Indian Statistical Institute, Kolkata

Migrating Internal Ratings Based Models from MATLAB to Python

A dissertation submitted in fulfillment of the
requirements for the award of
Master of Technology
in
Cryptology and Security
By: Sourav Ghosh

Primary Supervisors:
Ashok Kumar Dengla
& Arindam Adak
HSBC

Secondary Supervisor:
Dr. Anisur Rahaman Molla
INDIAN STATISTICAL
INSTITUTE,Kolkata

Acknowledgements

This internship has proved to be a great opportunity for me to expand my learning and professional horizon. I consider myself very lucky and honoured to have so many wonderful people lead me through, in this project.

I am thankful to ISI, Kolkata for providing me with this opportunity to cross the classroom boundaries and work in the corporate sector. The internship opportunity I had with HSBC is a great chance for learning and professional development. I am also grateful for having a chance to interact with our faculty, Dr. Anisur Rahaman Molla, who always ensured that I am on the right path and that I make the best use of this opportunity.

I want to use this opportunity to express my deepest gratitude and special thanks to Arindam Adak, Ashok Kumar Dengla, Soumya Boral, Dhiraj Laxman Patle and all my colleagues who in spite of being remarkably engaged and busy with their duties, took time out to guide me on the correct path and allowed me to carry out my project within the timeline. I am truly grateful for Dipanjan Sil's guidance and support throughout the process of writing this report

I would also like to thank my friends who have supported and motivated me during this time.

Abstract

Banking regulators mandate certain restrictions on banks to ensure that they are adequately capitalized and prepared for a variety of unforeseen adverse economic scenarios. The adequacy of the regulatory capital is decided based on multiple parameters, which are often determined by Internal Ratings Based (IRB) models. To ensure the accuracy and reliability of these models, banks must regularly assess these models based on defined parameters.

Currently, these models are implemented using MATLAB, a licensed software that results in significant costs for HSBC. In order to reduce such costs and upgrade to the latest technology stack, HSBC has taken the initiative to migrate these models from MATLAB to Python. Python is an open-source programming language that offers a variety of libraries and tools to support data analysis, machine learning, and other computational tasks.

This project involves replicating a model that calculates different parameters like Probability of Default (PD). To replicate the existing MATLAB code in Python, we will need to understand the underlying mathematical models and algorithms used. We will also need to ensure that the Python code produces the same results as the original MATLAB code. To check the sanity of the code, multiple test cases would be used to verify the results.

Once we have successfully migrated the code to Python, we will evaluate its performance and identify any areas for improvement. Overall, this project is significant because it enables HSBC to upgrade to the latest technologies and reduce software costs while maintaining the accuracy and reliability of their IRB models. It also demonstrates the benefits of open-source tools like Python for the financial services industry.

Contents

Chapter 1

Introduction

HSBC is a global banking and financial services organization with a strong presence in over 60 countries, including India. It offers a wide range of services to individuals and businesses, including retail and commercial banking, investment banking, and wealth management. Founded in 1865 in Hong Kong, HSBC has grown into one of the largest banks worldwide, serving millions of customers with a workforce of over 200,000 employees. HSBC's focus on international trade and finance has established its prominence in global markets, participating in major transactions. Despite its size, HSBC maintains a strong reputation for ethical and sustainable practices, earning numerous awards and recognition. Within HSBC, there is a dedicated credit risk department responsible for assessing and managing lending-related credit risk. The department employs various techniques, including internal ratings models, to evaluate credit risk based on factors like borrowers' financial history, industry sector, and economic conditions. Collaborating with other bank divisions, such as commercial and investment banking, the credit risk department ensures loans and investments are structured for optimal risk-adjusted returns. Playing a critical role in the bank's risk management framework, HSBC's credit risk department ensures prudent and effective risk exposure management.

I am currently working with the Wholesale credit risk model development team.

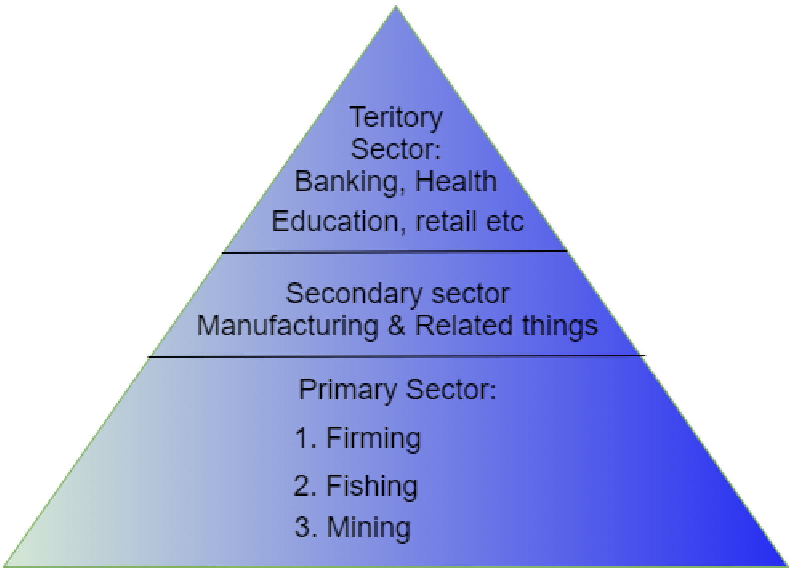
The background chapter provided an overview of the necessary prerequisites for the work, including essential terminologies used in the financial services. The second chapter delved into the Basel framework, highlighting different types of Internal Ratings-Based (IRB) models. Following that, we presented the problem statement. In the final chapter, we focused on discussing the outcomes and challenges encountered throughout the project.

Chapter 2

Background

2.1 Economy & Banking

The economy is commonly categorized into three sectors. The Primary sector involves activities such as farming, mining, and fishing, while the Secondary sector comprises manufacturing and related activities. The Tertiary sector, or service sector, includes a broad range of industries such as retail, banking, hotels, education, and health.



Money is the lifeblood of the economy and must be efficiently distributed among its various participants. Banks and other financial institutions serve as the essential connectors that facilitate the movement of funds. Banks act as financial intermediaries, collecting funds from depositors and lending them to borrowers in need. In this way, banks serve as a bridge between these two groups.

2.2 Business Models of Banks

- **Traditional Banking Model:** Banks that follow this model typically generate revenue by accepting deposits from customers and lending money out to borrowers at a higher interest rate. They also offer various services such as checking accounts, savings accounts, and loans.
- **Fee-Based Model:** Under this model, banks generate revenue by charging fees for various services such as ATM usage, overdrafts, wire transfers, and account maintenance. They may also charge fees for investment advisory services and wealth management.
- **Asset Management Model:** Banks that follow this model generate revenue by managing assets such as mutual funds, hedge funds, and other investment products. They typically charge a percentage of the assets under management as fees.
- **Investment Banking Model:** This model involves providing various investment banking services such as underwriting, mergers and acquisitions, and securities trading. Banks generate revenue through commissions and fees charged for these services.
- **Digital Banking Model:** This model involves offering banking services exclusively through digital channels such as mobile apps and online banking platforms. Banks generate revenue through fees charged for services such as money transfers, bill payments, and account management. They may also earn interest on deposits.

The traditional banking model is more relevant to credit risk since it involves the core activities of banking, which are accepting deposits from customers and lending money out to borrowers. Credit risk refers to the risk that a borrower will default on their loan or fail to repay it on time, which can lead to losses for the bank.
The traditional banking model is designed to manage this risk by using various credit risk management techniques, such as credit scoring, underwriting, and loan monitoring, to assess the creditworthiness of borrowers and ensure that loans are properly structured and priced based on the level of risk involved.
In contrast, other business models such as fee-based or investment banking models may not involve direct lending activities and may not be as focused on managing credit risk.

2.3 Risk

Banks face several types of risks in their operations. One common risk is the possibility that borrowers may not repay their loans or may repay them late, which can result in financial losses for the bank. This credit risk is an inherent part of the lending business, and effective credit risk management is essential to mitigate this risk.

Another risk that banks face is the possibility of faster withdrawals by customers than deposits, which can lead to a bank run. In such a scenario, customers rush to withdraw their deposits, which can deplete the bank's reserves and create a liquidity crisis. Banks must maintain adequate liquidity buffers and have contingency plans in place to address this risk.

Banks also face the risk that their investments may lose value, which can result in financial losses. This market risk is especially significant for banks that hold large portfolios of securities or engage in trading activities. Effective risk management practices such as diversification and hedging can help mitigate this risk.

Finally, interest rate fluctuations in the market can also pose a risk to banks. Changes in interest rates can impact the bank's profitability and financial stability. Banks must manage their interest rate risk by balancing their assets and liabilities and using hedging instruments to mitigate the impact of interest rate fluctuations.

Overall, effective risk management is critical for banks to ensure their financial stability and profitability in the face of these and other risks that may arise in the course of their operations.

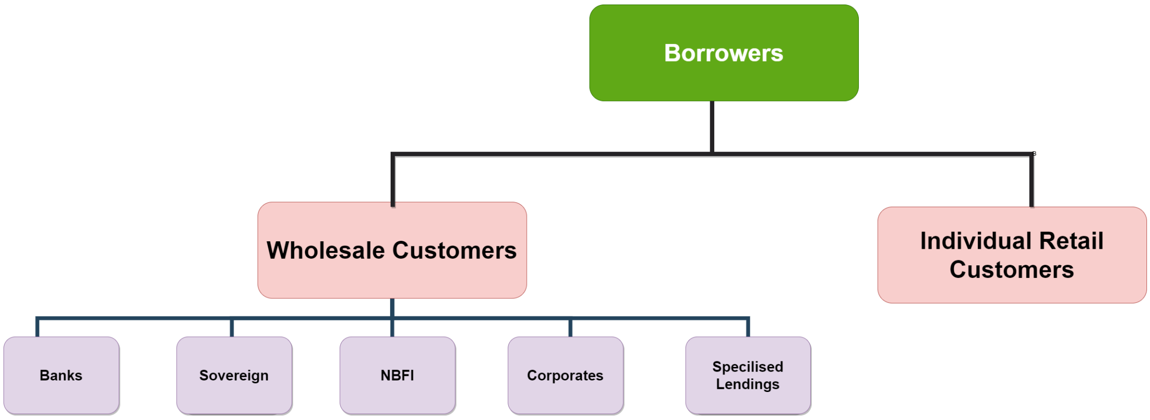
2.3.1 Types of Risk

- **Credit risk:** the risk of loss due to a borrower's failure to repay a loan or meet other credit obligations.
- **Market risk:** the risk of loss due to changes in market factors such as interest rates, exchange rates, commodity prices, and equity prices.
- **Operational risk:** the risk of loss due to inadequate or failed internal processes, systems, or external events such as fraud, human error, and system failures.
- **Legal risk:** the risk of loss due to lawsuits, adverse judgments, or legal sanctions.
- **Reputational risk:** the risk of loss due to damage to a bank's reputation, which can lead to loss of customers, business, and revenue.

2.4 Credit Risk

Credit risk refers to the risk that a borrower will default on their debt obligations or fail to repay the loan according to the agreed terms, which can result in financial losses for the lender. Credit risk is a significant concern for banks and other financial institutions that lend money to individuals and businesses.

2.4.1 Credit Risk Management



Credit risk management is a technique used by financial institutions to reduce the risk of losing money when lending to someone. This involves assessing the creditworthiness of the borrower and monitoring their ability to repay the loan.

Banks also establish loan loss reserves and maintain adequate capital to mitigate potential losses. By implementing effective credit risk management practices, financial institutions can minimize their exposure to credit risk and ensure the stability of their operations.

2.4.2 Wholesale & Retail Credit Risk

Wholesale credit risk and retail credit risk are two broad categories of credit risk that banks and financial institutions face.

Wholesale credit risk refers to the credit risk associated with lending to large corporate clients, financial institutions, and government entities. The loans in this category are typically large in size and involve complex financial transactions. Due to the large size of the loans and the complex nature of the borrowers, the credit risk involved in wholesale banking is higher than in retail banking.

In contrast, retail credit risk refers to the credit risk associated with lending to individual consumers, small businesses, and other retail customers. Retail loans are typically smaller in size and involve simpler financial transactions. Due to the smaller size of the loans and the lower complexity of the borrowers, the credit risk involved in retail banking is lower than in wholesale banking.

Effective credit risk management is crucial for both wholesale and retail banking. This includes credit analysis and underwriting, monitoring of loan performance, diversification of the loan portfolio, and the use of collateral and risk-based pricing to mitigate risk. The specific techniques used to manage credit risk may differ depending on the type of lending involved, but the overall goal is the same: to identify and manage credit risk in order to minimize financial losses and maintain a profitable lending business.

Wholesale credit risk is generally considered more important than retail credit risk due to the larger loan sizes and greater complexity of the transactions involved.

Lending to large corporate clients, financial institutions, and government entities in the wholesale banking sector involves much larger loan amounts than those in the retail banking sector. If one of these large borrowers' defaults, the potential financial losses for the bank can be significant, and may have a more pronounced impact on the bank's overall profitability and financial stability.

Furthermore, the complexity of the transactions involved in wholesale banking makes effective credit risk management even more crucial. The borrowers in wholesale banking typically have complex financial structures and business models, and their loans often involve complex financing arrangements such as syndicated loans or structured finance. This requires more detailed and specialized credit analysis and monitoring, which adds to the complexity and importance of wholesale credit risk management.

Overall, while both wholesale and retail credit risk are important for banks, wholesale credit risk is generally considered more significant due to the larger loan sizes and greater complexity of the transactions involved. Effective management of wholesale credit risk is therefore a key factor in the success and stability of a bank's lending business.

Chapter 3

Frame of Work

3.1 Three key components of credit risk

The probability of default (PD), loss given default (LGD), and exposure at default (EAD) models are the three key components of credit risk assessment. These models assist in estimating the likelihood of default, the potential loss in case of default, and the total exposure at the time of default.

- **Probability of Default (PD):**
  The Probability of Default (PD) refers to the likelihood that a borrower will default on a loan or credit obligation. PD models are employed to estimate this probability for individual borrowers or a portfolio of borrowers.

  By utilizing historical data and considering the borrowers' credit characteristics and other relevant information, PD models provide insights into the likelihood of default. Estimating PD and implementing internal ratings models are essential for effective credit risk management in banks.

  These models enable banks to make informed decisions, allocate capital efficiently, and maintain the stability of their loan portfolios. By understanding and managing PD, banks can enhance their risk assessment processes and mitigate potential credit losses.

- **Exposure at Default (EAD):**
  Exposure at default (EAD) is the total amount of credit exposure that a lender has at the time of default. In EAD modeling, financial institutions aim to accurately assess the potential loss they may incur if a borrower defaults.By considering the EAD, lenders can better understand and quantify their risk exposure, allowing them to allocate

capital efficiently, set appropriate credit limits, and make informed decisions regarding risk management and mitigation strategies.

The mathematical intuition behind EAD models is to use historical data on credit exposure to estimate the total exposure, based on the borrower's credit limit, utilization rate, and other relevant information.

- **Loss Given Default (LGD):**
  Loss given default (LGD) is the amount of loss that a lender incurs when a borrower defaults on a loan or other credit obligation. LGD models are utilized to estimate the potential loss for individual borrowers or a portfolio of borrowers.

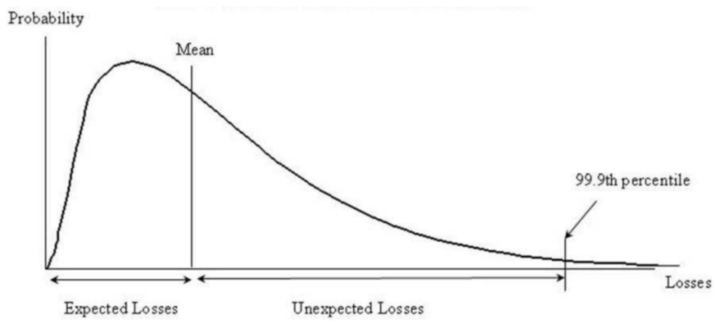
  In specific industries like e-commerce, LGD models help estimate potential losses arising from fraud or Disputes. Similarly, in the real estate sector, LGD models assist in estimating potential losses in case of project failure or market downturn.

  By accurately estimating LGD, financial institutions can better understand the potential financial impact of defaults, enhance their risk management practices, and make informed decisions regarding loan pricing, credit provisioning, and capital allocation. LGD modeling plays a critical role in assessing credit risk and ensuring the overall stability of lending portfolios.

3.1.1 Loan Example:Calculation of components

Let's consider a scenario where a bank has lent 6 lakhs to an individual. The borrower has already repaid 2 lakhs, leaving an outstanding loan amount of 4 lakhs, which represents the Exposure at Default (EAD) in this case.
Now, let's assume that the bank sells collateral associated with the loan, generating proceeds of 3 lakhs. This reduces the net loss to 1 lakh, as the collateral value partially offsets the outstanding loan amount.
To calculate the Loss Given Default (LGD), we divide the net loss (1 lakh) by the EAD (4 lakhs). In this case, the LGD is 25% (1 lakh / 4 lakhs).
Finally, to determine the expected loss, we multiply the Probability of Default (PD) by the EAD and the LGD. For instance, if the PD is 1 (indicating a certain default), the expected loss would be 1 lakh. Similarly, if the PD is 1/2 (indicating a 50% probability of default), the expected loss would amount to 50,000.
This calculation helps banks assess the potential loss they may face based

on the borrower's credit risk profile, the outstanding loan amount, the collateral value, and the probability of default. By understanding and monitoring these factors, banks can make informed decisions, manage credit risk, and ensure the soundness of their lending practices.
Unexpected loss, also known as unexpected credit loss or unexpected default loss, refers to the unforeseen or unexpected portion of potential losses that a financial institution may experience due to credit defaults. Unlike expected loss, which is based on estimated probabilities and predefined risk factors, unexpected loss represents the additional losses beyond what is anticipated.



While there isn't a specific equation for calculating unexpected loss, it is typically derived by subtracting the expected loss from the total potential loss or the worst-case scenario. Unexpected loss takes into account events or circumstances that may deviate from the expected probabilities or severity of default events.
Financial institutions consider unexpected loss as an important aspect of their risk management framework. It helps them quantify the potential impact of extreme or rare events that may have a significant adverse effect on their credit portfolios. By incorporating measures to mitigate unexpected losses, institutions aim to enhance their resilience and protect their capital base against unexpected credit risks.

3.2 Risk-Weighted Assets

Risk-weighted assets (RWA) play a crucial role in determining the minimum capital requirement for banks, taking into account the riskiness of their assets. Every asset held by a bank, including loans, investments, and off-balance sheet exposures, carries a specific risk weight that reflects its credit risk level.
For IRB Models the risk weights are determined from PD, LGD of the individuals/industrialist. Assets with higher risk are assigned higher weights, indicating the need for a greater allocation of capital as a buffer against

potential losses. This approach ensures that banks maintain sufficient capital to absorb risks and uphold their financial stability.
Calculating RWA provides a standardized framework for evaluating the risk sensitivity of a bank's balance sheet. This assessment assists regulators and stakeholders in assessing a bank's capital adequacy and facilitates comparisons between banks in terms of risk profiles. Ultimately, the objective of aligning capital requirements with risk-weighted assets is to foster a safer and more resilient banking system.

3.2.1 A Simplified Example

Let's assume a bank has the following assets with their respective risk weights:

- Cash and government securities: Risk weight of 0%
- Residential mortgages: Risk weight of 35%
- Corporate loans: Risk weight of 100%

Now, let's consider the values of these assets, each set to 10 lakhs:

- Cash and government securities: 10 lakhs
- Residential mortgages: 10 lakhs
- Corporate loans: 10 lakhs

To calculate the risk-weighted assets, we multiply the value of each asset by its respective risk weight:

- RWA for cash and government securities = 10 lakhs * 0% = 0
- RWA for residential mortgages = 10 lakhs * 35% = 3.5 lakhs
- RWA for corporate loans = 10 lakhs * 100% = 10 lakhs

Total risk-weighted assets (RWA) = 0 + 3.5 lakhs + 10 lakhs = 13.5 lakhs
So, in this example, the bank's total risk-weighted assets amount to 13.5 lakhs. This value represents the credit risk exposure of the bank's asset portfolio and is used to determine the minimum capital requirements.

3.3 Capital Adequacy Ratio

The Capital Adequacy Ratio (CAR) is a metric utilized to evaluate the financial soundness of a bank and its capacity to absorb potential losses. It is a regulatory standard aimed at guaranteeing that banks maintain an adequate level of capital to support their operations and safeguard the interests of depositors and creditors.

Expressed as a percentage, the CAR is calculated by dividing a bank's capital, which includes both Tier 1 and Tier 2 capital, by its risk-weighted assets (RWA). The CAR formula is as follows:

$$\frac{(Tier1\ Capital + Tier2\ Capital)}{RWA}$$

The components of capital in the CAR framework are categorized into Tier 1 and Tier 2 capital. Tier 1 capital primarily comprises the bank's core equity capital, such as common stock and retained earnings, representing the highest quality capital. Tier 2 capital encompasses subordinated debt, preferred stock, and other forms of capital that are comparatively less secure.

Regulatory bodies, like the Basel Committee on Banking Supervision, establish minimum CAR requirements that banks must adhere to in order to ensure financial stability. These requirements may vary across jurisdictions and are designed to ensure that banks possess a sufficient cushion to absorb losses and maintain the confidence of depositors and other stakeholders.

A higher CAR signifies a stronger capital position for a bank, indicating its ability to handle potential losses. It demonstrates the bank's resilience in times of financial stress and its ability to fulfill its obligations. Banks typically maintain CARs above the minimum regulatory thresholds to operate securely and soundly within the banking system.

3.4 Basel Norms

The Basel Committee on Banking Supervision (BCBS) is an international organization that establishes global standards for banking regulations. The Basel accords, developed by the BCBS, provide guidelines for banks in areas such as capital adequacy, risk management, and liquidity. These accords aim to enhance the strength and stability of banks, safeguard depositors, and promote financial stability.
Basel I, Basel II, and Basel III represent successive versions of the accords, addressing evolving risks and vulnerabilities in the banking sector. While the adoption of Basel norms is widespread, their implementation varies across countries as they incorporate them into their respective banking frameworks. The accords foster consistency, transparency, and a level playing field in international banking, contributing to the resilience and soundness of the global financial system.

3.4.1 Basel I

Basel I, introduced in 1988 by the Basel Committee on Banking Supervision (BCBS), is the initial set of global banking regulations. It was created to address concerns about insufficient capital levels in banks and the need for a standardized risk management approach. Basel I established a minimum capital requirement for banks, focusing primarily on credit risk in their lending activities.

Under these regulations, banks were mandated to maintain a minimum capital adequacy ratio (CAR) of 8%, with a portion of the capital reserved to cover credit risk. This framework provided a foundational structure for measuring and managing risk in the banking industry, fostering stability and instilling confidence among depositors and creditors.

Basel I played a crucial role in establishing a uniform international standard for capital adequacy and risk management. By setting minimum capital requirements, it aimed to mitigate the risk of bank failures, strengthen banks' ability to absorb losses, and contribute to financial stability. The objective was to create a level playing field among banks, improve the overall health of the global banking system, and safeguard the interests of depositors and stakeholders.

Basel I had limitations in risk-weighting assets, overlooking different risk levels within asset classes. It focused on credit risk and ignored market and operational risks. The inflexible system lacked customization and the 8% minimum capital adequacy ratio was seen as insufficient. Uneven implementation raised concerns about international competitiveness. Subsequent Basel versions aimed to address these issues and strengthen banking supervision.

3.4.2 Basel II

Basel II is the second set of international banking regulations developed by the BCBS to enhance the risk management and capital adequacy standards for banks. It was introduced in 2004 as an improvement over Basel I.
Basel II was necessary to address the limitations of Basel I and further strengthen the banking system. The key objectives of Basel II were to provide a more risk-sensitive framework, align regulatory capital requirements with actual risk profiles, and promote better risk management practices. It introduced three pillars:

- Minimum Capital Requirements: Basel II refined the calculation of risk-weighted assets (RWA) by introducing more risk-sensitive approaches, taking into account credit, market, and operational risks.
- Supervisory Review Process: It emphasized the importance of strong risk management practices and required banks to undergo a comprehensive assessment of their risk profiles and internal controls.
- Market Discipline: Basel II aimed to promote transparency and disclosure by encouraging banks to provide more comprehensive information about their risk profiles and capital adequacy.

Basel II had limitations including complexity in implementation, pro-cyclicality, variability in interpretation, and inadequate focus on liquidity risk. The framework's complexity made compliance challenging for banks and regulators. Pro-cyclicality meant banks had to raise capital during economic downturns, potentially worsening financial stress. Differences in interpretation and application across jurisdictions led to inconsistencies.
Basel II also placed less emphasis on liquidity risk, which became evident during the 2008 financial crisis. These limitations prompted the development of Basel III, introducing measures to strengthen capital adequacy, liquidity risk management, and leverage ratios, aiming to enhance stability and resilience in the banking sector based on lessons learned.

3.4.3 Basel III

Basel III is a set of international banking regulations developed by the BCBS as an enhancement to the previous Basel II framework. Introduced in response to the global financial crisis of 2008, Basel III aims to strengthen the resilience of the banking sector and promote financial stability.

The key objectives of Basel III are to improve the quality and quantity of capital held by banks, enhance risk management and governance practices, and address liquidity risk and leverage. It introduces stricter capital requirements, with a focus on common equity Tier 1 (CET1) capital, to ensure banks have a more solid capital base to absorb losses during periods of financial stress.

Basel III also introduces new liquidity standards, such as the Liquidity Coverage Ratio (regulatory requirement that measures a bank's ability to meet its short-term liquidity needs.) and the Net Stable Funding Ratio (regulatory measure that assesses a bank's long-term stability and funding adequacy.), to promote better liquidity risk management and reduce the risk of funding shortfalls during market disruptions.

By implementing Basel III, regulators aim to improve the overall resilience and stability of the banking industry, enhance risk management practices, and protect the interests of depositors and creditors. It provides a comprehensive framework that sets higher standards for capital adequacy, liquidity risk management, and leverage, contributing to a more robust and secure global banking system.

Chapter 4

Migration from MATLAB to Python

In this chapter, I addressed the problem statement of migrating Internal Ratings Based (IRB) models from MATLAB to Python at HSBC. The objective was to reduce software costs and upgrade to the latest technology stack while maintaining the accuracy and reliability of the models. I outlined the steps I took to solve the problem and discussed the challenges encountered during the solution process.

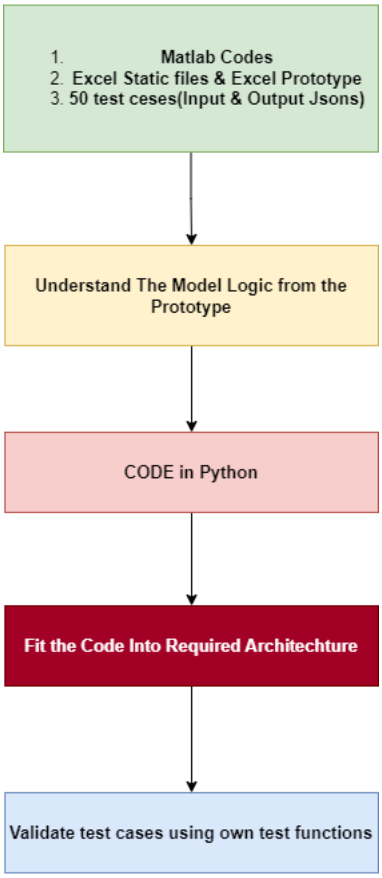
4.1 Problem Statement

The implementation of Internal Ratings Based (IRB) models is a crucial aspect of banking regulation to ensure adequate capitalization and preparedness for economic scenarios. To assess the reliability and accuracy of these models, banks must regularly evaluate their accuracy based on predefined parameters.
Traditionally, the bank has used proprietary software like MATLAB for implementing IRB models, resulting in significant costs. To reduce these costs and upgrade to the latest technology stack, HSBC has initiated a project to migrate their IRB models from MATLAB to Python. However, little research has been done on the appropriateness of Python in implementing IRB models, and it is unclear whether the migration to Python will maintain the accuracy and reliability of the IRB models.

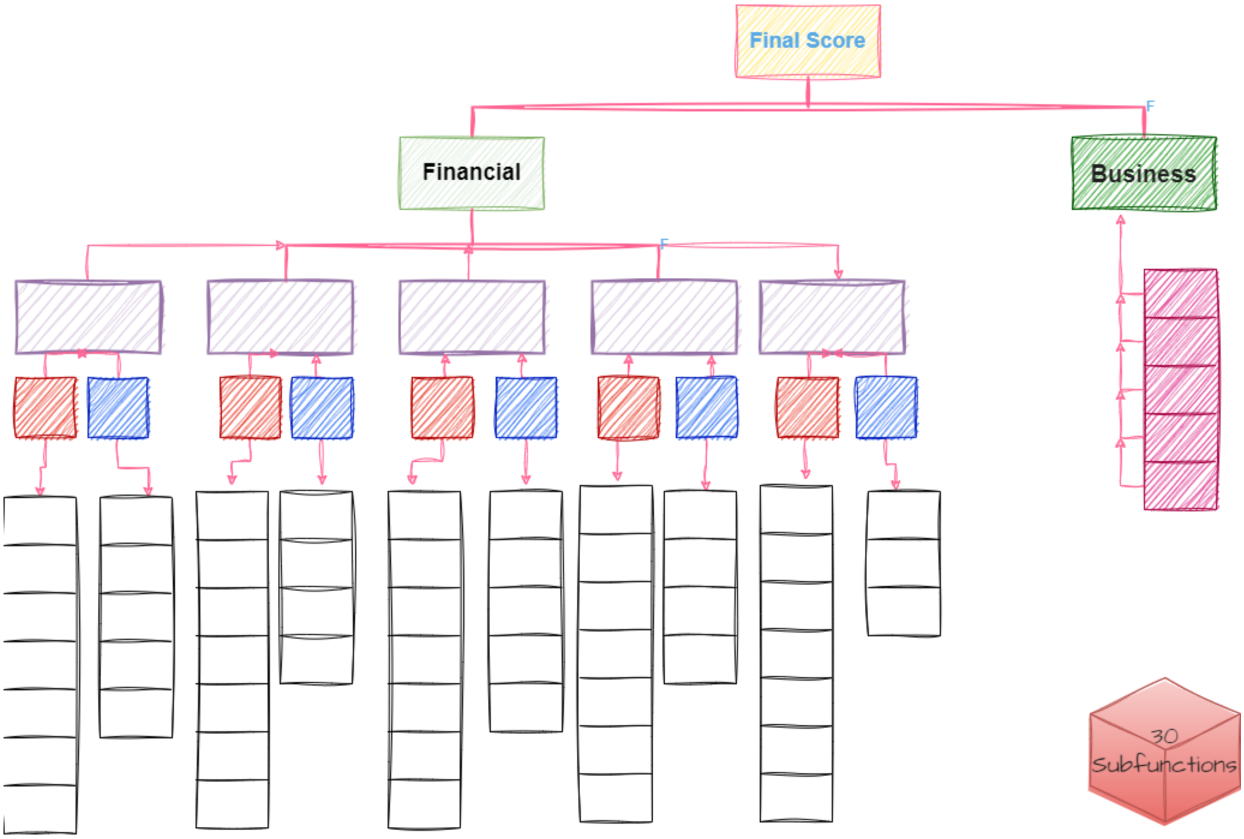
Thus, the problem statement of this thesis is to evaluate the effectiveness of Python as a tool for implementing IRB models and to assess the accuracy and reliability of the IRB models after the migration from MATLAB to Python. This research will contribute to the understanding of the role of open-source tools in the financial services industry and inform best practices for implementing IRB models using Python.

4.2 Key Steps

I was given a set of resources for my project, which included MATLAB codes, Excel static files, and an Excel prototype. Alongside these resources, I also received 50 test cases consisting of input and output JSON files. The initial task was to thoroughly comprehend the code implementation demonstrated in the prototype. Once I gained a clear understanding, the next step involved translating the code into Python. Finally, to ensure the accuracy and correctness of the Python implementation, extensive testing was conducted using the provided test cases. This comprehensive process aimed to validate the functionality and reliability of the Python code.



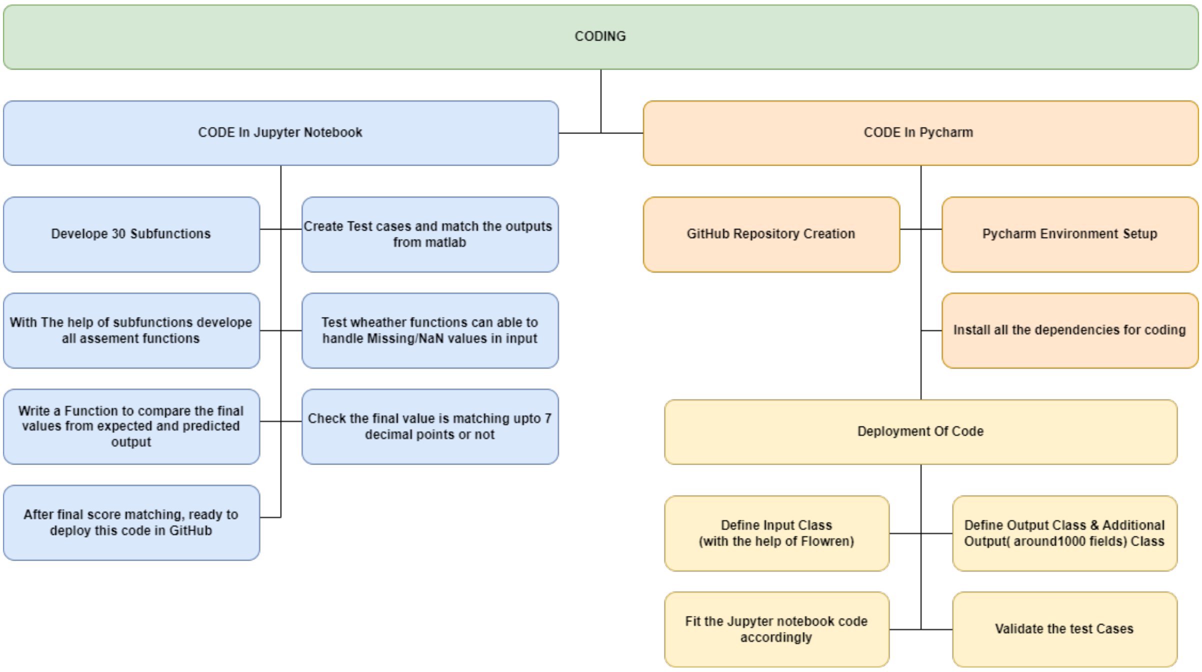
4.3 Model Structure Overview



While thoroughly analyzing the prototype, I came to realize that deriving the final score entails an extensive examination of the Financial and Business Analysis section. This critical component encompasses a series of intermediate assessments showcased in the comprehensive flowchart. Moreover, these internal assessments themselves encompass additional layers of intermediate assessments.

To accurately calculate the values required for these assessments, the utilization of specific sub functions becomes imperative. These sub functions serve as indispensable tools, enabling us to effectively compute the necessary values and contribute to the overall analysis process. By harnessing the power of these sub-functions, we can systematically calculate the values crucial for the intermediate assessments and ultimately determine the precise final score.

4.4 Coding Journey



First, I began coding in Jupyter Notebook and created 30 subfunctions to calculate various assessments. Next, I validated these subfunctions by comparing their outputs with existing MATLAB functions using test cases. I then developed the assessment functions to calculate the final score. After obtaining the final score, I checked its accuracy by comparing it with the expected value up to 7 decimal points. Finally, I deployed the code on GitHub for further collaboration and version control.

To deploy our code on GitHub, we collaborated with the Financial Engineering Team who provided instructions and architectural guidelines. They developed an architecture and a framework similar to Flowren (comparable to Pydantic) that assisted us in creating the Input and Output classes. We implemented our code within their architecture and utilized API calls to execute it. We defined the necessary classes and marked certain input fields as Optional based on whether they were mandatory for our calculations. We conducted a batch test to validate all intermediate and final scores, ensuring the accuracy of our implementation.

4.4.1 Pydantic

Pydantic is a Python library for data validation and parsing. It allows you to define data models using Python classes with type annotations, enabling easy validation and enforcement of data types and constraints.
With Pydantic, you can validate input data, perform data parsing and serialization, and generate interactive documentation. It supports optional fields, automatic parsing and serialization between different formats, and leverages Python's type hinting system for better code readability and IDE support.
Pydantic is widely used in web development, data processing, API development, and configuration management, thanks to its simplicity, flexibility, and integration with Python's ecosystem.

```python
from pydantic import BaseModel, validator
from typing import Optional


class Person(BaseModel):
    name: str
    age: int
    email: Optional[str]  # Optional field


@validator('age')
def validate_age(cls, age):
    if age < 0:
        raise ValueError("Age must be a positive integer.")
    return age


# Create an instance of the Person class
person_data = {
    'name': 'Sourav Ghosh',
    'age': 25,
    'email': 'sourav@example.com'
}
person = Person(**person_data)

print(person)
#name='Sourav Ghosh' age=25 email='sourav@example.com'
```

Listing 4.1: Pydantic example

4.5 Challenges

The challenges faced during the project included:

- Performance warning: DataFrame fragmentation due to multiple calls to frame.insert, resulting in poor performance. The suggestion was to use pd.concat(axis=1) to join all columns at once and improve performance.

- Division by zero: Unlike Matlab, Python does not handle 0/0 as NaN by default. Handling this required implementing appropriate mechanisms to avoid division by zero errors.

- Test case failures: The presence of performance warnings caused some test cases to fail on GitHub. Addressing the warnings allowed the test cases to pass successfully.

> Note
>
> **Cognitive Complexity:**
> Cognitive complexity refers to how difficult it is for humans to understand and reason about a piece of code. It measures the complexity of a program's logic, control flow, and overall structure.
>
> In software development, managing cognitive complexity is important because it affects the quality, maintainability, and readability of code. High cognitive complexity can make it harder to understand, debug, and modify code, increasing the chances of errors and reducing reliability.
>
> Cognitive complexity is influenced by factors like nested conditionals, loops, function length, variable scope, and code repetition. By reducing cognitive complexity through refactoring, simplifying code, and following best practices, developers can improve code clarity and ease of maintenance.
> .

- Cognitive complexity: The cognitive complexity of functions needed to be reduced below 15 to merge the branch with the master. This required refactoring and simplifying the code structure to improve readability and maintainability.

Chapter 5

Conclusion & Future Work

5.1 Conclusion

In conclusion, the problem of migrating Internal Ratings Based (IRB) models from MATLAB to Python at HSBC has been successfully addressed. The objective of reducing software costs and upgrading to the latest technology stack while maintaining the accuracy and reliability of the models has been achieved.

Throughout the solution process, I thoroughly understood the existing MATLAB code implementation and the underlying mathematical models used in the IRB models. By carefully translating the code into Python, I ensured that the functionality and results remained consistent.

Extensive testing using a comprehensive set of test cases was conducted to validate the correctness of the Python implementation. This testing phase confirmed that the migrated code produced accurate results, thereby maintaining the reliability of the IRB models.

However, the solution was not without its challenges. Performance warnings related to DataFrame fragmentation, division by zero errors, and cognitive complexity required attention and refinement. By addressing these challenges through appropriate optimizations and refactoring, the code's performance, readability, and maintainability were improved.

Overall, this project has provided HSBC with a cost-effective solution by migrating IRB models to Python. By leveraging the open-source nature of Python and its rich ecosystem of libraries and tools, HSBC can benefit from the latest technologies while ensuring the accuracy and reliability of their risk assessment models.

The successful completion of this project highlights the advantages of adopting open-source tools like Python in the financial services industry. It demonstrates how such technologies can help banks enhance their capabilities, reduce costs, and stay aligned with industry standards and best practices. The migration from MATLAB to Python sets a foundation for further technological advancements and innovations within HSBC's risk management practices.

5.2 Future Work

- **Documentation and Code Maintenance:** Creating comprehensive documentation that explains the code's functionality, structure, and usage can facilitate collaboration and future maintenance. Additionally, regular code reviews, refactoring, and adherence to coding best practices will contribute to maintainable and readable code.
- **Expand Test Coverage:** While 50 test cases were provided and validated, it is advisable to expand the test coverage to encompass a wider range of scenarios. This can help uncover edge cases and potential issues that were not initially considered. Comprehensive testing ensures the accuracy and reliability of the Python implementation.
- **Integration and Deployment:** Collaborating with the Financial Engineering Team to integrate the code within their architecture and utilizing API calls was a significant step. Further work can involve enhancing the integration process, ensuring smooth deployment, and considering automation techniques for continuous integration and delivery.

References

[1] https://www.eba.europa.eu/regulation-and-policy/model-validation/guidelines-on-pd-lgd-estimation-and-treatment-of-defaulted-assets

[2] https://www.bis.org/bcbs/index.htm

[3] HSBC Training Sessions and Seminars in Wholesale Credit Risk Management.