Transition Risk Models for Loan Portfolio Stress Testing with Respect to Climate Scenarios

Rui Qin, PhD **Ashwani Roy** **Dr. Arpad Borock**

Angel Lane Partners (ALP) Analytics Ma[illegible] 14, 2024

Contents

Chapter 1

Introduction

1.1 | Introduction to Climate Transition Risk

Climate change presents significant challenges to human society and global economic activity. Existing national and international climate related policies will result in a global temperature increase ranging from 2.1°C to 3.9°C by 2100, compared to pre-industrial levels. The global community is increasingly becoming aware that climate change not only poses risks to the environment but also creates significant financial risks and economic consequences, thereby affecting the overall stability of the financial system.

As extreme weather events become more frequent and intense, businesses face heightened vulnerabilities, from supply chain disruptions, property damages, escalating operational costs, etc. This is referred to as the physical climate risk. However, the transition risks, i.e. the financial stress that the companies face as they transition to a low carbon economy, are equally important and need to be identified, quantified and managed.

Financial institutions, including banks, have started to recognize the critical importance of proactive risk mitigation on their asset portfolios. In many ways, Climate Risk Stress Testing on financial institutions' asset portfolios is the first step in understanding and dimensioning the scale of the potential stress on balance sheets, which ultimately leads to a Capital Adequacy problem - something that the Central Banks have had intense focus on in recent times.

To limit global warming to 1.5°C above pre-industrial levels, society as a whole, must accelerate the decarbonization process, collectively achieving net-zero emissions by 2050. The Basel Committee also issued a consultation paper seeking feedback on a potential Pillar 3 disclosure framework for climate-related financial risks. The goal is to promote comparability of banks' climate risk profiles and enable market participants to better understand banks' exposures. Within the strategic initiative of attaining Net Zero by 2050, governments have undertaken to allocate substantial resources and create a robust policy framework to reduce greenhouse gas (GHG) emissions. This undertaking has significant implications for key economic sectors, encompassing energy, transportation, mining, manufacturing, etc.

However, modelling for transition and physical risk is a complex exercise and have a number of inter-dependent variables that will need to considered: For example, a financial institution will need to understand and project the following before an effective stress testing model can be implanted:

1. **Identifying potential transmission channels to financial stress**: This is a critical consideration. There could be multiple ways that a firm is exposed to transitional climate related financial stress. For example, following are some of the transmission channels:

- A firm may be subject to Carbon Prices, negatively impacting EBITDA
- Suffer a loss in market share because of changes in client preferences
- Significant capital outlay to transition to greener technologies without corresponding incremental revenues
- Stranded assets as a result of regulatory changes

2. **Project the sources of risk climate action pathways**: There are three major sources of transition risk that will need to be accounted for and projected. These are discussed below:

    - Local Regulatory Changes: This includes any regulatory changes that are enforced by a country, e.g., levying carbon prices generally or on specific sectors. These could also take the form of fossil fuel excise taxes etc. These changes can potentially be projected in the context of country's Nationally Determined Contributions ("NDC"). In other words, the Climate Action Pathways adopted or expected to be adopted by the local regulator are an important consideration.
    - International Regulatory Changes: Measures such the Carbon Border Adjustment Mechanism ("CBAM") that will become effective in the EU in 2026, essentially makes non-EU jurisdiction subject to these regulations and hence to the transition risk. Identifying clients, potentially exposed to international regulations in effect or becoming effective at a future date is, critical and must be incorporated in the stress testing model.
    - Climate Related Economic Trends: This area is quite broad but important to consider. This attempts to identify industry or consumer behaviour trends that could negatively impact a particular client. In other words, this is about identifying Early Warning Signals ("EWS") on how the global efforts to combat climate change and the rising awareness could impact an industry generally and a particular client specifically.

      For example, global banks decided to stop funding coal fired power generation plants restricting access to liquidity for the affected companies. Airline industry continues to be under pressure to reduce emissions albeit without any viable solution. It is important for risk managers at banks to be fully aware of such developments.

3. **Emissions data and clients' disclosures**

    - While the climate action pathways are important, the starting point is to make sure that the data related to current emissions is accurate to the extent possible. In addition to relying on publicly and privately reported data, the issue of gaps in data can be addressed by clustering and distance algorithm, which is discussed in later sections.

4. **Companies' plans and initiatives to reduce Emissions**

    - While the stress tests can initially be run on the basis of companies maintaining status quo in terms of efforts to reduce Emissions, this is not practical. It is fair to assume that the firms will undertake any number of steps or projects to reduce GHGs and these in turn, would have financial implications.

There could be several avenues to achieve this. For example:

i. A company can opt to borrow more funds to enhance its carbon emission reduction initiatives. This, however, leads to an increase in leverage and a subsequent rise in the probability of default (PD).

ii. Secondly, acquiring a company specializing in green practices emerges as a viable strategy, resulting in a reduced carbon-to-revenue ratio. Yet, this endeavor also often involves borrowing more, contributing to increased leverage and deteriorating PD.

iii. Regulatory change could force companies to offset their emissions by purchasing credits. This approach requires additional expenditures and, consequently, increased borrowing.

Regulators like Central Bank of United Arab Emirates (CBUAE) and Bank of Canada (BoC) are keen for the financial institutions, regulated by them, to undertake stress testing exercise and also disclose the results and approach as a first step to strengthen banks' risk management practices and build capabilities towards resilience to climate-related financial risks.

From the perspective of the financial institutions, the ultimate objective of these stress tests is to quantify the impact of transition risks on the PD of the firms they have exposure to. Like all other stress tests, the impact of climate risk will need to be assessed under different scenarios which typically reflect different action pathways to achieving net zero 2050 objective. Once quantified, navigating these financial risks requires a strategic and forward-thinking approach from financial institutions to foster a smooth transition towards a sustainable and low-carbon future.

1.2 | Introduction to Climate Risk Scenarios

Climate-related risks are subject to significant uncertainty in terms of their timing, frequency and severity. Despite this uncertainty, forward-looking assessment approaches are crucial to effectively address these risks. Scenario analysis is a pivotal tool for evaluating the potential impact of climate change on economies and financial systems. It's important to note that these scenarios are not intended to make forecasts or make comprehensive predictions. Rather, they explore different plausible global transition pathways consistent with achieving specific climate targets. These scenarios take a conservative approach and include technologies that are not currently commercially available or may face scalability challenges in the future.

Essentially, there could be any number of different way of conducting stress tests that should reflect all or some of the variables discussed above.

In this paper we discuss the implementation of two approaches that have also been proposed by Central Bank of United Arab Emirates ("CBUAE") and Bank of Canada ("BoC"). Both theses approaches rely on different climate scenarios developed by the Network of Central Banks and Supervisors for Greening the Financial System ("NGFS") but differ in terms of the transition mechanism. While BoC's approach is based on modelling the change in market share of firms operating in a particular sector in a certain geographic area, CBUAE looks to model the impact on EBITDA as a result of the carbon price.[1]
Climate Action Pathway Scenarios:

- Current Policies (+3.0°C): Baseline scenario, with no or little change of current policies to combat climate change, causing high physical risks but minimal transition risk with global temperature increase by more than 3 degrees;
- Net Zero 2050 (+1.5°C): Coordinated global policy implementation (Paris Agreement) to combat climate change and limit the global temperature increase to 1.5 degrees, implying moderate transition risks;
- Delayed Transition (+2.0°C): Delayed policy implementation creates a "Minsky Moment" with high transition risks.

[1] Please refer to the Network for Greening the Financial System https://www.ngfs.net/ngfs-scenarios-portal/

- Immediate Transition (+2.0°C): Immediate collective global action is taken to reduce emissions toward a target of below 2°C.

In all scenarios, the assumption is a global collective effort to reduce emissions. In the immediate transition scenario, this action is assumed to have commenced in 2020, while the delayed transition scenario envisions action starting only in 2030. Due to the postponement in the latter case, emissions must decrease rapidly, post 2030, to compensate for lost time and offset the additional emissions linked to the delay, necessitating a significant transition by mid-century. The emission trajectories for both the below 2°C immediate and delayed scenarios are derived from countries' Nationally Determined Contributions ("NDC") submissions, adjusted to align with the ambition and timing of each respective scenario.
In the following chapters, we will mainly focus on these scenarios, to analyse the climate risk impact on different portfolios and companies.

1.3 | Introduction to Methodologies

In assessing the impact of climate risk, regulatory authorities including CBUAE delineate two potential methodologies that banks can employ to conduct thorough analyses of transition risk scenarios. These strategic approaches aid banks in proactively identifying, assessing, and addressing potential risks arising from the transition to a low-carbon economy.

1. Bottom-up approach: Bottom-up analyses are based on issuer-specific data. The target is to process a detailed analysis on banks' top or largest non-financial corporate (NFC) exposures, within pre-defined economic sectors that are subject to high transition risk.

2. Top-down approach: Top-down analyses are characterized by estimating financial losses at the level of portfolios, sectors or institutions. No distinction is made between assets or issuers within the same sector. Banks can leverage and build on their IFRS 9 models to analyse the sectoral impact from climate risk scenarios.

Within our report, Chapter 2 will introduce a Top-Down Approach in transition risk scenario analysis. This chapter primarily centres on the market shock under different climate scenarios at the sector level and examining their impact on portfolio valuation. In Chapter 3, we will introduce a Bottom-Up Approach to assess the carbon cost sensitivity of individual companies across different climate scenarios. This chapter aims to illustrate how the carbon cost will influence default probability.

Chapter 2

Market Shock Approach : A Top-Down Approach in Transition Risk Scenario Analysis

2.1 | Valuation of the Loan Portfolio

We consider a certain loan portfolio. We denote $A_j(t_0, T_j)$ as the financial valuation of loan j at time t_0, and T_j is denoted as the maturity of the loan j. Then, the valuation of the loan portfolio at time t_0 is written as

$$A(t_0) = \sum_j A_j(t_0, T_j). \tag{2.1.1}$$

For simplicity, we consider the expected value of the loan at time t_0 as the valuation of the loan at that time,

$$A_j(t_0, T_j) = p_j(t_0, T_j) r_j F_j + (1 - p_j(t_0, T_j)) F_j = F_j - F_j(1 - r_j) p_j(t_0, T_j), \tag{2.1.2}$$

given the recovery rate as r_j, the face value of the loan j as F_j, and the probability of the default of loan j at time t_0 as $p_j(t_0, T_j)$.

Now, we introduce a policy shock at time t_0. Here, the policy shock implies that the economy switches from a business-as-usual scenario, which we denoted as the Baseline (B), to a new scenario P, because of the change of the policy. This will lead to the change in the valuation of the loan. Keeping same r_j and F_j, we have

$$\Delta A_j(t_0, T_j, P) = A_j(t_0, T_j, P) - A_j(t_0, T_j, B) = -F_j(1 - r_j) \Delta p_j(P), \tag{2.1.3}$$

where $\Delta p_j(P)$ denotes the difference between the default probability (PD) given a change in the scenario from B to P. This $\Delta p_j(P)$ change is the negative of the change in valuations for the loan. That is, if the value goes down, it is due to an increase in provisions driven by the change in probability.

Now the question is, HOW does the change of the scenario, or we say, change of the policy lead to the change in the PD $\Delta p_j(P)$, in other words, what is the transmission mechanism to the financial stress.

2.2 | Compute the Market Share and the Market Shock

Let's go back to the data obtained from Bank of Canada to see which variable or variables will be influenced by the change of the policy from B to P.

Bank Of Canada Climate Scenario Data (BoC dataset) contains energy information spanning multiple sectors across global regions. Within the BoC dataset, we concentrate on the Primary Energy market value, which encompasses seven distinct forms of primary energy, i.e. (i) bio-energy, (ii) coal, (iii) gas, (iv) hydro, (v) nuclear, (vi) oil, and (vii) renewable sources such as wind and solar. These energy sector market sizes are aggregated based on various geographic regions, including Africa, Canada, China, Europe, India, Japan, the United States, and the rest of the world.

Similar to the NGFS Scenarios discussed in Chapter 1, BoC dataset analysis introduces four policy scenarios that will be rewritten here as:

- *Baseline* (2019 policies) — baseline scenario consistent with climate policies in place at the end of 2019.
- *Below 2°C Immediate* — immediate policy action scenario to limit average global warming to below 2°C.
- *Below 2°C Delayed* — delayed policy action scenario to limit average global warming to below 2°C.
- *Net-Zero* 2050 (1.5°C) — more ambitious immediate policy action scenario to limit average global warming to 1.5°C that includes current net-zero commitments by some countries.

These policies are implemented over the period spanning from 2020 to 2050, with a 5-year interval. In addition to the current market size, BoC dataset also includes data on market sizes across diverse geographic regions under various policy scenarios for specific years. In Figure 2.2.1, we present select examples of the data. Taking the geographic sector of Africa as an example, we illustrate the market size of the bio-energy sector under various climate scenarios over the time-frame from 2020 to 2050. It is evident how the market size of the bio-energy sector changes, prompting an exploration into the potential impact on loans extended to the companies in this sector.

Knowing the market size of each energy sector, it is straightforward to obtain the Market Share of each energy sector. To be more specific, if we take the primary energy sector *Coal* in geographic region *Global* under *Baseline* policy in 2020, the Market Share of *Coal* is

$$\text{Market Share}(\mathit{Coal}, \mathit{Global}, \mathit{Baseline}, 2020) = \frac{\text{market size}(\mathit{Coal}, \mathit{Global}, \mathit{Baseline}, 2020)}{\text{total market size } (\mathit{Global},\ \mathit{Baseline},\ 2020)} \times 100, \quad (2.2.1)$$

where the total market size $(\mathit{Global},\ \mathit{Baseline},\ 2020)$ is the total market size of all primary sectors in *Global* under *Baseline* policy in 2020.

	CL_GEOGRAPHY	CL_VARIABLE	CL_YEAR	Baseline (2019 policies)	Below 2°C delayed	Below 2°C immediate	Net-zero 2050 (1.5°C)
0	Africa	Primary Energy | Bioenergy	2020.0	15.5018	15.5018	15.5018	15.5018
1	Africa	Primary Energy | Bioenergy	2025.0	15.3023	15.3023	15.3023	15.3023
2	Africa	Primary Energy | Bioenergy	2030.0	15.2206	15.2206	15.2027	15.2916
3	Africa	Primary Energy | Bioenergy	2035.0	15.0715	15.0841	15.0422	15.2178
4	Africa	Primary Energy | Bioenergy	2040.0	15.0163	15.2305	15.0554	15.6312

Figure 2.2.1: BoC dataset example

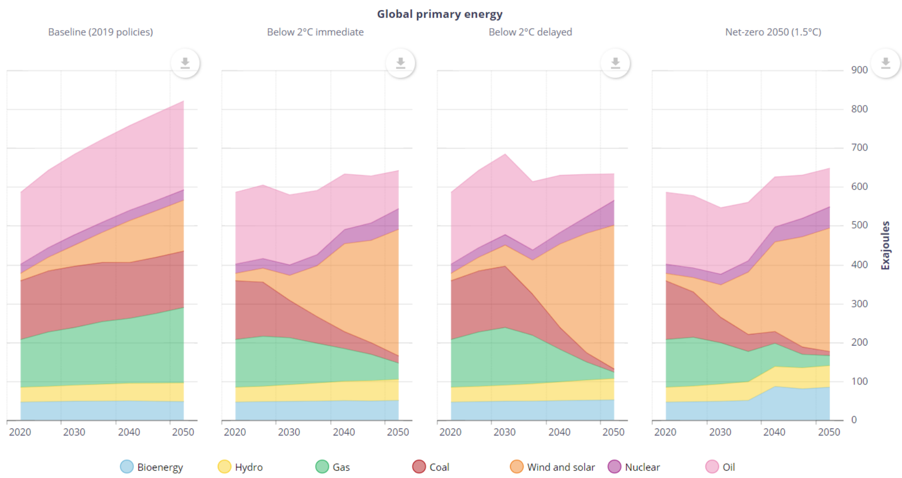


Figure 2.2.2: Global primary energy by source in four policy scenarios.
Image source: https://www.bankofcanada.ca/2022/01/climate-transition-scenario-data/

Figure 2.2.2 extracted from the Climate transition scenario data report published by BoC shows global changes in the primary energy Market Share over time under different policy scenarios. It is evident that adherence to the *baseline* policy will lead to a gradual expansion in the overall market size over time, with the Market Share of traditional energy sectors such as oil, coal, and gas also showing growth. Conversely, opting for the *Net-Zero* policy will result in fluctuations in the total market size, while the Market Share of traditional energy sectors experiences a notable decline, accompanied by an increase in the Market Share of clean energy sources like renewable energy (wind and solar) and bio-energy.

So it is important that we quantify the change of the Market Share under different policy scenarios compared with the *Baseline* scenario. To derive this we introduce the Market Shock, which is defined as the change in a particular sector's Market Share under different scenarios and is given by the formula 2.2.1:

$$\text{Market Shock}(P) = \frac{\text{Market Share}(P) - \text{Market Share}(B)}{\text{Market Share}(B)}, \tag{2.2.2}$$

where we compute the Market Share(P) and the Market Share(B) using the data of the same energy sector from the same geographic region in the same year. Thus, the Market Shock(P) represents the percentage change in the Market Share if policy scenario P is adopted compared with the Market Share under the *Baseline* policy.

2.3 | From Market Shock to the Change in Valuation

Now we are going to discuss how the Market Shock will influence the change in the PD. Intuitively, a decrease in the Market Share of a sector and consequently the companies in that sector, will lead to some level of financial stress, potentially leading to a rise in the PD. This, in turn, will negatively impact the valuation of loans to the sector.

In other words, a relative change in the PD is inversely proportional to the shock in Market Share. What must be kept in mind is that the PD of every company will not respond to the Market Shock to the same extent. Put differently and borrowing the concept from physics, the elasticity of the transmission mechanism from Market Shock to change in PD is critical to determine, recognizing the fact that the PD of every firm, even within the same sector can exhibit different elasticity. Determining this elasticity parameter, currently sits outside of our model. However, we are in the process of developing advanced models for the elasticity parameter χ, aiming to capture the market dynamics and accommodate more complex scenarios.

We introduce the elasticity parameter, χ, to denote this elasticity, and make the simplifying assumption that each firm responds to the market stress it encounters by yielding the same proportion of changes in PD.

$$\Delta p_j(P) = -\chi \text{Market Shock}(P). \tag{2.3.1}$$

Then, plugging the equation 2.3.1 into equation 2.1.3, we obtain the change in the expected value of the loan, conditional to a Market Shock when we change the policy scenario B to scenario P:

$$\Delta A_j(P) = F_j(1 - r_j)\chi \text{Market Shock}(P). \tag{2.3.2}$$

Summing all the loans j in the portfolio, we obtain the total valuation change as

$$\sum_j \Delta A_j(P) = \sum_j F_j(1 - r_j)\chi \text{Market Shock}(P). \tag{2.3.3}$$

2.4 | Results

We apply the valuation model to the assumed Climate Loan Portfolio Data to demonstrate the result. This data set contains simulated loan data of two fictitious portfolios: Portfolio A and Portfolio B. The portfolio of loans are from different geographies and sectors. Figure 2.4.1 is a subset of the dataset, showing the loan information of different portfolios.

By referring to Figure 2.2.2, it becomes evident that under the *Below 2°C Delayed* policy, the Market Share of oil remains relatively stable until 2030, after which it exhibits a continuous decline. We can obtain the Market Shock of oil with respect to the *Below 2°C Delayed* policy using equation 2.2.1 and equation 2.2.2. Then, we select the sample loan *LoanID626UnOi*, which is in oil sector, to compute its valuation under *Below 2°C Delayed*, given the face value $F_j = 5,940,895$ and the recovery rate $r_j = 0.4$, set $\chi = 1$, we can calculate the loan's valuation spanning from 2020 to 2050, with intervals of 5 years. Figure 2.4.2 illustrates change in loan values to Market Shock. It becomes apparent that the loan value remains unchanged until the Market Shock emerges in 2030.

Loan Portfolio Data:

LoanID	BorrowerCreditRating	LoanType	Portfolio	InterestRate	InterestType	BorrowerRegion	BorrowerSector	OriginationDate	MaturityDate	FaceValue	FairValue
LoanID626UnOi	A2	Term	Portfolio B	0.0325	Fixed	United States	Oil	31-Aug-20	31-Jul-28	5940895	4922178
LoanID136EuOi	Aa3	Term	Portfolio A	0.027	Fixed	Europe	Oil	30-Apr-20	30-Apr-29	8329151	7211622
LoanID446AfCo	Aa1	Term	Portfolio A	0.017	Fixed	Africa	Coal	30-Nov-19	31-May-31	5404480	6711562
LoanID507UnOi	Aaa	Term	Portfolio A	0.018	Fixed	United States	Oil	31-May-19	30-Apr-29	13907276	10171201
LoanID262InOi	A1	Term	Portfolio A	0.03	Fixed	India	Oil	31-Mar-19	30-Apr-30	11956296	6936505
LoanID769JaOi	Baa2	Term	Portfolio A	0.041	Fixed	Japan	Oil	29-Feb-20	30-Apr-27	17942452	13524261
LoanID622AfCo	A2	Term	Portfolio A	0.031	Fixed	Africa	Coal	30-Nov-18	31-Jul-27	7567771	8758214
LoanID948AfCo	Aaa	Term	Portfolio A	0.017	Fixed	Africa	Coal	30-Apr-19	28-Feb-26	12280372	8187601
LoanID695EuOi	Baa1	Term	Portfolio A	0.037	Fixed	Europe	Oil	31-Dec-19	31-Dec-29	10275597	11135715
LoanID594InOi	Aa1	Term	Portfolio A	0.02	Fixed	India	Oil	31-Mar-18	29-Feb-32	10460813	9670102

Previous Page Next Page

Figure 2.4.1: Climate Loan Portfolio Data example

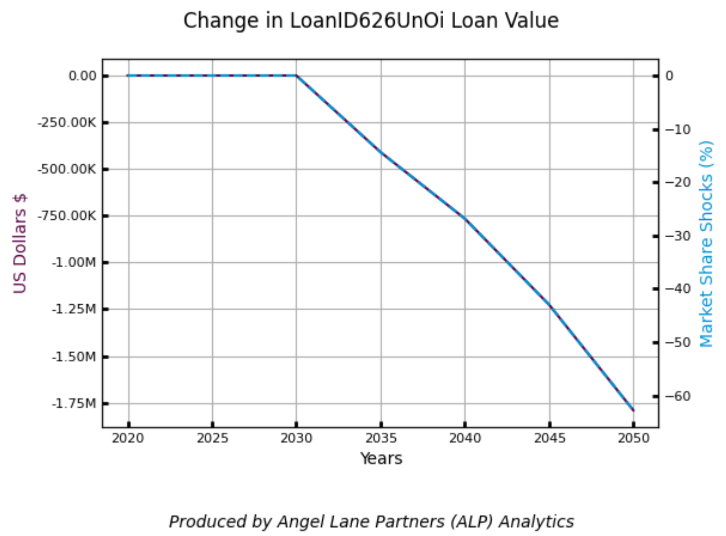


Figure 2.4.2: Compare the change in loan value with the Market Shock

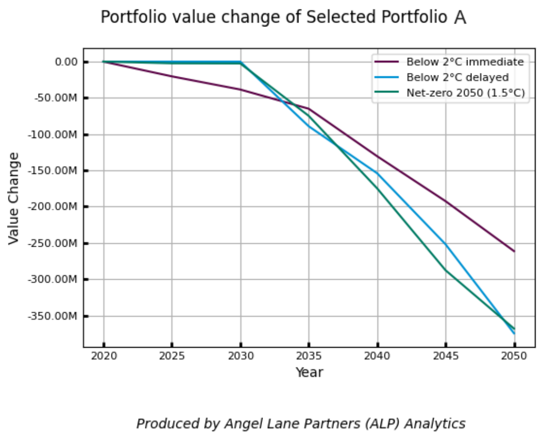
Stress Testing Results

LoanID	Portfolio	BorrowerRegion	BorrowerSector	FaceValue	Years	PD change Below 2°C immediate	value change Below 2°C immediate	PD change Below 2°C delayed	value change Below 2°C delayed	PD change Net-zero 2050 (1.5°C)	value change Net-zero 2050 (1.5°C)
LoanID626UnOi	Portfolio B	United States	Oil	5940895.0	2020.0	-0.0	0.0	-0.0	0.0	-0.0	0.0
LoanID626UnOi	Portfolio B	United States	Oil	5940895.0	2025.0	0.02	-76654.63	-0.0	0.0	-0.03	122402.01
LoanID626UnOi	Portfolio B	United States	Oil	5940895.0	2030.0	0.03	-124453.91	-0.0	0.0	-0.08	299219.12
LoanID626UnOi	Portfolio B	United States	Oil	5940895.0	2035.0	0.08	-274468.04	0.14	-512110.59	0.09	-317301.3
LoanID626UnOi	Portfolio B	United States	Oil	5940895.0	2040.0	0.22	-773286.56	0.27	-956631.05	0.26	-918677.23
LoanID626UnOi	Portfolio B	United States	Oil	5940895.0	2045.0	0.29	-1041645.59	0.43	-1531664.99	0.44	-1583091.29
LoanID626UnOi	Portfolio B	United States	Oil	5940895.0	2050.0	0.42	-1501998.48	0.63	-2237966.41	0.54	-1907960.74

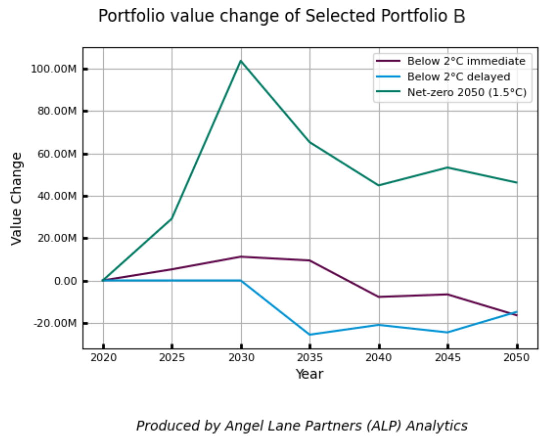
Previous Page Next Page

Figure 2.4.3: The result of change in PD and value of the selected loan portfolio

Additionally, we offer users a selection panel to choose the portfolios, regions, and sectors of their interest, according to the data they have uploaded. To present the results more visually, we will specifically add the option for our uploaded example loan data worksheet from both Portfolio A and Portfolio B, focusing on loans within the Oil, Hydro, and Bioenergy sectors, and spanning across the European and USA regions.



(a) Loan Portfolio A.



(b) Loan Portfolio B.

Figure 2.4.4: The aggregate change in value of the different selected loan portfolios under different climate scenarios.

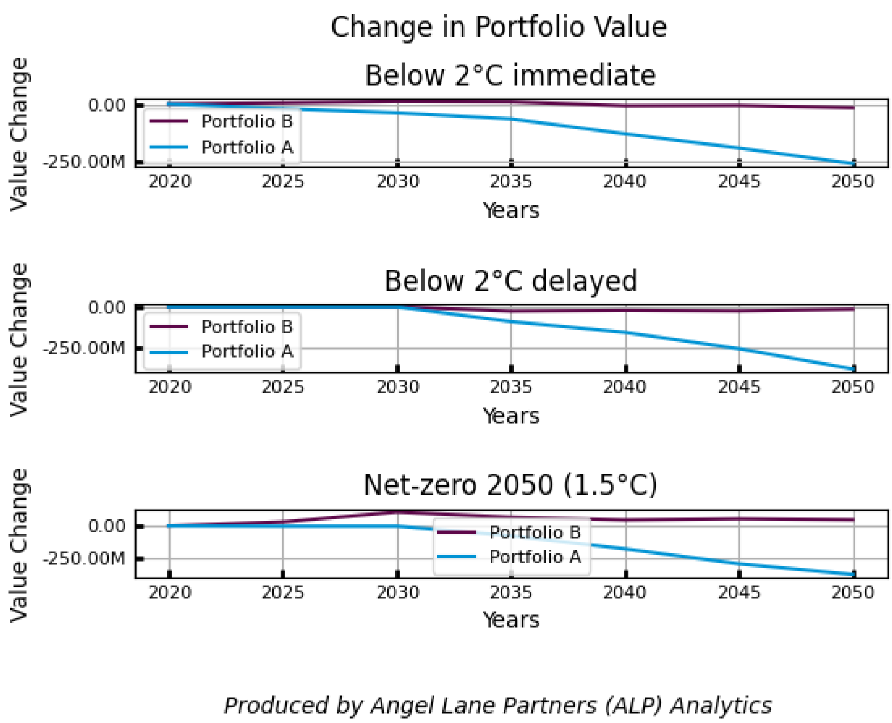
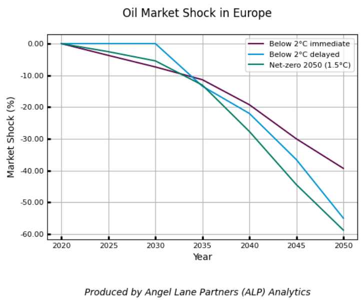
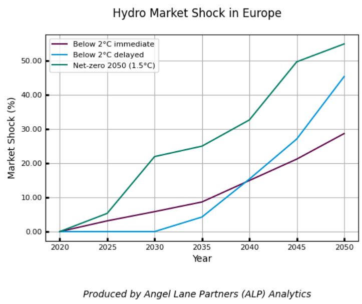


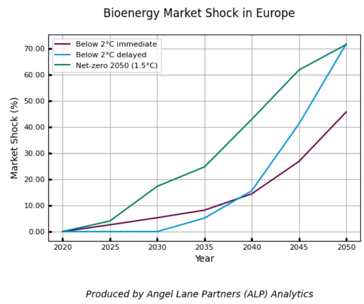
Figure 2.4.5: The aggregate change comparison between Portfolio A and Portfolio B under different climate scenarios.



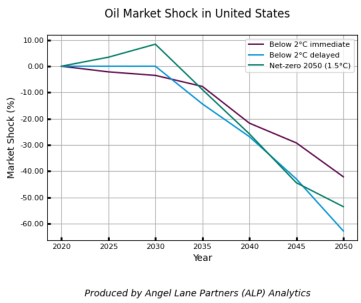
(a) The Market Shock in oil sector in Europe.



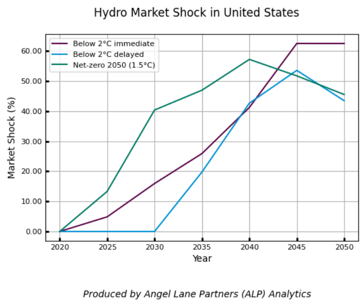
(b) The Market Shock in hydro sector in Europe.



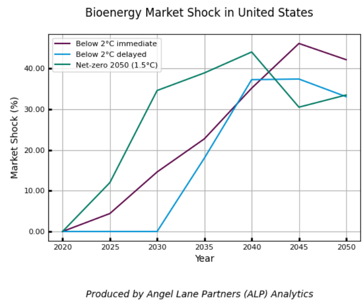
(c) The Market Shock in bio-energy sector in Europe.



(d) The Market Shock in oil sector in USA.



(e) The Market Shock in hydro sector in USA.



(f) The Market Shock in bioenergy sector in USA.

Figure 2.4.6: Market Shocks of different regions and sectors under different scenarios..

We can subsequently present the outcomes of change in the PD and loan value for each loan across various years and scenarios, as depicted in a table similar to Figure 2.4.3. Each page will display the results for an individual loan, allowing users to navigate through pages to assess different loans. Additionally, Figure 2.4.4 will demonstrate the cumulative value changes in the selected loan portfolio across different climate scenarios. Users can further conduct a direct comparison between the portfolios of Portfolio A and Portfolio B using Figure 2.4.5.

Furthermore, we will present Market Shocks corresponding to various climate scenarios, determined by user selections, as illustrated in Figure 2.4.6. This allows for a direct comparison of Market Shocks, offering a clear perspective on the potential changes in PD and values. This proves beneficial even when specific loan data for that sector and region is unavailable.

2.5 | Conclusion

In summary, to derive the changes of the portfolio valuation under different policy scenarios, we should proceed with the following steps:

- Read the loan portfolio dataset and classify the loan into different sectors.
- Obtain the market size of each sector under different policies. The data may be obtained from central banks and other regulatory bodies.
- Calculate the Market Shock of each sector under different policies using equations 2.2.1 and 2.2.2.
- Estimate the elasticity parameter χ. Then equation 2.3.1 gives the change of the PD.
- Now we can obtain the change in the valuation of loan using equation 2.3.2 and 2.3.3.

Chapter 3

Carbon Cost Sensitivity: A Bottom-Up Approach in Transition Risk Scenario Analysis

3.1 | Introduction of the Carbon Cost

The NGFS transition scenarios, including Delayed Transition and Net Zero 2050 as introduced in Chapter 1, presuppose a significant rise in carbon prices, leading to a decrease in CO2 emissions.

- Current Policies (+3.0°C): Baseline scenario, with no or little change in current policies to combat climate change, causing high physical risks but minimal transition risk with global temperature increase by more than 3 degrees;
- Net Zero 2050 (+1.5°C): Coordinated global policy implementation (Paris Agreement) to combat climate change and limit the global temperature increase to 1.5 degrees, implying moderate transition risks;
- Delayed Transition (+2.0°C): Delayed policy implementation creates a "Minsky Moment" with high transition risks.

Carbon pricing, often referred to as the shadow price of carbon, serves as a proxy for carbon policies. It can be narrowly construed as a carbon tax per tonne of CO2 emissions, added to the selling price of a product based on the quantity of greenhouse gases emitted during its production and/or use. Certain carbon taxes may directly apply to a company's or sector's emissions.

From a company's standpoint, the carbon tax will impact the company's EBITDA. In the following sections, we will show the structural transmission from carbon price to the PD of a company.

3.2 | From Carbon Price to Total Asset Value Shock

Denote $CE(i,j,t)$ to be the emissions in tons of CO2 equivalent emitted by the borrower company i in a given region j at date t. Each region j has a representative carbon price (CP or $CP(j,k,t)$) for each date t in each climate scenario k. Normally, CP will change along with the the time t in scenarios of Delayed Transition and Net Zero 2050, however, the baseline scenario $k = 0$, which is the current policy, will keep the carbon prices stable:

$$CP(j,k=0,t) = CP(j,k=0,t=0). \tag{3.2.1}$$

Each year, each borrower company has the carbon cost (CC) derived from the company activities in each region j in the set $\mathcal{G}$ regions where the company has reported emissions:

$$CC(i,k,t) = \sum_{j\in\mathcal{G}} CE(i,j,t) \times CP(j,k,t). \tag{3.2.2}$$

To simplify the carbon cost model, we make the assumption that each borrower company will report emissions in a single region exclusively. Thus, equation 3.2.2 can be rewritten as:

$$CC(i,k,t) = CE(i,j,t) \times CP(j,k,t). \tag{3.2.3}$$

Now the question is, how can we obtain the carbon emission $CE(i,j,t)$?

As for companies who report the emission data to aggregators, such as Bloomberg, we can read the Scope 1 and Scope 2 carbon emission directly. However, for some companies who do not report carbon emission data, we refer to a proxy measure based on industry level emission multipliers. CO2 emission multipliers are introduced in the IMF Climate Report Dashboard[1]. These multipliers signify the volume of CO2 released into the atmosphere due to both Scope 1 and Scope 2 fuel combustion per million USD of output. For instance, the emissions associated with one million USD of electricity encompass the Scope 1 emissions from the electricity producer as well as the Scope 2 emissions embedded in inputs used during electricity production (such as transportation services for transporting fossil fuel to the electricity plant).

If we denote the carbon multiplier as $M(j,s,t)$ in a given region j and sector s at time t, and the revenue of each borrowing company as $Rev(i,j,s,k,t)$, we can calculate the carbon emissions using the formula:

$$CE(i,j,t) = M(j,s,t) \times Rev(i,j,s,k,t)/(\$100,0000). \tag{3.2.4}$$

However, it is unfeasible for us to ascertain the revenue of each borrowing company under various climate scenarios. To address this challenge, the NGFS introduced the G-cubed model, designed to comprehend the sectoral impact on a company's output projections concerning climate change, utilising NGFS scenarios. The revenues of a company i in the respective sector j can be proportionately reduced based on the decline in the product of output and price for the corresponding sector under different climate scenarios k. Consequently, the future revenue $Rev(i,j,s,k,t)$ can be derived from the baseline revenue $Rev(i,j,s,k=0,t)$ by

$$Rev(i,j,s,k,t) = Rev(i,j,s,k=0,t) \times (1+O(s,k,t)) \times (1+P(s,k,t)), \tag{3.2.5}$$

where the $O(s,k,t)$ is the output trajectories under the transition scenarios k as a percentage deviation from the Current Policies scenario $k=0$, the $P(s,k,t)$ is the producer price trajectories under the transition scenarios k as a percentage deviation from the Current Policies scenario $k=0$.

Next, we can calculate the carbon cost by substituting equations 3.2.5 and 3.2.4 into equation 3.2.3. Given that the carbon cost is incorporated in the cost of goods sold, integrating this carbon cost results in a shock to EBITDA, defined as

$$\xi(i,k,t) = \frac{CC(i,k,t}{\text{EBITDA}(i,k=0,t=0)}. \tag{3.2.6}$$

[1] https://climatedata.imf.org/pages/re-indicators/#re3

The impact of the variation of EBITDA on total asset value V_A can be computed through $\mathcal{R}_i$ (V_A-EBITDA ratio). The assumption we take is that, this $\mathcal{R}_i$ ratio remains constant over time and across different climate scenarios. In this case, the shock is directly transmitted to the V_A. The economic shock transmission to the total asset value of each company i, in each scenario k and at each date t is

$$\begin{aligned} V_A(i,k,t) &= (1-\xi(i,k,t)) \times V_A(i,k=0,t=0) \\ &= (1-\xi(i,k,t)) \times R_i \times \text{EBITDA}(i,k=0,t=0). \end{aligned} \tag{3.2.7}$$

The V_A-EBITDA ratio will be different for individual firms and will depend on factors such capital structure, accounting policies etc. We propose that the $\mathcal{R}_i$ is calculated with reference to past three years of company specific data, taking note of any specific changes that may have a significant impact on this ratio and making adjustment accordingly.

In the next section, we look at Merton's model and illustrate its application to the quantification of corporate PD in the context of transition risk analysis.

3.3 | From Total Asset Value Shock to Probability of Default

Merton (1974) developed the model of calculating the probability of default P_d from the total asset value V_A. This Merton's model establishes connections among market values of equity, assets, and liabilities within an option pricing framework. It assumes a single liability L with maturity T, typically a one-year period. At time T, the firm's value to the shareholders equals the difference $V_A - L$ when the asset value V_A is greater than the liabilities L. However, if the liabilities L exceed the asset value V_A, then the shareholders receive no compensation. The value of the equity E at time T is related to the value of the assets and liabilities by the following formula:

$$E = \max(V_A - L, 0). \tag{3.3.1}$$

Assuming a log-normal distribution for the asset returns , we can use the Black-Scholes-Merton equations to relate the observable market value of equity E, and the market value of assets V_A, at any time prior to the maturity T:

$$E = V_A \mathcal{N}(d_1) - Le^{-rT}\mathcal{N}(d_2), \tag{3.3.2}$$

where r is the risk-free interest rate, $\mathcal{N}$ is the cumulative standard normal distribution, and d_1 and d_2 are given by:

$$d_1 = \frac{\ln(\frac{V_A}{L}) + (r + 0.5\sigma_{V_A}^2)T}{\sigma_{V_A}^2\sqrt{T}}, \tag{3.3.3}$$

$$d_2 = d1 - \sigma_{V_A}\sqrt{T}, \tag{3.3.4}$$

and σ_{V_A} is the volatility of assets. Then, the theoretical probability of default (P_d) can be derived as:

$$P_d = 1 - \mathcal{N}(d_2) = \mathcal{N}(-d_2). \tag{3.3.5}$$

However, the asset volatilities are not directly observable and therefore need to be calculated. Depending on the availability of specific data, the determination of asset volatility requires different approaches. We propose the following to achieve this:

- **For Private Firms**: one way of tackling this is to analyse the total asset values over the past years, taking into account the fact that not all assets are carried at market value and will need to be restated.

  Alternatively, the probability of default measure can be extracted from the company's credit rating assigned by the bank. The asset volatility can the be reverse calculated using the company's debt and total assets using the Merton's model.

- **For Private Rated Firms**: We calculate company's baseline PD from S&P or Moody's rating. Merton's model is then used to calculate asset volatility. In other words, we essentially derive the asset volatility by equating the PD under the Merton model, assuming normal distribution, with the PD derived from external rating.

- **For Listed Firms**: When observable equity volatility σ_E is readily available, it serves as a direct input for solving the asset volatility σ_{V_A} as

$$\sigma_E E = \mathcal{N}(d_1)\sigma_{V_A} V_A. \tag{3.3.6}$$

  That is, the typical change in equity value is equal to the typical change in asset value, adjusted for the probability of the assets surviving as $\mathcal{N}(d_1)$ indicates. Utilising equation 3.3.3, 3.3.4, and the formula of E as 3.3.2, the σ_{V_A} is solved by

$$\sigma_{V_A} = \frac{\left(V_A \mathcal{N}(d_1) - Le^{-rT}\mathcal{N}(d_2)\right)\sigma_E}{\mathcal{N}(d_1)V_A} \tag{3.3.7}$$

Once we have all the required parameters, the final step is to calculate the stressed PD based on the Merton's model, using the updated total assets, debt, risk free rate, and the asset volatility.

In summary, as we move forward with our transition risk analysis, we maintain the assumption that the V_A-EBITDA ratio ($\mathcal{R}i$) and asset volatility (σ_{V_A}) remain constant across different climate scenarios. With this foundation, we leverage available carbon emission data to compute the carbon cost ($CC(i, k, t)$) using Equation 3.2.3 under each climate scenario.

This computed carbon cost serves as a pivotal input, instigating the Total Asset Value Shock ($V_A(i, k, t)$) through the application of Equations 3.2.6 and 3.2.7. Subsequently, we employ Equations 3.3.3, 3.3.4, and 3.3.5 to determine the PD (P_d) for each company. These sequential steps establish a robust methodology, integrating Merton's model and previous calculations, to quantify the potential default risk under different carbon cost in different climate scenarios.

3.4 | Further Discussion: Parameter Uncertainties

In the preceding sections, we introduced our model, utilizing the carbon cost and Merton's model to assess the PD of companies under various climate scenarios. However, there are still a number of uncertainties

	Revenue (USD)	EBITDA (USD)	Carbon Emission (Real, tonnes)	Carbon Cost (Real, USD)	Carbon Emission (Multiplier, tonnes)	Carbon Cost (Multiplier, USD)
Toyota	276.02B	0.53B	557M	116B	131M	27.5B
Nissan	79.5B	0.08B	121M	25.4B	37.9M	7.96B

Table 3.3.1: The financial status, carbon emission, and carbon cost of Toyota and Nissan companies.

that require further exploration. One of the biggest challenges is the availability of accurate carbon emission data.

For some of the companies, obtaining carbon emission data may be possible through sources like Bloomberg or official company reports. However, in instances where direct carbon emission data is unavailable, a proxy measure based on industry-level emission multipliers, as discussed in the previous section, may be utilized.

Nevertheless, delving deeper into the reliability of carbon emission multipliers reveals their variability, as illustrated by the comparison of Toyota and Nissan. Both Japanese car manufacturing companies share a same carbon emission multiplier 477.2 [2]. Table 3.3.1 presents their financial status sourced from Bloomberg, alongside the carbon emission and cost, utilizing an average carbon price of 210 USD per ton of CO2. The apparent gap between the cost of carbon and EBITDA highlights a potential default risk, particularly if the carbon price comes under immediate pressure.

This underscores the inherent instability and potential limitations of relying solely on industry-level emission multipliers. To address these uncertainties, seeking peer references and employing clustering methods becomes imperative. By aligning companies with similar profiles and characteristics, we can enhance the reliability and accuracy of our carbon emission assessments in situations where direct data is either not readily available or otherwise needs to be verified. A detailed exploration of this method is given in 4.

3.5 | Results

To demonstrate our carbon cost sensitivity model, we searched the financial data values and the carbon emission data of EMAAR Properties from Bloomberg and the Pitchbook. Then we calibrated the V_A-EBITDA ratio and the asset volatility to match the S&P rating of EMAAR as BBB-. The detailed data is shown in Figure 3.5.1.

In Figure 3.5.2, we observe the dynamic change in EMAAR's PD across distinct scenarios throughout the designated time span. Notably, the Delayed scenario, characterized by postponed implementation of climate policies, would necessitate a swift reduction in emissions to compensate for the temporal time loss. As a consequence, starting from 2035, the PD in the Delayed scenario exhibits a more accelerated increase when contrasted with the NetZero scenario. This underscores the critical impact of the timing of climate policy implementation on EMAAR's PD dynamics, highlighting the intricate interplay between climate scenarios and financial risk.

Additionally, we employed the derived PD to assess the impact on Capital and the Expected Loss (EL) for EMAAR, with $EL = PD \times LGD \times EAD$, where $LGD = 0.45$ and $EAD = 100$. Figure 3.5.3 provides a comprehensive depiction of the expected loss and the corresponding change in capital across different climate scenarios. This visual representation offers banks and analysts valuable insights into the carbon cost sensitivity, facilitating a deeper understanding of the financial implications associated with different climate scenarios for EMAAR.

[2] multipliers signify the tonnes of CO2 released into the atmosphere per million USD of revenue.

Loan Portfolio Data:

name	sector	current_pd	revenue	ebitda	debt	asset_vol	ebitda_date	ev_ebitda_ratio	risk_free_rate	carbon_emission
EMAAR Properties LLC	Construction	0.016	6429075000	2218493000	3423223000	0.18	2023-06-30	5.5	0.03	500263

Previous Page Next Page

Figure 3.5.1: Financial Data of Emaar Properties

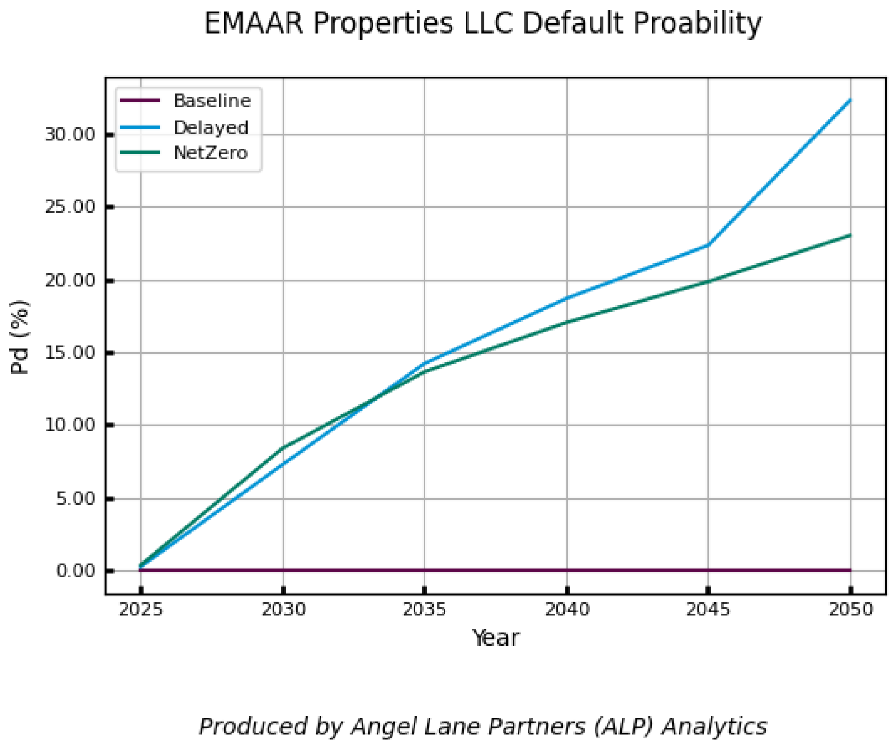
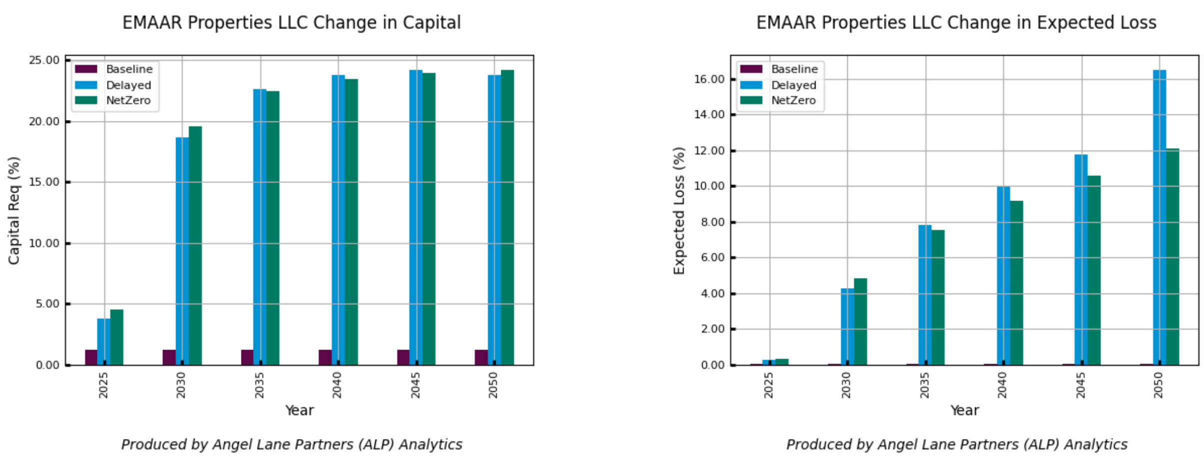


Figure 3.5.2: The change of PD of the EMAAR Properties under different climate scenarios.



(a) The Capital Change

(b) The Expected Loss

Figure 3.5.3: Capital Change and the Expected Loss of the EMAAR Properties under different climate scenarios.

3.6 | Conclusion

In summary, to obtain the change in PD of a selected obligor under various climate scenarios, we should proceed with the following steps:

- Retrieve the current Revenue ($Rev(i, j, s, k = 0, t)$), EBITDA (EBITDA$(i, k = 0, t = 0)$), Debt (L), asset-volatility (σ_{V_A}), Time to Maturity(T), EV-EBIDTA ratio ($\mathcal{R}_i$) from the borrowers i, and also classify the borrower companies into different sectors s and regions j.
- Obtain the output and product trajectories $O(s, k, t)$ and $P(s, k, t)$ from the NGFS G-cubed model. Calculated the shocked revenue ($Rev(i, j, s, k, t)$) of the borrower company by using equation 3.2.5.
- Obtain the carbon emission multiplier ($M(j, s, t)$) of each region j and sector s at time t. Calculate the carbon emission ($CE(i, j, t)$) by using equation 3.2.4.
- Calculate the carbon cost ($CC(i, k, t)$) of each borrower company using equation 3.2.3.
- Then we can calculate the total asset value shock ($V_A(i, k, t)$) by using equation 3.2.6 and 3.2.7.
- Now we can obtain the PD (P_d) of the borrower company from the total asset value shock $V_A(i, k, t)$ by using equations 3.3.3, 3.3.4 and 3.3.5.

Chapter 4

Applying Cluster Analysis for Data Calibration

4.1 | Introduction

Clustering methods play a pivotal role in data analysis and calibration by categorising entities into groups based on shared features or patterns, providing valuable insights into underlying relationships within the data. In the context of our analysis, clustering becomes instrumental in handling uncertainties related to carbon emission data.

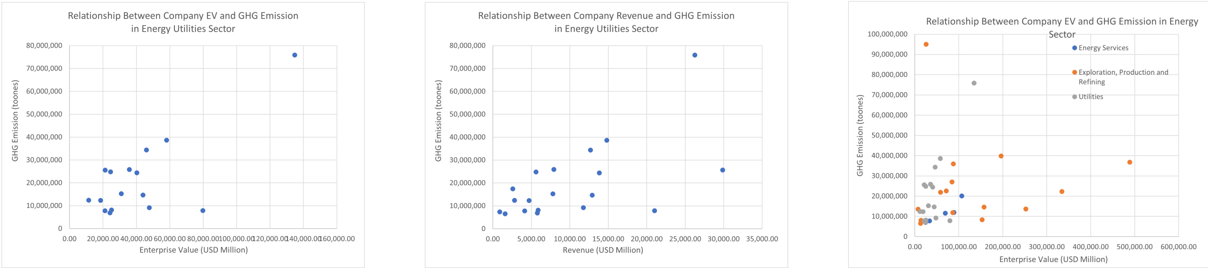
When confronted with unavailable or unreliable direct carbon emission data, clustering aids in identifying comparable entities, thereby enhancing the precision of our carbon emission assessments. We take the assumption that companies within the same sectors, sharing similarities in EBITDA, Revenue, Sector, Number of Staff, and other features, are likely to exhibit comparable carbon emission patterns. This assumption is derived from the logic that companies within a sector that use similar technologies and energy sources to generate revenue will have similar emission profiles. Although an approximation, this approach is more reliable than traditional emissions multipliers.

To execute these clustering methodologies, we employ k-means clustering and a cosine distance metric based on relevant features. Through this approach, we conduct cluster analysis and proxy selection, further refining our calibration process for carbon emission data.

4.2 | Data Preparation, Feature Selection, and Scaling

In the first steps of our data preparation process, we start by collecting extensive information about the companies. This dataset includes vital financial indicators such as external rating, Default Probability (PD), EBITDA, Total Revenue, Employee Count, Enterprise Values, Debt, and, if available, carbon emission or greenhouse gas (GHG) emission. Following this, we move on to the creation of a DataFrame, achieved either by generating synthetic data or by integrating existing data from various sources.

In the process of preparing our data, we adopt meticulous feature selection techniques. Feature selection involves identifying and choosing variables in a dataset that exhibit a significant correlation with or impact on the target variable. This becomes especially crucial in larger datasets, where there could be numerous



(a) GHG Emission vs. EV for Energy Utilities Sector

(b) GHG Emission vs. Revenue for Energy Utilities Sector

(c) GHG Emission vs. EV for Overall Energy Sector

Figure 4.2.1: EAD example on GHG Emission and Financial Data for Energy Sector.

features, some of which may lack relevance or connection to the desired output. Thus, the process of feature selection is essential before modelling to ensure the highest level of accuracy.

We employ Exploratory Data Analysis (EDA) as a foundational step in our feature selection process. This encompasses a comprehensive exploration of our dataset's features and variables. Through EDA, we derive key statistical measures such as count, mean, standard deviation, minimum, and maximum for quantitative variables. Additionally, we visualize dynamic trends between various variables and identify correlation factors among different features. Figure 4.2.1, a robust correlation emerges between the greenhouse gas (GHG) emission data and both the Enterprise Value and Revenue of the companies. Figure 4.2.1c further illustrates that distinct sub-sectors exhibit diverse patterns in the association between GHG emission data and Enterprise Value.

Thus, for this dataset we will strategically choose Enterprise Value and Revenue as our central features for the subsequent clustering analysis of GHG emission data. These observations underscore the intricate relationships within the dataset, highlighting the need for nuanced analyses that consider sector-specific dynamics, and ensuring a well-informed and targeted approach to feature selection.

Before we apply the clustering model to the data, we will also leverage the data scaling process. As for financial dataset, the data range and distribution among all features are relatively different from one another, not to mention some variables bearing with outliers. It is essential that we apply feature scaling to the entire dataset consistently for the purpose of making it more digestible to clustering algorithms.

4.3 | K-Means Clustering Using Cosine Distances

K-Means clustering stands out as a widely adopted unsupervised machine learning algorithm, effectively partitioning data points into distinct clusters. It has the ability to unveil underlying patterns and groupings within datasets, offering valuable insights into the inherent structure of the data. In mathematical terms, this algorithm aims to partition n observations into k clusters, assigning each observation to the cluster with the nearest mean. The algorithm achieves this by minimizing the within-cluster variance. The objective function is given by:

$$J = \sum_{i=1}^{k}\sum_{j=1}^{n_i} dis(x_{ij} - \mu_i) \tag{4.3.1}$$

where x_{ij} is the j-th observation in cluster i, μ_i is the mean of cluster i, and n_i is the number of observations in cluster i, and function $dis()$ is the distance function that we used to calculate the distance between the observations and the centre mean of each cluster. Here we adopt the cosine distance.

Cosine distance calculates the cosine of the angle between two vectors, providing a measure of similarity regardless of their magnitude. In the context of K-Means clustering, cosine distance is employed to determine the similarity between feature vectors of different entities. If the cosine distance is close to 1, the vectors are similar, indicating a smaller angle between them. Conversely, a cosine distance closer to -1 signifies dissimilarity.

Mathematically, for two vectors A and B, the cosine similarity is given by:

$$\text{cosine\_similarity}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

where $\cdot$ represents the dot product and $\|\|$ denotes the Euclidean norm.

To apply K-Means clustering with cosine distances, we initialize cluster centroids and iteratively assign data points to the nearest centroid based on cosine distances. This process continues until convergence, resulting in distinct clusters of entities with similar feature vectors.

4.4 | Calibrate the data using K-Nearest Neighbours

K-Nearest Neighbours (K-NN) is a versatile algorithm used for both classification and regression tasks. In the context of data calibration, K-NN can be employed to estimate missing or uncertain data points, such as GHG Emission or Asset Volatility, based on the values of their neighbouring data points.

The algorithm firstly calculates the similarity between data points, often using distance metrics like Euclidean or cosine distances. Here, since we use the cosine distance to do the clustering, we will keep the cosine distance to do the K-NN calibration.

For each data point with a missing GHG emission value as an example, the K-NN algorithm identifies the K-nearest neighbours based on the available features. Once the nearest neighbours are identified, the missing GHG emission value is imputed by considering the values of the feature from its K-nearest neighbours. A common approach is to take the mean or weighted mean of the neighbours' GHG emission values.

By leveraging the information from neighbouring data points, the K-NN method provides a data-driven approach to imputing missing values, contributing to a more complete and reliable dataset for subsequent analyses. The choice of K (number of neighbours) and the distance metric are crucial parameters that can be fine-tuned based on the characteristics of the dataset and the desired imputation accuracy.

In summary, K-Means clustering with cosine distances helps categorize entities with similar features, while K-Nearest Neighbours aids in data calibration by estimating missing values based on the characteristics of neighbouring data points. Together, these techniques contribute to a more robust and refined dataset for subsequent analyses.

4.5 | Application: GHG Emission Data Calibration

To enhance the demonstration of our model, we generated a thousands of synthetic company data, encompassing key financial metrics such as EBITDA, Total Revenue, Employee Count, Enterprise Value, and GHG Emission (when available). To enhance the sector difference and feature correlations, we defined a sensible relationship between Total Revenue and GHG Emission based on the sector. Then, in order

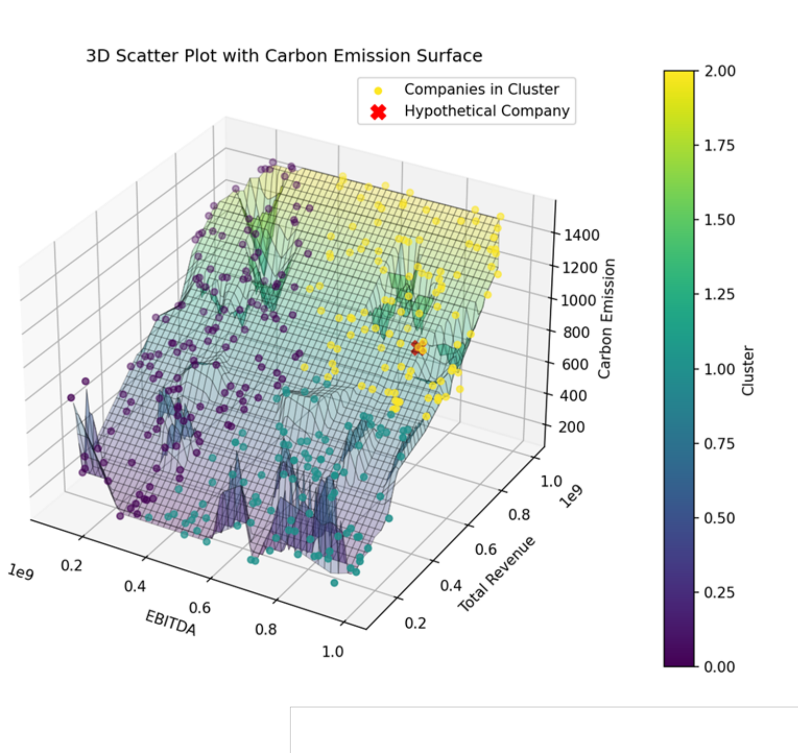


Figure 4.4.1: K-Means Clustering and K-NN for hypothetical carbon emission data calibration.

to introduce an element of uncertainty, we introduced a hypothetical company with the missing GHG Emission data. The target is to calibrate the missing GHG Emission data for this hypothetical company, using K-Means clustering with cosine distances and the K-Nearest Neighbours approach.
The outcomes of our analysis are presented in Figure 4.4.1 and Figure 4.5.1, offering a comprehensive insight into the clustering patterns based on EBITDA and Total Revenue features.

Additionally, the relationship between the hypothetical company and its neighbouring entities is visually represented, providing a detailed interpretation in Figure 4.5.1. Notably, by exclusively correlating Total Revenue with Carbon Emission, the calibrated Carbon Emission data aligns closely with Company_575, which exhibits a similar Total Revenue to our hypothetical company. This observation elucidates the nuanced relationships within the clustering results, underscoring the impact of specific financial features on the calibrated data.

4.6 | Conclusion

In this chapter, we presented the application of the K-Means Clustering and K-Nearest Neighbourhood approach as a robust method for calibrating missing values. The alignment of companies with similar profiles and characteristics enhances the reliability of the data calibration, as demonstrated in our application examples. Additionally, our approach not only provides a valuable framework for handling data uncertainties but also enables informed estimations, contributing to more sustainable decision-making processes grounded in reliable insights and well-founded approximations.

```
Nearest Company Information:
Company              Hypothetical_Company
Sector                         Technology
EBITDA                 799481280.527223
Total_Revenue          800766011.847275
Employee_Count                     1000
Carbon_Emission              802.042487
Cluster                               2
Name: 1000, dtype: object
Interpolated Carbon Emission: 802.0424874359172

EBITDA, Revenues, Employee Count, and Carbon Emission of 5 Closest Neighbors:
         Company        EBITDA  Total_Revenue  Employee_Count  Carbon_Emission
346  Company_346  9.240709e+08   1.163009e+08             202       179.461586
960  Company_960  6.973690e+08   1.004920e+08             117       155.890414
939  Company_939  6.846718e+08   1.000000e+08            3820       155.156896
351  Company_351  9.743462e+08   1.438573e+08            2660       220.548137
575  Company_575  1.252364e+08   8.492933e+08            3951      1272.354244
```

Figure 4.5.1: Calibration result of the hypothetical company, and relevant data of the 5-nearest neighbour companies.

4.7 | Future Work: Ricci Curvature for Clustering

Ricci curvature is utilized as a geometric measure to guide the transformation of graph structures, with the aim of improving clustering outcomes in subsequent machine learning tasks. The approach involves capturing and leveraging geometric information encoded in the graph's curvature to influence the definition of distances and, consequently, the clustering process.

Key Steps Envisioned

1. **Graph Representation:** The graph is represented as $G = (V, E)$, where V is the set of vertices (data points) and E is the set of edges.

2. **Ricci Curvature:** Ollivier-Ricci curvature for an edge (i, j) in the graph is denoted as $\kappa(i, j)$.

3. **Ricci Flow:** The evolution of edge weights in Ricci flow can be represented as a differential equation, for example:
$$\frac{dw_{ij}}{dt} = -\kappa(i, j)w_{ij}$$
where w_{ij} is the weight of the edge (i, j) and t is time.

4. **Distance Measure:** The distance measure induced by Ricci flow might be expressed as a function of the evolved edge weights, for example:
$$d_{ij} = f(w_{ij})$$

5. **Clustering Process:** Given an embedding X of the graph into Euclidean space, and a clustering algorithm like K-means:
$$\text{argmin}_{\text{clusters}} \sum_{i=1}^{n} \|x_i - \mu_{c_i}\|^2$$
where x_i is a data point, μ_{c_i} is the centroid of the cluster c_i, and n is the number of data points.

6. **Comparison Before and After Ricci Flow:** The comparison involves evaluating the clustering performance using an appropriate metric, for example:
$$\text{Performance Metric} = \text{Compare}(\text{Ground Truth}, \text{Clustering Results})$$

7. **Evaluation Metrics:** The evaluation metrics may include measures like Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), or others, depending on the specific goals.

Utilizing Ricci curvature and Riemannian geometry in clustering offers a unique approach that goes beyond traditional Euclidean distance-based algorithms. By incorporating geometric properties of graphs and capturing curvature information, this methodology has the potential to provide several advantages:

- **Incorporation of Intrinsic Geometry:** Ricci curvature considers the intrinsic geometry of graphs, providing a more nuanced measure than Euclidean distances. This can be especially beneficial when dealing with complex data structures and non-linear relationships.
- **Preservation of Local Structures:** Ricci flow transformations aim to preserve local structures in the graph. This can be advantageous when dealing with data that exhibits intricate neighborhood relationships, which might be overlooked by traditional methods relying solely on global Euclidean distances.
- **Sensitive to Graph Topology:** Riemannian geometry inherently takes into account the topology of the underlying space. This sensitivity to graph topology allows for a more accurate representation of data relationships, potentially leading to improved clustering performance.
- **Enhanced Discriminative Power:** The ability of Ricci curvature to capture subtle geometric features can enhance the discriminative power of clustering algorithms. This is particularly valuable when dealing with datasets where class boundaries are complex and non-linear.
- **Potential for Improved Cluster Separation:** The utilization of Ricci curvature may contribute to better separation of clusters by considering the intrinsic structure of the data. This can lead to clusters that are more representative of the underlying patterns in the data.

In summary, incorporating Ricci curvature and Riemannian geometry in clustering algorithms presents a promising avenue for improving performance, especially in scenarios where the Euclidean distance-based approaches may fall short in capturing the inherent complexity of the data.

Appendix A

Bibliography

Battiston, S., Mandel, A., Monasterolo, I., Sch¨utze, F., and Visentin, G. (2017). A climate stress-test of the financial system. Nature Climate Change, 7(4):283–288.

BCBS (2023). Disclosure of climate-related financial risks.

Bouchet, V. and Le Guenedal, T. (2020). Credit risk sensitivity to carbon price. Available at SSRN 3574486.

Carney, M. (2015). Breaking the tragedy of the horizon–climate change and financial stability. Speech given at Lloyd's of London, 29:220–230.

CBUAE (2023). 2023 climate risk scenario analysis.

Chen, H., Ens, E., Gervais, O., Hosseini, H., Johnston, C., Kabaca, S., Molico, M., Paltsev, S. V., Proulx, A., and Toktamyssov, A. (2022). Transition scenarios for analyzing climate-related financial risk. Technical report, Bank of Canada Staff Discussion Paper.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics), 28(1):100–108.

Hull, J., Nelken, I., and White, A. (2004). Merton's model, credit risk, and volatility skews. Journal of Credit Risk Volume, 1(1):05.

Jost, J., & Liu, S. (2014). Ollivier's Ricci curvature, local clustering and curvature-dimension inequalities on graphs. Discrete & Computational Geometry, 51(2), 300-322.

Monasterolo, I., Zheng, J. I., and Battiston, S. (2018). Climate-finance and climate transition risk: An assessment of china's overseas energy investments portfolio. Available at SSRN 3163335.

NGFS (2022). Running the NGFS scenarios in G-cubed: A tale of two modelling frameworks.

Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2):1883.

Team, P. G. C. (2022). The global GHG accounting and reporting standard, part a.23