Aalto University

MS-E2177
Seminar on Case Studies in Operations Research

COVID-19 impact on credit loss modelling

Final Report

Aki Malinen

Ricardo Möll

Aleksi Pelttari

Thong Tran (Project Manager)

June 7, 2021

Contents

1 Introduction

1.1 Background

Our client is SEB, a leading Nordic financial services group founded in 1856. SEB is serving and advising 4.4 million customers in Nordics, Baltics, and Germany with a strong focus on institutional and corporate clients.

Our case study is about the lending business, one of banks' major areas of business. In general, most loans will be paid back in time. However, there are also cases where customers default, meaning they are unable to pay back parts or the full amount of the loan due to financial distress. After customers default, there are two possible outcomes. They may resolve their financial problems and continue with the payments, or they cannot resolve the issues and go bankrupt. In the case of bankruptcy, banks may be able to recover the remaining debt by selling collateral or from other possible sources, and the credit that is not retrieved are written down as losses [10]. Credit losses can have serious impacts on the bank's business.

To protect their business from potential losses, banks estimate the probability for such events and set aside capital to absorb losses (provisions). The International Accounting Standards Board has issued International Financial Reporting Standard 9 – Financial Instruments (IFRS9) which introduced the "expected credit loss" (ECL) framework on how banks should recognize and provide for credit losses for financial statement reporting purposes. IFRS9 standard defines principles but gives freedom in choosing what models and approaches banks use to estimate their losses. These estimates are then used to account for financial buffers to protect the losses [15].

In general, expected credit losses are the weighted average of credit losses with the probability of default (PD) as the weight. For this project, we are focusing on the PD and will not examine procedures in bankruptcy.

1.2 Motivation

Modeling credit losses is extremely important for the financial stability of the banking sector. By having an estimate for the upcoming losses, banks are able to set aside enough capital to absorb the losses. The importance of such buffers was shown during the financial crisis. After the crisis, different regulations and standards have been implemented to control the capital and provision requirements. In short, the requirements aim to ensure that the more significant the risk a bank is exposed to, the more capital the bank needs to hold to safeguard its solvency and overall economic stability.

Probability of default (PD) is one of the major components of credit loss estimation. Currently, one method used in SEB to estimate PDs is the Merton-Vasicek (MV) one-factor model [11, 19]

$$\mathrm{PIT} = \Phi\left[\frac{\Phi^{-1}(\mathrm{TTC}) + \sqrt{\rho}Z}{\sqrt{1-\rho}}\right]. \tag{1}$$

The model estimates point-in-time (PIT) PD under certain economic circumstances by considering three main elements:

- Through-the-cycle (TTC) idiosyncratic risk in neutral economic conditions

- Systematic risk Z
- Sensitivity of the customer to the economical environment ρ

For certain sub-portfolios, SEB uses a single economic driver and a single sensitivity variable. This means the model assumes that there is a way to measure the large-scale systematic conditions applicable to the whole portfolio [8].

Although the above-mentioned method works "reasonably" in normal economic conditions, the COVID-19 pandemic has created additional complications. Some sectors, such as leisure and transportation, are having major difficulties due to government restrictions. On the other hand, sectors such as technology, online retail, and utilities have experienced rising demand, which has resulted in higher rating affirmations [6]. It has also been noted that the size and rating are also factors as smaller firms with lower credit ratings are more affected by economic downturn [1, 7]. Additionally, the government support measures are significant. These highly segregated economic conditions pose challenges for the one-factor MV model, in which there is no single systemic factor that explains the general economic conditions for all sectors.

2 Objectives

Our task was to improve the existing Merton-Vasicek model to account for the differences in sensitivities across sectors, rating, and size, as well as the impact from COVID-19 related government supports.

Currently, the model is used to calculate the probability of default under certain economic circumstances in a one-year horizon, and for certain portfolios, SEB is using the single economic driver and a single sensitivity variable. Our focus is to find a way to implement changes in the model to improve results during periods of changing correlations. Changes will be explored on a level of economic environment, sectors, and credit quality of individual clients. We also seek to include indirectly the effects of the COVID-19 pandemic and government support on PD prediction.

Our main goals are the following:

- Calibrate of the MV model on a risk rating level
- Calibrate of the MV model on industry/sector level
- Calibrate of the multi-factor MV model
- Implement a framework for using multiple indicators to predict PD under different economic scenarios both in normal condition and in adverse conditions such as the current COVID-19 pandemic

3 Literature Review

Modeling credit losses largely revolves around modeling the probability of default, which is the probability that a loan defaults over some specified time horizon. This probability usually comes in two types: (i) point-in-time (PIT) which describes the probability in the standing state of the economy at that particular time, and (ii) through-the-cycle (TTC) which is the average probability of default over a business cycle.

The general idea of PD modeling relies on using empirical data of observed credit ratings and default rates together with variables on relevant economic conditions to fit a model. The data is usually represented as a transition matrix that describes the evolution of credit ratings at the beginning and at the end of a selected time period. The industry standard for measuring creditworthiness is based on a discrete scale ranging from A to D where D represents defaulted status. In addition, the tables usually involve a column for resolved defaults and an exit column for loans that are fully paid.

Usually, the available data is limited because the number of companies in the portfolio is usually limited. This means that the empirical data available is often sparse and the grade-level observed default frequencies, referred to as ODF, are often non-monotonic.

Figure 2 in Section 4 showcases the irregularity of real-life portfolio data. However, the general assumption is that for a suitably developed rating model based on carefully selected risk factors, the associated PD curve describing the dependence of default rates with the credit rating should be monotonic for economic reasons; that is, that a lesser credit rating value correlates with a higher probability of the default state. Due to this, modeling PDs involves smoothing the ODF to create a positive and monotonic PD curve. For this purpose quasi moment matching (QMM) has been proposed in [18]. QMM aims to smooth the ODF by utilizing a robust logistic curve

$$\mathrm{PD}(x) = \Pr(D|X = x) \approx \frac{1}{1 + \exp(\alpha + \beta \Phi^{-1}(F_N(x)))} \tag{2}$$

$$F_N(x) = \frac{\Pr(X > x|N) + \Pr(X \geq x|N)}{2} \approx \Pr(X \geq x|N), \tag{3}$$

where x is a given rating, Φ^{-1} is the inverse of the normal cumulative probability distribution, and α and β are parameters to be optimized. The smoothed curve in Figure 1 should match the ODF in marginal PD and accuracy ratio

$$\mathrm{AR} = \Pr(X_D > X_N) - \Pr(X_D < X_N) = 2\Phi\left(\frac{\mu_D - \mu_N}{\sqrt{2}\sigma}\right) - 1.$$

X_D, μ_D and X_N, μ_N are random variables denoting continuous rating and mean ratings given default and not default respectively,

$$X_D \sim N(\mu_D, \sigma) \tag{4}$$

$$X_N \sim N(\mu_N, \sigma). \tag{5}$$

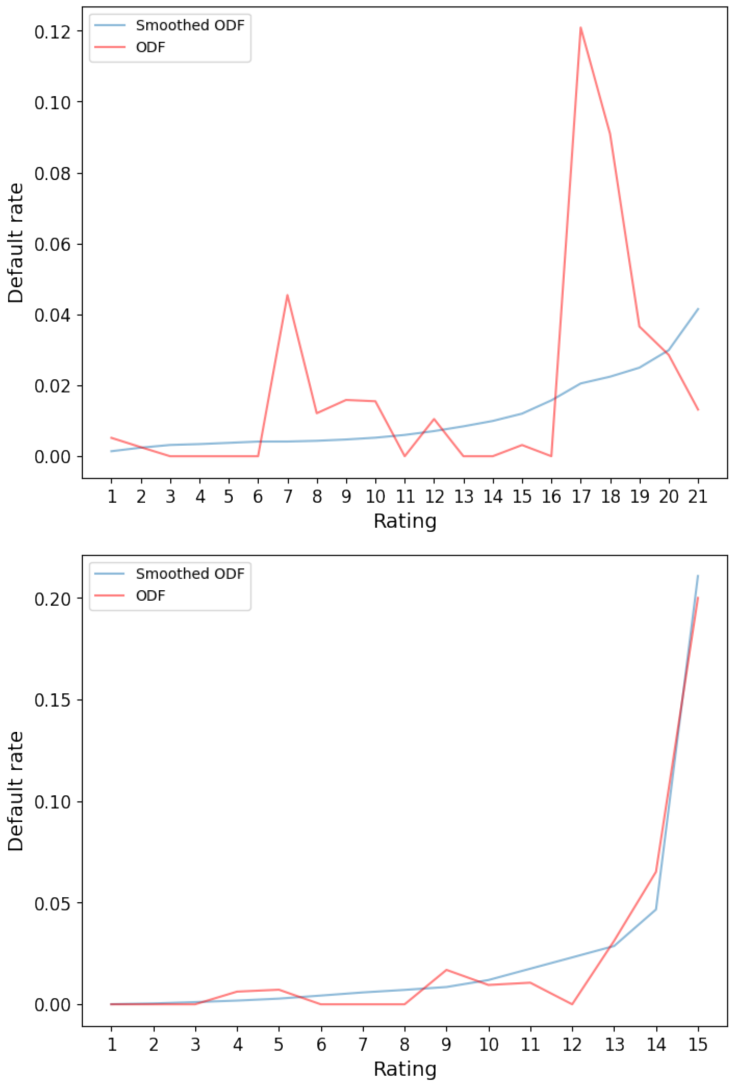

Figure 1: Smoothed ODF using quasi moment matching for an original sector (top), and an aggregated sector (bottom).

The single factor model currently used as a base case is in equation (1). A common approach to utilizing it is to select a very general economic indicator such as GDP and use a static sensitivity factor. However, it is easy to see how this approach can fail in complex economic situations if used on a large portfolio. One economic indicator simply cannot explain behaviors of several different sectors that are subjected to very different economic conditions. Because of this, it is likely that the solution is to incorporate a model which can consider more economic indicators.

Literature on the MV framework shows that the single-factor MV model can be considered a special case of the multi-factor MV model. Using the multi-factor approach, the predicted PD at time t is

$$\mathrm{PD}_{i,t} = \Phi\left[\frac{\Phi^{-1}(\mathrm{TTC}_i) + \sqrt{\rho_i}\alpha^T Z_t}{\sqrt{1-\rho_i}}\right], \tag{6}$$

where α is the indicator weights vector. Z_t is a vector of normal distributed systemic indicators. With this model, we assume that with in a portfolio or sector, companies with different ratings i have same indicator weights vector α [17].

While in the single factor case the systemic indicator Z represents the state of the economy using an indicator such as GDP, the standard approach to multi-factor systemic indicators using the sector level GDPs Z_i representing the state of the economy in each sector. However, we do not yet know what is the best sector level division, and information of very low-level sector level GDP may be challenging to obtain. The focus on total production can also be very problematic in the case of complex economic settings caused by the COVID pandemic since GDP is an indicator with a very high level of aggregation. According to [3] where the effect of macroeconomic factors was compared to corporate defaults and credit rating transitions, somewhat promising indicators were related to industrial activity, consumer sentiments, inflation, unemployment, and overall economic development including real GDP growth.

It is also worth noting that the quality of the credit ratings have been questioned in the literature. In particular, the ratings before the financial crisis has been questionable [4]. There appears to be a widespread agreement that the performance of the credit rating agencies has been unreliable and that their ratings and downgrades played pivotal roles in the global financial crisis and the European sovereign debt crisis [12]. The criticism motivated by the financial crisis also induced new policies, regulations, and guidelines which in turn has affected the methods and usage of credit ratings. It is thus possible that this has also largely affected data so that there are disparities between the credit ratings before and after the crisis.

4 Data

Data in this project was obtained from several sources: SEB and Global Credit Data (GCD)[1] rating migration data, free access economic data from Eurostat and the Wharton Business School database. Section 4.1 and 4.2 explain rating migration and economic indicators respectively in more detail, Section 4.3 introduces steps taken to process raw data in our experiments.

[1] https://www.globalcreditdata.org/

4.1 Rating migration data

Original aggregated data was received from the SEB and it was subject to our NDAs with the bank. Therefore, specific information of the data can not be explained in detail, and part of the information, such as the relevant sector codes or the number of companies in a specific rating level is hidden or the numbers are. The data included specific industry groups, countries, number of customers, and rating transition during the given timeline. Besides the original aggregated data from the SEB, we used the GCD as another source for comparative data. The data from the GCD included parameters that fit well into our case study. They provided specific industry classifications, countries, ratings from multiple providers, and default information.

Code	Industry
A	Agriculture, Hunting, Forestry and Fishing
B	Oil, Gas and Mining
C	Manufacturing
D	Electricity, Gas and Water Supply
E	Construction
F	Wholesale, Retail, and Services
G	Transportation and Storage
H	Finance and Insurance
I	Real Estate, Rental and Leasing
J	Public Administration and Defense
K	Individual and Household
X	Other

Table 1: Aggregated industry code

A custom industry code mapping shown in Table 1 was created so that data from SEB and GCD can be used together. In addition, the mapping reduces the overall number of industries thus alleviating the sparsity in the sector-segmented data.

Figure 2 shows a transition matrix obtained from a portfolio consisting of several company sectors from an early stage of a economic cycle after the financial crisis. The T0 refers to the credit rating at the beginning of the one-year period and T1 is the credit rating at the end. Figure 3 shows the same but at a later time after the financial crisis. We observe that the overall pattern remains similar, but a larger portion of the companies have the A rating. The biggest difference is in the B rating: It seems that a lot of the companies from the C to CCC-rated companies have moved to the single B rating. This indicates that the credit ratings do not behave "linearly".

These graphs also highlight another major problem: The prevalence of defaults in the D column is extremely low. Many of the columns are empty or the number of defaults is in the single digits even though we combine several sectors and representing a large portfolio of thousands of companies large companies for a single bank. This means that a massive amount of default data is needed to make meaningful generalizations.

The empirical data also suggests that combining multiple company sectors or using too large sectors can cause problems since the data is very non-smooth and there are large disparities. It is also possible that the non-smooth behavior of the ratings is caused by not

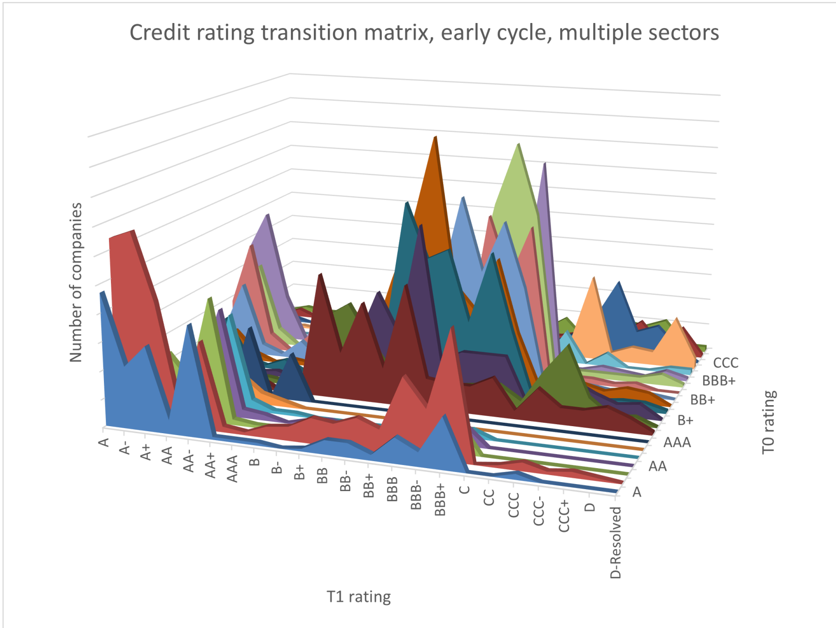

Figure 2: Transition matrix of a portfolio consisting of several company sectors from an early stage of economic cycle after the financial crisis. The graph represents a large portfolio consisting of thousands of companies.

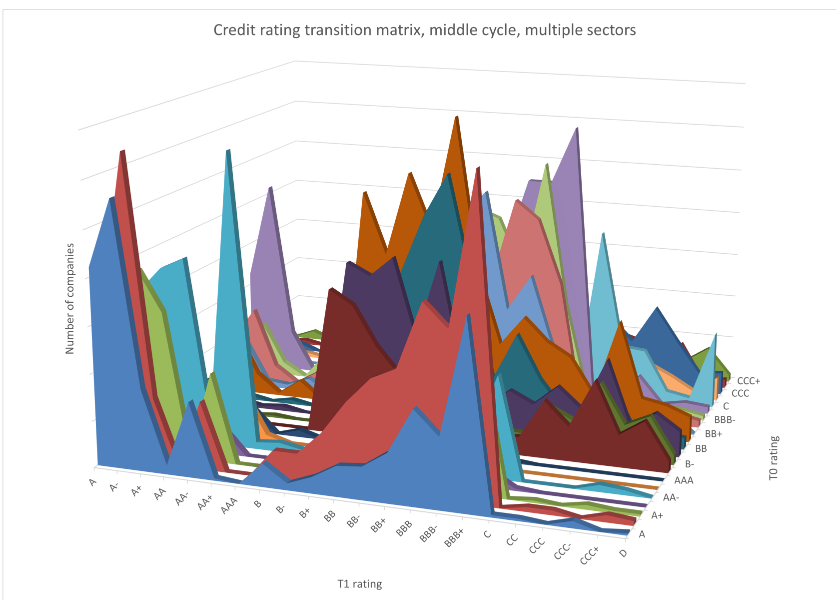

Figure 3: Transition matrix of a portfolio consisting of several company sectors from a mature stage of an economic cycle

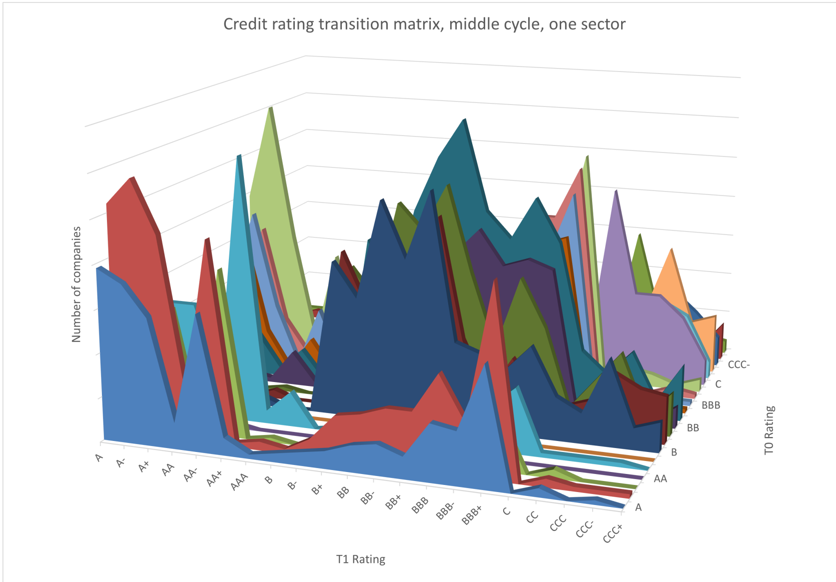

Figure 4: Transition matrix of a portfolio consisting of one company sector from a mature stage of an economic cycle

having enough data, and if enough company data was observed in each sector, these sectors as shown in Figures 2, 3 and 4 would also turn out smooth.

Figure 4 shows the transition matrix of a single large sector. We still see the strange division between the A and B sectors. This can simply be explained by too general sector division. It is important to highlight that the sector divisions have a lot of sub-categories. Here the "one sector" refers to a large group of different sub-sectors. For example, the "manufacturing" sector contains very different businesses such as the manufacturing of medical instruments or the manufacturing of consumer electronics. It is hard to estimate how many sector categories would be needed to accurately model economic conditions in each, but by looking at how vastly different economic areas are combined in the largest categories of NACE segmentation, it appears that sector division probably needs to contain tens of hundreds of different sub-sectors to produce meaningful results. This of course further worsens the problem with not having enough data: More data is needed to populate the matrix and particularly the default column if each sector has its own transition matrix.

Another possible explanation for the graph is that credit ratings are not used in a systematic way. For example, companies with A rating are much more likely to move to BBB+ rating than the other B ratings. It is possible that some of the rating levels are simply used less often and a change in credit rating occurs only due to major disruptions. By looking at the data we also see there is a significant difference when comparing the low end of the A to the high end of B.

To sum up the conclusions regarding data, we probably need to have more empirical credit data than was was provided by SEB to produce meaningful to results.

4.2 Indicators

To implement a multi-factor version of the model, we collected various macroeconomic indicators from public sources. We used two main sources: The World Bank and Eurostat. The purpose of these indicators is to extend the accuracy of the model over simple GDP as an only macroeconomic indicator. Therefore, we selected multiple different indicators that can be applied to separate countries in the data. Based on the research, macroeconomic indicators were selected to represent overall economic activity and changes in the macroeconomic state. For example, studies [3] examined the relationships between the effect of macroeconomic factors and to corporate defaults and credit rating transitions, mildly promising indicators include industrial activity, consumer sentiments, inflation, unemployment, and overall economic development including real GDP growth. Thus, we selected indicators that represent these areas from multiple perspectives.

We have selected from the World Bank open data source consumers' Final consumption expenditure and gross domestic savings to represent consumer sentiments. Exports of goods and services, changes in inventories, imports of goods and services, merchandise exports, net trade in goods and services, and energy use to represent economic and industrial activity. In addition, we have also used different unemployment measures (overall unemployment, 15-24 years old unemployment), GDP, GDP Growth, and inflation rate. The large dataset from World Bank is represented in Table 2.

Finally, Table 3 shows example data of indicators in the smaller set with Eurostat main macroeconomic measures for Finland from 2009 to 2019. Table 3 represents the annual changes of the chosen indicators from Eurostat. We have added also public indicator public consumption and real estate market indicator for housing prices to get better results in line with the overall economic state of the country.

For most of the indicators, we had data available from 1990 to 2019. Data for 2020 was not completely released yet. Therefore, it is important to notice that the effects of the current COVID-19 pandemic may not be accurately represented in the data. However, previous crises situations in observed countries are inside the data collection period representing different unusual and high-risk time periods including the financial crisis and the 1990's depression. However, the systematic collection of public data is relatively new, many of the indicators particularly in the Eurostat set are simply not available. Also, we can see data disparities between non-EU country Norway when comparing to other Nordic countries. Figure 5 shows how the columns become increasingly depopulated towards the 1990s. Due to this, our data indicator selection was largely limited.

These indicators may not portray the best possible combination of indicators, since their selection was affected by limitations such as the availability of such data for all of the countries. The further we go back in history, it seems that not only the data becomes more sparse, but the definition and collection methods may not be identical in all countries. However, we expect that using these indicators we can beter illustrate improve function of the multi-factor solution.

4.3 Data preprocessing

The data came from a few countries that are closely related. This was to ensure that country-level economic indicators are more likely to have good correlations with the default

Table 2: Large indicator set obtained from World Bank

	Indicator		Indicator
1	Trade in services (% of GDP)	26	Goods and services expense (% of expense)
2	Foreign direct investment, net outflows (% of GDP)	27	Interest payments (% of revenue)
3	Current account balance (% of GDP)	28	Interest payments (% of expense)
4	Foreign direct investment, net inflows (% of GDP)	29	Expense (% of GDP)
5	Domestic credit to private sector by banks (% of GDP)	30	General government final consumption expenditure (annual % growth)
6	Claims on central government (annual growth as % of broad money)	31	Households and NPISHs Final consumption expenditure (annual % growth)
7	Monetary Sector credit to private sector (% GDP)	32	Households and NPISHs Final consumption expenditure per capita growth (annual %)
8	Claims on private sector (annual growth as % of broad money)	33	Final consumption expenditure (annual % growth)
9	Broad money growth (annual %)	34	Gross national expenditure (% of GDP)
10	Consumer price index (2010 = 100)	35	Exports of goods and services (annual % growth)
11	Inflation, consumer prices (annual %)	36	Gross fixed capital formation (annual % growth)
12	Claims on central government, etc. (% GDP)	37	Gross capital formation (annual % growth)
13	Domestic credit to private sector (% of GDP)	38	Imports of goods and services (annual % growth)
14	Net lending (+) / net borrowing (-) (% of GDP)	39	External balance on goods and services (% of GDP)
15	Social contributions (% of revenue)	40	Trade (% of GDP)
16	Revenue, excluding grants (% of GDP)	41	Inflation, GDP deflator (annual %)
17	Taxes on goods and services (% of revenue)	42	GDP growth (annual %)
18	Taxes on goods and services (% value added of industry and services)	43	GDP per capita growth (annual %)
19	Customs and other import duties (% of tax revenue)	44	Gross domestic savings (% of GDP)
20	Taxes on international trade (% of revenue)	45	Gross savings (% of GNI)
21	Other taxes (% of revenue)	46	Gross savings (% of GDP)
22	Tax revenue (% of GDP)	47	DEC alternative conversion factor (LCU per US$)
23	Taxes on income, profits and capital gains (% of revenue)	48	Official exchange rate (LCU per US$, period average)
24	Taxes on income, profits and capital gains (% of total taxes)	49	Real effective exchange rate index (2010 = 100)
25	Compensation of employees (% of expense)	50	Unemployment, total (% of total labor force) (national estimate)

Table 3: Selected economic indicators of Finland from 2011 to 2019

Indicator, R=Real	2011	2012	2013	2014	2015	2016	2017	2018	2019
RGDP	2.5	-1.4	-0.9	-0.4	0.6	2.8	3.2	1.3	1.3
Private Consumption	-2.6	-0.6	1.3	3	3.6	4	3.4	4.2	2.9
Public Consumption	1.9	-0.4	0.3	-1.8	-1	1.5	-1.1	1.1	0.1
RGF Capital Formation	-15.1	-7.3	-5.8	3.9	4.9	6.4	8.5	2.9	3.6
R Exports	-24	-2.9	-4.6	15.5	11	3.6	18.5	-2.8	2.2
R Imports	-19.2	-10.5	-3.6	6.9	11.4	10.8	13.5	7.1	0.2
CPI	1.6	3.2	2.2	1.2	-0.2	0.4	0.8	1.2	1.1
Unemployment	7.8	7.7	8.2	8.7	9.4	8.8	8.6	7.4	6.7
Retail and Trade	2.7	1.6	-0.8	-0.9	0.4	1.5	3.1	2	2.6
Industrial Production	2	-2.2	-3.1	-1.8	-1	4.2	3.4	3.3	1.7
House Price	3.2	2.4	1.2	-0.4	0	0.8	1.6	1	1

rates. Moreover, we omitted time periods with the data sparse and few observed defaults. Following the guidelines from SEB, we also removed transitions where there are changes in sector, entrance, or exit (missing starting or ending rating). Ratings in the GCD dataset were mapped to the custom rating of SEB to allow using both in our experiments. There were mismatches in sector codes from the two datasets. Therefore, we aggregated some of the sectors based on their NACE codes to form our own 12 industry codes in Table 1. However, sector K was later removed as it is either missing or has very few data points in both datasets.

Transition matrices for the calibration experiments are compiled from SEB and GCD data and can contain rating migration in certain years, countries, and industries. At a higher level of granularity, the data is usually more sparse, making calibration more prone to numerical errors. To counteract this, we added $\epsilon_1 = 10^{-10}$ to every element of the transition matrices and $\epsilon_2 = 10^{-4}$ to the row sum when calculating rating specific PD.

Aside from the conventionally used GDP growth (denoted in the Table 4 and 5 as 'Single'), we experimented with several other choices of indicators. As introduced in Section 4.2, we had two sets of indicators. One was hand-selected indicators in Table 3 (denoted as 'Small'), while the other 2 (denoted as 'Large') was only partially filtered to leave out the most irrelevant indicators. For both indicator sets, the data spans several economic cycles, as it is required that the indicators must be normalized to applied to the MV model

$$\bar{Z}_t = \frac{Z_t - \mu(Z)}{\sigma(Z)}.$$

For the larger set of indicators, we used principal component analysis (PCA) to reduce the number of indicators that would be used to fit the MV model and make use of the correlations between indicators (shown in the heatmap in Figure 6). We would use different numbers of PCA components, ranging from 1 to 39, as indicators (denoted as 'PCA-n-components') for each experiments. This sheds light on the effect of indicators on the performance of the model, whether more indications would improve prediction performance or make fitting more challenging.

Indicator	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Real gross domestic product	4.2	5.0	5.3	2.7	2.0	3.2	2.0	1.3	1.0	4.0	2.7	2.4	2.9	0.2	-1.3	0.4	1.1	2.6
Real private consumption	3.7	6.0	3.1	2.9	3.8	4.2	2.3	2.9	3.0	5.2	4.9	5.1	5.3	1.9	-0.2	3.7	2.3	3.5
Real public consumption	1.0	2.5	3.2	3.5	3.0	2.1	4.3	2.8	1.3	0.8	2.4	2.0	2.0	2.2	4.3	2.1	1.0	1.5
Real gross fixed capital formation	3.1	9.2	14.6	12.2	-5.0	-3.3	-0.5	-0.6	0.3	10.1	12.0	9.2	12.2	1.1	-6.8	-6.4	7.5	7.5
Real exports of goods and services	5.0	10.0	7.7	0.6	2.8	3.2	4.4	-0.2	-0.1	1.1	0.4	-0.8	1.3	0.7	-3.6	0.2	-1.5	1.6
Real imports of goods and services	5.8	8.7	12.4	8.9	-1.6	2.0	1.7	0.9	1.3	9.0	7.8	9.2	10.1	3.5	-10.5	8.7	3.7	2.9
Gross domestic product	7.4	9.2	8.5	2.0	8.7	19.1	3.9	-0.4	3.8	10.0	11.7	11.5	5.9	11.1	-6.8	6.6	7.9	6.0
Consumer price index	2.5	-15.9	2.5	2.0	2.2	3.1	2.6	0.8	2.0	0.6	1.5	2.4	0.8	3.4	2.3	2.3	1.3	0.3
Unemployment	5.5	4.8	3.9	3.1	3.0	3.2	3.4	3.7	4.2	4.3	4.5	3.4	2.5	2.7	3.3	3.7	3.4	3.3
Retail trade							1.8	4.3	3.7	2.8	3.1	5.5	6.6	1.7	0.1	2.0	2.1	3.2
Industrial production		5.4	3.8	-1.1	-1.3	3.1	-0.5	-0.1	-1.8	-1.2	-0.2	-2.2	-1.3	0.2	-3.6	-5.6	-4.3	2.7
Government balance																		
Government debt						29.9	27.1	34.7	42.1	44.9	44.0	53.5	51.5	48.5	43.2	43.4	29.8	29.9
Current account balance																		
Net international investment position																		92.0
International investment position: Assets																		269
International investment position: Liabilities																		177
Long term yield																		
House price												13.7	12.6	-1.1	1.9	8.3	8.2	6.7
Oil production								3 104	2 971	2 955	2 673	2 526	2 239	2 030	2 017	1 874	1 759	1 602
Population															1.3	1.2	1.3	1.3

Figure 5: Country profile indicators available on Eurostat for Norway.

5 Methods

The aim of the MV model is to predict PD given historical observed default frequencies ODF_h. The general procedure for predicting PD with the MV model is as follows:

1. Compute the ODF_{TTC} by averaging the historical ODF_h
2. Calibrate the ρ_i and indicator weights α_i of the MV model with ODF_{TTC}
3. Smooth the ODF_h using QMM if separate ρ value for each rating class is used
4. Remove the trend from ODF_h by applying the MV model in reverse
5. Compute PD_{TTC} by averaging the reversed MV ODF_h
6. Apply MV model with the PD_{TTC} and economic indicator Z_t to get predictions for PD_t.

For our project, optimization was done using tools from the tensor-based computing package PyTorch [13]. The procedures and technical details for calibrating the multi-factor MV model and QMM smoother are in Section 5.1 and 5.2. The code for implementing QMM and MV can also be found in the Appendices.

5.1 Merton Vasicek calibration

For our project, we implemented the multi-factor MV model introduced in Section 3. For the MV model, economic indicators Z should be positively correlated to PD, which is not always the case. Therefore, this condition was enforced for each indicator by multiplying

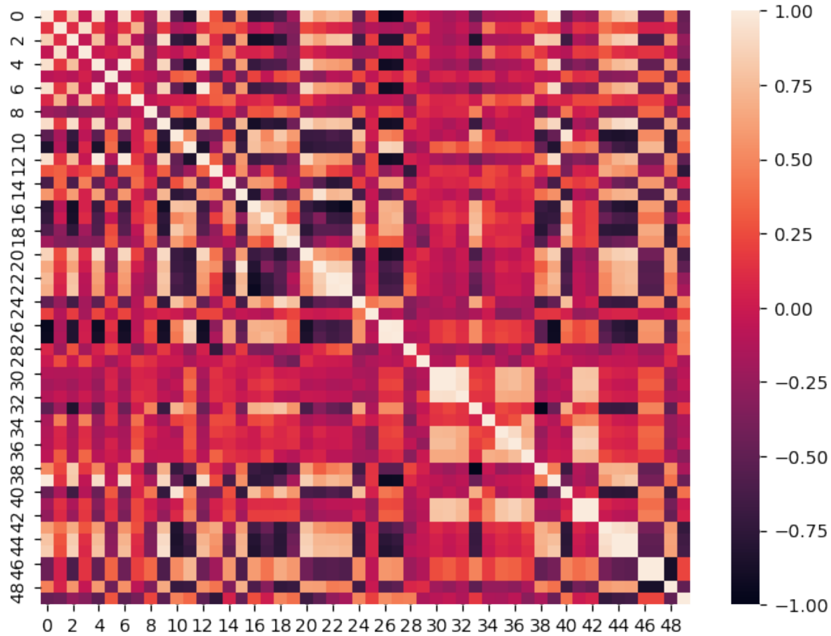

Figure 6: Heatmap for the large set of indicators

the normalized $\bar{Z}$ with the sign of its Pearson correlation coefficient

$$\bar{Z}^{+} = \text{sign}[r(\bar{Z}, \text{ODF}_h)]\bar{Z}.$$

This sign adjustment process was done for every set of transitions.

For α, we have the constraints $\alpha_k \geq 0$ and $\sum \alpha_k^2 = 1$ for $\alpha^T Z_t$ to have unit variance [16, 17]. In our implementation, the α is defined as the square root of the Softmax function of raw weights $w \in R^K$

$$\alpha_k = \sqrt{\frac{e^{w_k}}{\sum_{i=1}^{K} e^{w_i}}}.$$

The sensitivity parameterρ is initialized at 0.05 while w is initialized as a uniform vector of 1. For the optimization of ρ and α, we used Adam optimizer [9] with separate learning rates of 0.001 for α and 0.00001 for ρ. The objective function is the negative log-likelihood loss (NLLL) over the time period T and across all rating i

$$NLLL = -\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\text{ODF}_{i,t}\log(\text{PD}_{i,t}) + (1 \text{ - } \text{ODF}_{i,t})\log(1 - \text{PD}_{i,t})\right].$$

The optimal value for the loss function is achieved when the predictions are equal to the observations $\text{PD}_{i,t} = \text{ODF}_{i,t}$. The actual NLLL-$\rho$ curve during optimization can be seen in Figure 7.

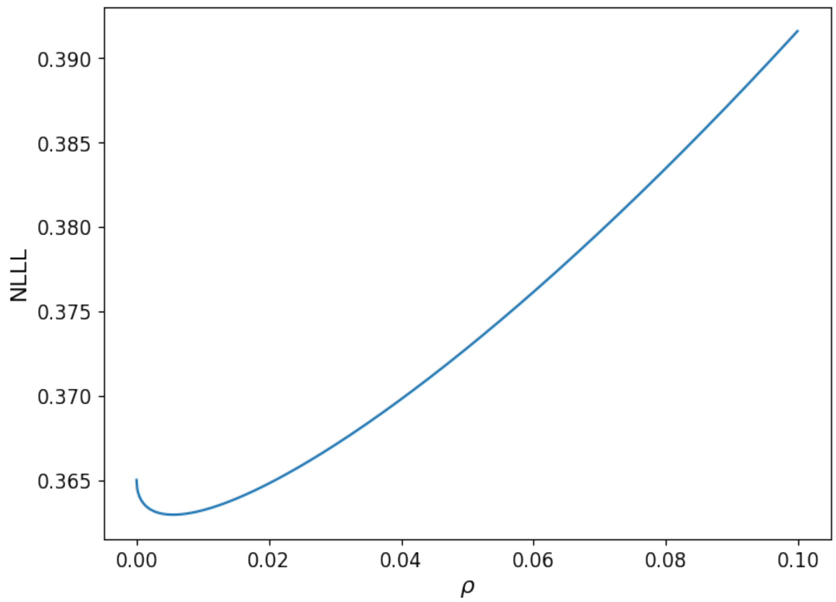

Figure 7: NLLL for the case of single ρ factor and economic indicator

5.2 Quasi moment matching calibration

The QMM smoothed PD curved is calibrated by optimizing α and β in Equation 2. The procedure for QMM smoothing was implemented according to [18]:

1. From the given transition matrix, compute $F_N(x)$ for each rating class x as shown in Equation 3
2. Compute accuracy ratio

$$\text{AR} = \sum_{x=1}^{k-1} P(X = x|N)P(X \geq x+1|D) - \sum_{x=2}^{k} P(X = x|N)P(X \leq x-1|D). \tag{7}$$

3. Calculate the terms

$$\begin{aligned} \mu &= \frac{1}{k}\sum^{k} \Phi^{-1}(F_N(x_i)) \\ \tau &= \text{Var}(\Phi^{-1}(F_N(X)))^{0.5} \\ c &= \sqrt{2}\Phi^{-1}\left(\frac{\text{AR}+1}{2}\right) \\ \sigma^2 &= \frac{\tau^2}{1+p(1-p)c^2} \\ \mu_N &= \mu + p\sigma c \\ \mu_D &= \mu - (1-p)\sigma c, \end{aligned}$$

where p is the observed marginal default rate.

4. Initialize α and β as

$$\alpha_0 = \frac{\mu_D^2 - \mu_N^2}{2\sigma^2} + \log(\frac{1-p}{p})$$
$$\beta_0 = \frac{\mu_N - \mu_D}{\sigma^2}.$$

5. Optimize α and β using Adam optimizer with learning rate of 0.1. The loss function is the mean square error between observed vector $(p_{true}, \text{AR}_{true})$ and the $(p_{qmm}, \text{AR}_{qmm})$ obtained by recalculating $PD(x)$ using Equation 2 and AR using Equation 7.
6. The smoothed PD for each rating grade is obtained using Equation 2.

5.3 Principal Component Analysis

Because we are aim to include a large set of parameters, many of our indicators are likely correlated. This means there is a risk of overfitting. To counter this, we used Principal Component Analysis (PCA) for dimensionality reduction. PCA is a multivariate technique that analyzes a data matrix in which observations are described by several inter-correlated quantitative dependent variables. We then change the basis of all of the vectors from our random distributions to represent the data using new orthogonal variables called principal components which maximize the variation along the axis. The components are ordered by total variation along the new axis. Thus, PCA removes excess inter-correlation for representing the most meaningful relations of the data with fewer variables. PCA has been shown to be a robust method for avoiding overfitting and improving predictions in machine learning models. [5]

6 Experiments

Each of our experiments was tested both with a single ρ – i.e. one factor for the whole portfolio – and separate ρ – i.e. one factor for each rating class. In the case of separate ρ, we consolidated the calculation for the whole portfolio by taking the weighted average of the different rating classes with respect to their share in the whole portfolio.

The average performance of the MV model over the chosen time period was measured by computing the root mean square error (RMSE) between the predicted PD and the ODF of the given SEB portfolio over a number of periods

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(\text{PD}_t - \text{ODF}_t)^2}{T}}.$$

As the core objective of our project is improving the MV model, RMSE gives us a good overview of the model performance. For more specific use-cases, other measures of performance can be more appropriate.

6.1 Portfolio level calibration

At portfolio level calibration, we had two separate experiments. In the first experiment (portfolio level - train), the whole migration data from the SEB portfolio was used to calibrate the MV model, which is the current practice. In the second experiment (portfolio level - test), data from the last year was left out for testing purposes while the rest of the data was used to calibrate the model. The result from the second experiment served as a reference to determine whether the first setup overfits the data. Furthermore, leaving out the latest year would also reflect the conditions in real applications where transitions of the latest year are unknown, and only economic indicator predictions are available for predicting future PD.

6.2 Sector level calibration

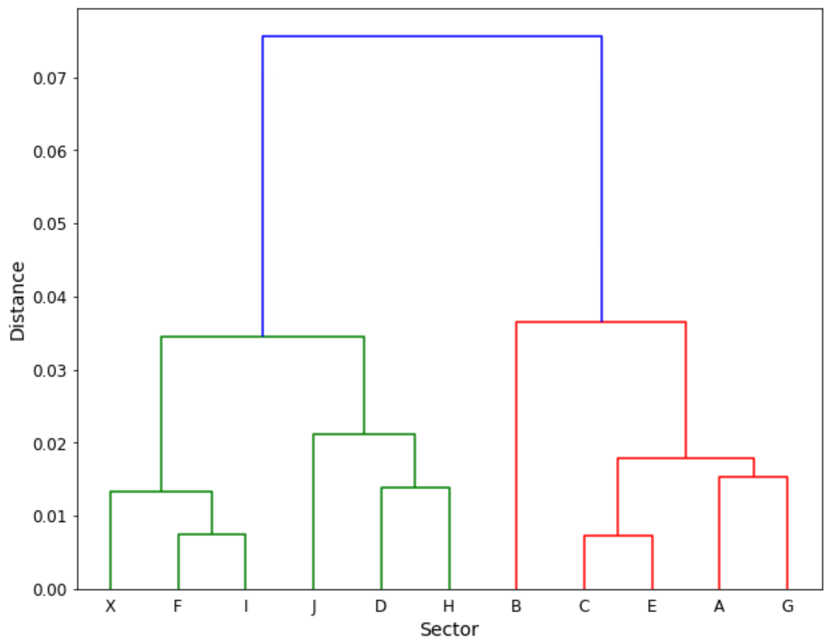

Figure 8: Dendrogram for sector clustering

At sector level, we produced the transition matrices and calibrated the MV model separately for each sector or sector group using GCD data. The calibrated models would then be used to compute the predicted PD for the SEB portfolio based on the respective sector weights in the portfolio.

Two different groupings were used to calibrate the model in terms of market sectors. Those include the sector breakdowns into a fully differentiated sector analysis (referred to as *sector level - 11*) as shown in Table 1 and a grouped sector arrangement (referred to as *sector level - 4*)

$$\text{Sector level - 11:} \ (A)\ (B)\ (C)\ (D)\ (E)\ (F)\ (G)\ (H)\ (I)\ (J)\ (X),$$
$$\text{Sector level - 4:} \ (A, C, E, G)\ (D, H, J)\ (B)\ (F, I, X),$$

where the parentheses encode the sector grouping. The more general grouping can be seen in Figure 8 The second grouping was obtained with agglomerative hierarchical clustering

[2] with ward distance implemented in the Scikit-learn package [14]. The features used for clustering were the historical ODF of each sector. This aimed to cluster sectors that have the most similar movement in ODF through the observation period. The larger sector groups alleviates the issue with data sparsity that occurs when the data is segmented.

7 Results

Due to a small data sample, the model did not converge in some cases. These are shown as blanks in the resultant data tables.

We start with the analysis of the results at portfolio level. The RMSE results for the experiments are in Table 4. In the portfolio-train experiment, the best result for single ρ is achieved with the single PCA component as the indicator. Multiple indicators yield worse results overall compared to single and PCA-1. Also, the RMSE increases with the number of PCA components in this case. On the other hand, fitting separate ρ for each rating gave significantly more accurate results overall, with improvement to the RMSE when the number of PCA components used is higher.

The results from the portfolio-test experiment are less clear-cut. All prediction RMSEs in the test experiment are low compared to those of the train experiment RMSE. In the single ρ test experiments, the results were mixed, with PCA of 1, 10, and all component performed worse than the highlighted base case of single ρ and single GDP growth indicator while PCA of 2, 5, and 20 components vastly outperform the base case. In the separate ρ test experiment, the results were consistently better than or comparable to the base case, with the best results from PCA-5 and PCA-20.

Indicators	Portfolio level (train)		Portfolio level (test)	
	Single ρ	Separate ρ	Single ρ	Separate ρ
Single	0.00283	0.00139	0.00116	0.000500
Small	0.00329	0.000961	0.000325	0.00107
Large	0.00366	0.00110	-	-
PCA-1	**0.00248**	-	0.00128	-
PCA-2	0.00356	0.00122	**0.0000366**	0.000785
PCA-5	0.00361	0.00110	0.0000421	**0.000145**
PCA-10	0.00363	0.000805	0.00232	0.001217
PCA-20	0.00375	0.000800	0.000180	0.000430
PCA-All	0.00384	**0.000492**	0.00227	0.00128

Table 4: RMSE of predicted PD with portfolio level calibration

For the sector level experiments, despite the more prevalent issue of convergence in the tests, Table 5 shows that the RMSEs in both cases of 4 and 11 sectors are lower than the base case of the portfolio-train experiment. The RMSE results at sector level are relatively consistent with the PCA-2 and PCA-5 indicators giving the best overall results. We can also observe that grouping the sector reduces the number of convergence failure but give slightly higher RMSE. In both sector level experiments, separate ρ gave higher RMSE compared to single ρ and were also more likely to fail to converge.

Indicators	Sector level - 11		Sector level - 4	
	Single ρ	Separate ρ	Single ρ	Separate ρ
Single	0.00183	-	0.00191	0.00133
Small	0.00123	**0.00106**	0.00129	0.00140
Large	-	-	-	-
PCA-1	-	-	-	-
PCA-2	0.00105	-	**0.00109**	**0.00120**
PCA-5	**0.00103**	0.00115	**0.00109**	**0.00120**
PCA-10	0.00116	0.00126	0.00124	0.00142
PCA-20	-	0.00186	0.00144	-
PCA-All	-	0.00237	0.00143	-

Table 5: RMSE of predicted PD with sector level calibration

8 Discussion

Although the portfolio-test experiments yielded results significantly better than those of the portfolio-train experiments, their main goal was to determine whether training on the whole dataset overfits the model. From the results obtained, we can reasonably say that it is unlikely that overfitting happens, and therefore, the portfolio-train results should be a more reliable benchmark for the performance of the MV model under a variety of economic conditions.

The experiments suggest that RMSEs improve with higher data granularity up to a certain point. Both separate ρ and sector calibration gave better results than the conventional portfolio level single ρ calibration. This confirms the original assumption that a single ρ factor fails to capture how companies of different ratings and sectors react to different economic indicators. By appropriately segmenting the migration data and apply the MV model separately, more accurate PD prediction can be achieved.

Similarly, the use of more diverse economic indicators also improves the quality of PD prediction. With a broader set of indicators, changes in PD can be more reliably predicted. Also appropriately applying PCA on a large set of indicators gives comparable or better results than a single or a few hand-selected indicators.

Lastly, although the results are incomplete, it has shown us that there appears to be a trade-off between higher level granularity of the data (which can produce better results) and the sparsity that occurs at that level of granularity. In theory, calibrating separate ρ for different sectors would give the best result as the subset of companies of the same sector and the same rating should react uniformly to the economic conditions. However, in reality, the lack of data limits the capability to fully model the tendencies of companies at that level of segmentation.

9 Conclusions and future direction

Our original main objective was to enhance the existing MV model of SEB to better work with the challenging macro-economic environment under the COVID-19 pandemic. We have successfully shown the flaws of the single factor model by examining the quality of the credit data and how it can be significantly improved by utilizing multiple indicators. While the results are not perfect and there is probably room for improvement by focusing on better data collection and data engineering, it is justified to claim that a large share of the bad estimates given by the single factor model is based on too general systemic indicators and poorly populated transition matrices and bad sector division. It is thus less likely that the exaggerated default projections are mainly caused by government subsidies or other supportive actions. While the government-induced restrictions have highly affected parts of the service industry bringing the general economic indicators such as GDP down, such systemic indicator does not explain the changes in a portfolio consisting of large corporate and institutional loans well enough. By including more indicators such as government spending or consumer index, we can create a more complete economic picture and improve the prediction of PDs in uncertain times such as the COVID-19 pandemic.

In conclusion, it is easy to see why institutions such as GCD have been established. The portfolios of individual banks simply are not comprehensive enough to make accurate modeling. The existing modeling frameworks are still functional for the current complex economic environment, though they can be improved by better utilizing economic data and data engineering methods such as PCA.

The results that we obtained can be improved in the future. Firstly, the calibration of the MV model should be made to be more robust to limit or eliminate cases of failure to converge. Secondly, to check for overfitting proper cross-validation should be carried out. Moreover, grouping of ratings could also be done to limit the data sparsity in some cases. Lastly, some hyperparameters such as the number of PCA components or the number of sector clusters could be done to determine the best setting for calibrating PD.

Acknowledgement

This project would not have been possible without the guidance and collaboration from numerous people and organizations. We want to thank to Professor Ahti Salo who gave us the opportunity to work on this meaningful project. We would like to express our gratitude to SEB and GCD for generously providing us the crucial data upon which this project was built.

References

[1] S Agarwal et al. "Did the Payment Protection Program Help Small Businesses? Evidence from Commercial Mortgage-backed Securities". In: (2020).

[2] W HE Day and H Edelsbrunner. "Efficient algorithms for agglomerative hierarchical clustering methods". In: *Journal of Classification* 1.1 (1984), pp. 7–24.

[3] S Figlewski, H Frydman, and W Liang. "Modeling the effect of macroeconomic factors on corporate default and credit rating transitions". In: *International Review of Economics & Finance* 21.1 (2012), pp. 87–105.

[4] K Galil. "The Quality of Corporate Credit Rating: An Empirical Investigation". In: *SSRN Electronic Journal* (Oct. 2003). DOI: 10.2139/ssrn.406681.

[5] A George. "Anomaly Detection based on Machine Learning Dimensionality Reduction using PCA and Classification using SVM". In: *International Journal of Computer Applications* 47 (June 2012), pp. 5–8. DOI: 10.5120/7470-0475.

[6] International Organization of Securities Commission. *Observed Impact of COVID-19 Government Support Measures on Credit Ratings*. 2021. URL: https://www.iosco.org/library/pubdocs/pdf/IOSCOPD671.pdf.

[7] T Jensen, D Lando, and M Medhat. "Cyclicality and Firm-size in Private Firm Defaults". In: *International Journal of Central Banking (Forthcoming)* (2016).

[8] M Kalkbrener and A Onwunta. "Validating structural credit portfolio models". In: *Model risk—identification, measurement and management. Risk Books, London* (2010), pp. 233–261.

[9] D P Kingma and J Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[10] Allan M Malz. *Financial risk management: Models, History, and Institutions*. Vol. 538. John Wiley & Sons, 2011.

[11] R C Merton. "On the pricing of corporate debt: The risk structure of interest rates". In: *The Journal of finance* 29.2 (1974), pp. 449–470.

[12] I Moosa. "The regulation of credit rating agencies: A realistic view". In: *Journal of Banking Regulation* 18 (Apr. 2017). DOI: 10.1057/jbr.2015.9.

[13] A Paszke et al. "Pytorch: An imperative style, high-performance deep learning library". In: *arXiv preprint arXiv:1912.01703* (2019).

[14] F Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[15] PwC. *In depth: A look at current financial reporting issues*. 2014. URL: https://www.bis.org/fsi/fsisummaries/ifrs9.pdf.

[16] M Pykhtin. "Multi-factor adjustment". In: *Risk* 17 (2004), pp. 85–90.

[17] H Ribom. "Consolidating Multi-Factor Models of Systematic Risk with Regulatory Capital". PhD thesis. KTH Royal Institute of Technology School of Engineering Sciences, 2018.

[18] D Tasche. "The art of probability-of-default curve calibration". In: *Journal of Credit Risk* 9.4 (2013), pp. 63–103.

[19] O Vasicek. "The distribution of loan portfolio value". In: *Risk* 15.12 (2002), pp. 160–162.

Appendices

A.1 Updated results

The results in Table 6 and 7 was updated with ρ factor parameterized as

$$\rho_i = \frac{1}{1 + e^{-w_{\rho_i}}}.$$

The learning rate for $-w_{\rho_i}$ is 0.01 with Adam optimizer. Furthermore, the portfolio level test experiment was done with 5-fold cross-validation.

The new results show that there the model was likely overfitted in the case of *Large* and *PCA-All* indicator sets. Despite this, fitting with an appropriate set of indicator and separately for rating was shown to improve the quality of the calibration at portfolio level.

With these updated results, we can conclude that the best performing method for calibration the MV model would be calibration on the sector level using a single ρ factor across all rating and 2 to 10 PCA factors of a large indicator set. However, it should also be noted that these results were produced using solely the SEB portfolio as the test data, and further tests should be done to accurately assess the performance of the model.

Indicators	Portfolio level (train)		Portfolio level (test)	
	Single ρ	Separate ρ	Single ρ	Separate ρ
Single	0.00283	0.00139	0.00266	0.00212
Small	0.00353	0.000983	0.00263	0.00173
Large	0.00374	0.00113	0.00403	0.00303
PCA-1	**0.00248**	0.00197	0.00223	0.00154
PCA-2	0.00356	0.00488	**0.00183**	0.00210
PCA-5	0.00361	0.00506	0.00206	0.00193
PCA-10	0.00364	0.000905	0.00275	**0.00132**
PCA-20	0.00380	0.00467	0.00352	0.00322
PCA-All	0.00385	**0.000614**	0.00262	0.00312

Table 6: RMSE of predicted PD with portfolio level calibration (updated).

Indicators	Sector level - 11		Sector level - 4	
	Single ρ	Separate ρ	Single ρ	Separate ρ
Single	0.00183	0.00124	0.00191	0.00133
Small	0.00115	**0.00111**	0.00129	0.00140
Large	0.00135	0.00250	0.00140	0.00232
PCA-1	0.00235	0.00190	0.00243	0.00194
PCA-2	0.00105	0.00115	**0.00109**	**0.00120**
PCA-5	**0.00103**	0.00119	**0.00109**	**0.00120**
PCA-10	0.00114	0.00138	0.00124	0.00143
PCA-20	0.00145	0.00335	0.00144	0.00184
PCA-All	0.00149	0.00347	0.00141	0.00235

Table 7: RMSE of predicted PD with sector level calibration (updated).

A.2 Python codes

Listing 1: Merton-Vasicek model fitting code

```
class MertonVasicek(torch.nn.Module)
    '''
    Simple class for calibrating the rho factor for (multifactor) Merton-Vasicek model

    Attributes:
    rho_dim: int
        1 for the case of single factor for the whole portfolio, otherwise muber of ratings
    z_dim: int
        Number of indicators, 1 for single factor model
    I: ndarray
        Sign indicator for correlation between PD and ZeroDivisionError
    w: ndarray
        Weights for indicators (weights will go throgh softmas transformation so that they sum to 1)
    rho: float or arraylike
        Sensitivity factor
    '''

    def __init__(self, rho_dim, z_dim, I, rho0):
        '''
        Constructor for the class

        Parameters:
        rho_dim: int, required
        z_dim: int, required
        I: ndarray , required
        rho0: float, required
            initial value for rho in (0,1)
        '''

```

```python
        super(MertonVacisek, self).__init__()
        self.rho_dim = rho_dim
        self.z_dim = z_dim
        self.I = IOError
        self.w = torch.nn.Parameter(torch.ones(1, self.z_dim).double())
        self.rho = torch.nn.Parameter(torch.ones(1, self.rho_dim).double() * rho0)

    def forward(self, odf_ttc, z):
        '''
        Forward function for calculating PD

        Parameters:
        odf_ttc: arraylike, flat, required
            Trough-the-cycle mean ODF or PD, float if dim = 1, array of length dim otherwise
        z: arraylike, ndarray, required
            Economic indicators (see fit_mv function)
        '''

        # Transform numpy inputs to tensors
        if not torch.is_tensor(odf_ttc):
            odf_ttc = torch.tensor(odf_ttc.copy()).reshape(-1)
        if not torch.is_tensor(z):
            z = torch.tensor(z.copy()).reshape(-1, self.z_dim)

        # Applying softmax to raw weights
        w_n = torch.sqrt(torch.nn.functional.softmax(self.w, dim=1))

        # MV, 10e-10 is added to the pd to make sure the opimization process is more numerically stable
        N = torch.distributions.normal.Normal(0,1)
        pd = N.cdf((N.icdf(odf_ttc) + (z * w_n) @ self.I * torch.sqrt(self.rho)) / torch.sqrt(1 - self.rho))

        return pd

    def get_rho(self):
        return self.rho.detach().numpy()

    def get_weight(self, odf, z):
        return torch.sqrt(torch.nn.functional.softmax(self.w, dim=1)).detach().numpy()

    def get_pd_ttc(self, odf, z):
        '''
        Reverse function to eliminate Z factor from historic PD

        Parameters:
        odf: ndarray, required
            Observed default frequency (smoothed by qmm in the case
```

```
                of separate rho)
        Z: ndarray, required
        '''

        if not torch.is_tensor(odf):
            odf = torch.tensor(odf.copy())
        if not torch.is_tensor(z):
            z = torch.tensor(z.copy()).reshape(-1, self.z_dim)

        w_n = torch.sqrt(torch.nn.functional.softmax(self.w, dim=1))

        N = torch.distributions.normal.Normal(0,1)
        odf_trend_removed = N.cdf(N.icdf(odf) * torch.sqrt(1 - self.
            rho) - (z * w_n) @ self.I * torch.sqrt(self.rho))

        if self.rho_dim == 1:
            pd_ttc = torch.mean(odf_trend_removed)
        else:
            pd_ttc = torch.mean(odf_trend_removed, axis=0)

        return pd_ttc.detach().numpy()


def nllloss(pd, odf):
    '''
    Negative loglikelihood loss

    Parameters:
    pd: ndarray, required
        Predicted PD
    odf: ndarray, required
        Historical ODF
    '''

    if odf.ndim ==1:
        pd = pd.reshape(-1)
    loss = - torch.sum(odf * torch.log(pd) + (1- odf) * torch.log(1
        - pd))
    return loss


def fit_mv(transitions, z, rating_level_rho=False, rho0=0.05, lr
    =0.00001, max_epochs=100000)
    '''Function for fitting MV model

    Parameters:
    transitions: list, required
        List of yearly transition matrices in chronological order
    z: ndarray, required
        Economic indicators, array of length N (years) for single
            indicator, matrix of N (year) x M (indicators) for
            multiple indicators
```

```
    rating_level_rho: bool, optional
        Set to True to fit separate rho for each rating_level_rho
    rho0: float, optional
        Starting value for rho
    lr: float, optional
        Learning rate for rho
    max_epochs: int, optional
        Maximum number of training iterations
    '''

    #Setting inputs for MV (get dimension, and correlation sign for
       Z factors, calculate historical ODF)
    if z.ndem == 1:
        z_dim = 1
        z = z.reshape(-1,1)
    else:
        z_dim = z.shape[1]

    if not rating_level_rho:
        odf_h = np.array([get_pd(m) for m in transitions])
        odf_ttc = np.mean(odf_h)
        rho_dim = 1
        I = torch.tensor(np.sign([np.corrcoef(z[:,i], odf_h)[0,1]
            for i in range(z_dim)])).reshape(-1,1)
    else:
        odf_h = np.vstack([get_pdx(m) for m in transitions])
        odf_ttc = np.mean(odf_h, axis=0)
        rho_dim = odf_ttc.shape[0]
        I = torch.tensor(np.sign([[np.corrcoef(odf_h[:,i], z[:,j])
            [0,1] for i in range(rho_dim)] for j in range(z_dim)]))

    # Historical ODFodf_h = torch.tensor(odf_h)
    odf_h = torch.tensor(odf_h)

    MV = MertonVasicek(rho_dim, z_dim, I, rho0)
    MV.train()

    # Training
    optimizer = torch.optim.Adam([{'params': MV.w 'lr':0.001}, {'
       params': MV.rho}], lr=lr)
    #lr_lambda = lambda x: np.exp(x * np.log(0,.1) /max_epochs)
    #scheduler = torch.optim.lr_scheduler.LambdaLR(optimizer,
       lr_lambda)

    epoch = 0
    stop = Falseloss_prev = 10

    while not stop:
        epoch = epoch + 1
        optimizer.zero_grad()
        pd = MV(odf_ttc, z)
        loss = nllloss(pd, otf_h)
```

```python
        loss.backward()
        optimizer.step()
        #scheduler.step()
        loss_current = loss.item()
        stop = np.abs(loss_prev - loss_current) < 1e-10 * np.abs(loss_prev) or epoch >= max_epochs
        loss_prev = loss = current

    return MV
```

Listing 2: QMM code

```python
class Tasche(torch.nn.Module):
    '''
    Class for optimizing parameters (alpha and beta) for quasi moment matching based on Tasche, 2013.
    '''
    def __init__(self, a_0, b_0):
        super(Tasche, self).__init__()
        self.alpha = torch.nn.Parameter(torch.tensor(a_0))
        self.beta = torch.nn.Parameter(torch.tensor(b_0))

    def forward(self, px, fn_inv):
        x = torch.tensor(fn_inv)
        px = torch.tensor(px)

        pdx = 1 / (1 + torch.exp(self.alpha + self.beta * x))
        pd = torch.sum(pdx * px)
        pxd = pdx * px / pd
        pn = 1- pd
        pnx = 1 - pdx
        pxn = pnx * px / pn

        ar = torch.sum(pxn[:-1] * torch.flip(torch.cumsum(torch.flip(pxd[1:], [0]), dim=0), [0])) - torch.sum(pxn[1:] * torch.cumsum(pxd[:-1], dim=0))

        return torch.hstack((pd, ar))

def qmm(m):
    '''
    Function for quasi moment matching. The formulation can be found in Tasche, 2013.

    Parameters:
    m:  ndarray
         Transition matrix (count)

    Return:
    Smoothed PD vector (arraylike)
    '''
    px = get_px(m)
    pd = get_pd(m)
```

```
    pn = 1 - pd

    pdx = get_pdx(m)
    pnx = 1 - pdx
    pxn = 0 if pd == 0 else pdx * px / pi
    pxn = 0 if pn == 0 else pnx * px / pn

    # Accuracy ratio
    ar = np.sum(pnx[:-1] * np.flip(np.cumsum(np.flip(pxd[1:]))) - np.cumsum(pxd[:-1]))

    fn = (2* np.flip(np.cumsum(np.flip(pxn))) - pxn) / 2
    fn_inv = norm.ppf(fn)

    mu = np.mean(fn_inv)
    tau = np.std(fn_inv)
    c = np.sqrt(2) * norm.pff((ar + 1) / 2)
    sigma = np.sqrt((tau ** 2) / (1+ pd * pn * c ** 2))
    mu_n = mu + pd * sigma * c
    mu_d = mu - pn * sigma * c
    alpha_0 = (mu_d ** 2 -mu_n ** 2) / (2 * sigma ** 2) + np.log(pn / pd)
    beta_0 = (mu_n - mu_d) / sigma **2

    tasche_smoother = Tasche(alpha_0, beta_0)

    mse = torch.nn.MSELoss()
    optimizer = torch.optim.Adam(tasche_smoother.parameters(), lr=0.1)
    epoch = 0
    stop = false
    loss_prev = 0

    while not stop:
        epoch = epoch + 1
        optimizer.zero_grad()
        res = tasche_smoother(px, fn_inv)
        loss = mse(torch.tensor([pd, ar]), res)
        loss.backward()
        optimizer.step()
        loss_current = loss.item()
        stop = np.abs(loss_prev - loss_current) < 1e-10 * np.abs(loss_prev) or epoch >= 10000
        loss_prev = loss_current

    alpha = tasche_smoother.alpha.item()
    beta = tasche_smoother.beta.item()

    return 1 / (1 + np.exp(alpha + beta * fn_inv))
```

Self Assessment

Project progress

Scope Although there were major delays to the plans, it was executed on time for the delivery of the final result. Although the scope of the project has changed from the original, the broadest objective, which is to find ways to improve the original MV model. The biggest deviation from the original objective would be that instead of explicitly includes COVID-19 and government support measure metrics, a broad set of economic indicators was employed. We believe that this can better as those indicators should work for normal economic scenarios as well as unusual events like depressions or pandemics. It should also be mentioned that we were only provided data up to 2019, which does not cover the pandemic. However, it should also be a simple matter to extend the analysis if given the new data for 2020.

Risks The late data delivery was the most severe as highlighted in the interim report. Technical difficulties have also occurred, but fortunately, was dealt with in a timely manner.

Schedule and Execution With the aforementioned delay, much of the plan was actually executed in the last period. However, the preparation the first two periods helped us complete the project on time.

Workload The workload was reasonable but should be more evenly distributed over the three periods.

Success and failings

Success The project achieved its objective of improving the MV model. It also opened up several directions through which future works can explore.

Failings The biggest unresolved issue of the implemented model is its failure to converge when the data is sparse or the sensitivity is very low. One way this can be solved is to parameterize the ρ factor similarly to the sector weights α. Unfortunately, this was realized too late and we cannot feasibly rerun the experiments in time.

Takeaways

Team Communication could have been better and more assertive. We should have set up clearer goals and deadlines to follow.

Teaching staff and client organization We think that it would be better if a dataset or open data source is ready at the inception of the project.