_i^R - CCF_i^E) - \frac{1}{R}\sum_{j=1}^{R}(CCF_j^R - CCF_j^E)\right)^2}{R-1},$$

where:

R is the number of facilities

$CCF^{E}{}_i$ is the estimated CCF for facility i

$CCF^{R}{}_i$ is the realised CCF for facility i

Calculate the p-value 1 − $S_{R-1}(T)$, where S_{R-1} is the cumulative distribution function of the Student's t-distribution evaluated at the test statistic (T) with (R − 1) degrees of freedom. The test is performed at α = 95% and is conclusive only if we have at least 20 number of observations.

- **<u>CCF backtesting using a Wilcoxon signed Rank test</u>**

The objective of the CCF backtesting via a Wilcoxon signed-rank test for paired observations is to ensure that the realized CCFs are less than the estimated CCFs.

The Wilcoxon signed-rank test for paired observations is a nonparametric test of the null hypothesis that the estimated CCFs are greater than the true ones (one-sided hypothesis test). The test compares the estimated CCFs with the realised CCFs under the assumption that the observations are paired, come from the same population and the responses are ordinal (i.e. of any two observations, it is possible to say which is the greater).

The relevant data basis for the test consists of all facilities defaulted during the relevant observation period. Data cleansing and treatment of outliners is performed in the same way as for the estimation of conversion factors.

Calculates the Wilcoxon signed-rank test as follows:

1. Calculate for each pair, the difference of realised and estimated CCFs:

$$d_i = CCF_i^R - CCF_i^E$$

2. Remove from the total sample of observations (N) those with a d_i equal to zero. The new number of paired observations is then N_0.

3. Rank the absolute differences $|d_i|$ (i.e. assign rank R = 1 to the smallest $|d_i|$, rank R = 2 to the next, etc.). Where there are groups of tied values (i.e. several observations sharing the same $|d_i|$), a rank equal to the midpoint of the unadjusted rankings is assigned

4. label each rank with its sign, according to the sign of d_i

$$label_i = sgn(CCF_i^R - CCF_i^E)$$

5. calculate W^+ as the sum of the ranks of the positive d_i (labelled positive):

$$W^+ = \sum_{i=1}^{N0} R|_{(di > 0)}$$

Under the null hypothesis and for sufficiently large number of observations/pairs is large enough, the test statistic z is normal distribution with Parameters μ_w and σ_w

Calculate as follows

$$Z = \frac{(W^+ - \mu_w)}{\sigma_w}$$

$$\mu^w = \frac{N_0(N_0 + 1)}{4}$$

$$\sigma_w = \sqrt{\left(\frac{N_0(N_0 + 1)(2N_0 + 1)}{24}\right) - \frac{\sum_t(f_t^3 - f_t)}{48}}$$

where:

a) CCF_i^E *is* the estimated CCF for facility i

b) CCF_i^R is the realised CCF for facility i
c) N is the number of facilities in default at the beginning of the relevant observation period whose workout has been closed within the relevant observation period
d) N_0 is equal to the total number of the facilities N, minus the number of observations whose difference between realised and estimated CCFs is equal to zero ($d_i = CCF_i^R - CCF_i^E = 0$).
e) t varies over the set of tied ranks and f_t is the number of times (i.e. frequency) the rank t appears

Calculate the p-value $1 - \phi(Z)$, where ϕ denotes the cumulative distribution function of the standard normal distribution evaluated at the test statistic

3.2 Monitoring LGD models

LGD model is developed by my team-mates. Here, I have included short introduction about LGD and the models. LGD is the loss, expressed as a percentage of the EAD, on a credit facility if the credit defaults. LGD for a non-defaulted facility can be defined as the estimate of loss conditional on the default, expressed as a percentage of the EAD. The LGD associated with a non-defaulted facility can be viewed as a random variable. LGD for a defaulted facility is the realized loss expressed as a percentage of the exposure at the time of default. If there is complete information on all of the losses related to a facility, and a method to calculate losses has been chosen, we can directly calculate realised LGD. If there is not complete information on the losses related to a defaulted facility, for example if the facility is in the process of workout, LGD is a random variable. We can calculate an estimate of realized LGD for these defaulted facilities by using complete information from a sample of similar facilities.

I have performed the LGD models monitoring using backtesting using the two same tests as of EAD models i.e. t-test and Wilcoxon. In addition, discriminatory power analysis is also performed and discussed in detail.

- **LGD back-testing using a t-test**

The one-sample t-test for paired observations compares estimated LGD with realised LGD under the null hy pothesis that estimated LGD is greater than true LGD (one-sided hypothesis test) assuming independent observations. Under the null hypothesis, the test statistic is asymptotically Student-t distributed with $(N - 1)$ degrees of freedom, where N denotes the number of facilities (backtesting).

Calculate the t-test statistic as follows:

$$T = \sqrt{N}\,\frac{\frac{1}{N}\sum_{i=1}^{N}\left(LGD_i^R - LGD_i^E\right)}{\sqrt{s_{LGD}^2}}\ ,$$

$$s_{LGD}^2 = \frac{\sum_{i=1}^{N}\left(\left(LGD_i^R - LGD_i^E\right) - \frac{1}{N}\sum_{j=1}^{N}\left(LGD_j^R - LGD_j^E\right)\right)^2}{N-1},$$

where:

a) N is the number of facilities (back-testing)
b) LGD_i^E denotes the estimated LGD for facility i
c) LGD_i^R denotes the realised LGD for facility i

Calculate the p-value $1 - S_{N-1}(T)$, where S_{N-1} is the cumulative distribution function of the Student t-distribution evaluated using the test statistic (T) with $(N - 1)$ degrees of freedom.

- **<u>LGD backtesting using a Wilcoxon signed Rank test</u>**

The objective of the LGD backtesting via a Wilcoxon signed-rank test for paired observations is to ensure that the realized LGDs are less than the estimated LGDs.

The Wilcoxon signed-rank test for paired observations is a nonparametric test of the null hypothesis that the estimated LGDs are greater than the true ones (one-sided hypothesis test). The test compares the estimated LGDs with the realised LGDs under the assumption that the observations are paired, come from the same population and the responses are ordinal (i.e. of any two observations, it is possible to say which is the greater).

The relevant data basis for the test consists of all facilities in default at the beginning of the relevant observation period whose workout has been closed within the relevant observation period. A workout agreement means the process in which lender renegotiate with the defaulted counterparty about the terms of a loan and recovery of loan.

Calculates the Wilcoxon signed-rank test as follows:

1. Calculate for each pair, the difference of realised and estimated CCFs:

$$d_i = LGD_i^R - LGD_i^E$$

2. Remove from the total sample of observations (N) those with a d_i equal to zero. The new number of paired observations is then N_0.

3. Rank the absolute differences $|d_i|$ (i.e. assign rank R = 1 to the smallest $|d_i|$, rank R = 2 to the next, etc.). Where there are groups of tied values (i.e. several observations sharing the same $|d_i|$), a rank equal to the midpoint of the unadjusted rankings is assigned

4. label each rank with its sign, according to the sign of d_i

$$label_i = sgn(LGD_i^R - LGD_i^E)$$

5. calculate W^+ as the sum of the ranks of the positive d_i (labelled positive):

$$W^+ = \sum_{i=1}^{N0} R|_{(di>0)}$$

Under the null hypothesis and for sufficiently large number of observations/pairs is large enough, the test statistic z is normally distributed with Parameters μ_w and σ_w

Calculate as follows

$$Z = \frac{(W^+ - \mu_w)}{\sigma_w}$$

$$\mu^w = \frac{N_0(N_0+1)}{4}$$

$$\sigma_w = \sqrt{\left(\frac{N_0(N_0+1)(2N_0+1)}{24}\right) - \frac{\sum_t(f_t^3 - f_t)}{48}}$$

where:

f) LGD_i^E *is* the estimated LGD for facility i
g) LGD_i^R is the realised LGD for facility i
h) N is the number of facilities in default at the beginning of the relevant observation period whose workout has been closed within the relevant observation period
i) N_0 is equal to the total number of the facilities N, minus the number of observations whose difference between realised and estimated CCFs is equal to zero ($d_i = LGD_i^R - LGD_i^E = 0$).
j) t varies over the set of tied ranks and f_t is the number of times (i.e. frequency) the rank t appears

Calculate the p-value 1 – ϕ(Z), where ϕ denotes the cumulative distribution function of the standard normal distribution evaluated at the test statistic

- **Discriminatory Power Analysis of LGD Models**

The analysis of discriminatory power is aimed at ensuring that LGD models are able to discriminate between facilities with high and low values for LGD. The measure used in this section to assess the discriminatory power of LGD models is the generalised AUC. That validation tool is based on a generalisation of the classical AUC that can be applied to multi-class problems as shown.

Statistics of the discriminatory power of LGD/CCF models Calculation

Let LGD_i^E and LGD_i^R be estimated and realised LGD respectively for facility i.

Since the model is continuous and based on more than 20 facility grades/pools, an ordinal segmentation of LGD is applied using the LGD segments defined by the European central Bank guidelines which are described here in italics. *The institution should use 12 predefined "LGD segments" on the basis of the following criteria:*

Segment 1: facilities i with 0% ≤ LGD_i^E < 5%;

Segment 2: facilities i with 5% ≤ LGD_i^E < 10%;

Segment 3: facilities i with 10% ≤ LGD_i^E < 20%;

…

(10% LGD steps from Segment 3 to Segment 11);

…

Segment 12: facilities i with 100% ≤ LGD_i^E

The 12 segments are then ordered from low to high (i.e. from Segment 1 to Segment 12). The basis for the test is a two-way contingency table (12 times 12) with all possible combinations of discretised LGD_i^E (12 possible segments as rows) and LGD_i^R (12 possible segments as columns) and the observed frequencies for each combination for all pairs of defaulted facilities within the sample (see the corresponding LGD section for details of the relevant data basis for this test).

Let a_{ij} denote the observed frequency in cell (i,j) (i.e. segment i for LGD^E and segment j for LGD^R) in the 12x12 contingency table described above. Let $r_i = \sum_j a_{ij}$ be the total for row i, $c_j = \sum_i a_{ij}$ be the total for column j and $F = \sum_j \sum_i a_{ij}$ be the total frequency. Let

$$A_{ij} = \sum_{k<i} \sum_{l<j} a_{kl} + \sum_{k>i} \sum_{l>j} a_{kl}$$

$$D_{ij} = \sum_{k>i} \sum_{l<j} a_{kl} + \sum_{k<i} \sum_{l>j} a_{kl}$$

and

$$P = \sum_i \sum_j a_{ij} A_{ij}$$

$$Q = \sum_i \sum_j a_{ij} D_{ij}$$

P can be understood as twice the number of agreements (i.e. for a given

combination of estimated "i" and realised "j" discretised LGD, the total frequency of observations with both indices greater or smaller than the given combination) in the ordering of the cell indices when all pairs of observations are compared. Similarly, Q is twice the number of disagreements (i.e. for a given combination of estimated "i" and realised "j" discretised LGD, the total frequency of observations with at least one index greater or smaller than the given combination).

The following definition of generalised AUC(gAUC) assumes that the row variable LGD_E is regarded as an independent variable, while the column variable LGD_R is regarded as dependent. The gAUC is estimated as:

$$D = \frac{P - Q}{w_r}$$

$$gAUC = \frac{D + 1}{2}$$

and the gAUC's standard deviation (s) can be estimated as:

$$s = \frac{1}{w_r^2} \sqrt{\sum_j \sum_i a_{ij}(w_r d_{ij} - (P - Q)(F - r_i))^2}$$

Where

$$w_r = F^2 - \sum r_i^2$$

$$d_{ij} = A_{ij} - D_{ij}$$

Now, the generalised AUC (gAUC) for the relevant observation period is compared with the gAUC at the time of the initial validation during development via hypothesis testing based on a normal approximation, assuming a deterministic gAUC at the time of development. The null hypothesis of the test is that the gAUC at the time of development is smaller than the gAUC for the relevant observation period.
Calculate the test statistic:

$$S = \frac{gAUC_{init} - gAUC_{curr}}{s}$$

where $gAUC_{init}$ denotes the gAUC at the time of the initial validation, $gAUC_{curr}$ denotes the gAUC for the relevant observation period and s, denotes the estimated standard deviation of $gAUC_{curr}$.

Calculate the p-value 1 – ϕ(s), where ϕ denotes the cumulative distribution function of the standard normal distribution evaluated at the test statistic

4 ANALYSIS

The relevant processed data for back testing purpose for both EAD and LGD models is shared from the respective model developers. Note, EAD is modelled is developed by me and the monitoring is also performed on the same processed data while LGD data is shared by my team mate. I have practically applied the discussed backtesting approach and performed the analysis using the R programming language. The findings about the backtesting results are shared with the concerned parties and where we are observing a failure in model i.e. while conducting test we have enough evidence that realised EADs/LGDs are greater than the estimated one. In these cases, appropriate add on are decided on the estimates so that the required capital requirement can be adjusted accordingly.

5 RESULT

Backtesting Results: - Both EAD and LGD successfully passes all the relevant backtesting tests and it can be safely concluded that the realised estimates are not greater the estimates. Hence, the estimated capital requirement is sufficient to cover the bank credit exposure.

Discriminatory Power Analysis: - Moderate Discriminatory Power is observed for LGD model. The feedback and analysis result is shared with Model developer and decided, the results are optimal for this year estimates and will incorporate changes for next year LGD model.

6 CONCLUSION AND RECOMMENDATION

Monitoring exercises are completely based on the ECB regulatory guidelines and both EAD/LGD models passes the relevant test & constraints and models can be safely used for estimating risk parameters. Since the guidelines are defined by regulators, the analyst is not recommended to make any changes. However, I have read in many literatures that before conducting Wilcoxon signed Rank test, we should first check whether the distribution of the difference of realized and estimated parameters needs to be symmetric in nature. I just want to recommend a small symmetric distribution check to be included before applying Wilcoxon test.

7 FOLLOW UP

The need of symmetric distribution check before applying Wilcoxon signed rank test is important and hence, discussed with the team mates. We are thinking to implement these check and get approved from the regulators for next year analysis.