KENYA BANKERS
ASSOCIATION
One Industry. Transforming Kenya.

WPS/09/22

Asset Quality Assessment in the Absence of Quality Data towards Optimal Credit Intermediation

Andrew Kioi Njeru

February 2022

KBA Centre for Research on Financial Markets and Policy®
Working Paper Series

63

KENYA BANKERS
ASSOCIATION
One Industry. Transforming Kenya.

Working Paper Series

Centre for Research on Financial Markets and Policy

The Centre for Research on Financial Markets and Policy® was established by the Kenya Bankers Association in 2012 to offer an array of research, commentary, and dialogue regarding critical policy matters that impact on financial markets in Kenya. The Centre sponsors original research, provides thoughtful commentary, and hosts dialogues and conferences involving scholars and practitioners on key financial market issues. Through these activities, the Centre acts as a platform for intellectual engagement and dialogue between financial market experts, the banking sector and the policy makers in Kenya. It therefore contributes to an informed discussion that influences critical financial market debates and policies.

The Kenya Bankers Association (KBA) *Working Papers Series* disseminates research findings of studies conducted by the KBA Centre for Research on Financial Markets and Policy. *The Working Papers* constitute "work in progress" and are published to stimulate discussion and contribute to the advancement of the banking industry's knowledge of matters of markets, economic outcomes and policy. Constructive feedback on the *Working Papers* is welcome. *The Working Papers* are published in the names of the author(s). Therefore their views do not necessarily represent those of the KBA.

Asset Quality Assessment in the Absence of Quality Data towards Optimal Credit Intermediation[1].

By Andrew Kioi Njeru

Abstract

The COVID-19 pandemic has posed a significant challenge for credit managers and risk management of financial institutions and regulators worldwide. The challenge faced arises because the consequences of outbreaks and epidemics are not distributed equally throughout, with some sectors of the economy suffering disproportionately. Businesses within the same sector are not affected the same way. This paper uses high-frequency transaction data and data obtained through web scraping to simulate firm's behaviour and performance during a crisis to estimate the sectoral impact of the pandemic and its pass-through to the portfolio of financial institutions and ultimately on economic growth. This proactive approach is critical due to the rapidly evolving nature of the crisis and delays by customers submission of the books of accounts and the impact of various measures such as lockdown and selected sector shutdowns undertaken by authorities that may have diverse implications for different businesses in various sectors of the economy thereby compromising he ability of risk managers to accurately forecast the performance of their portfolios.

1 I acknowledge the support and the very valuable forum afforded by the Kenya Bankers Association (KBA) in the form of the Annual Research Conference.We appreciate the input of the Director of Centre for Research on Financial Markets and Policy at KBA, Dr. Samuel Tiriongo and Josea Kiplangat.

1.0 Introduction

The COVID-19 pandemic has posed a significant challenge for credit managers and risk management for financial institutions and regulators worldwide due to the pace of its impact on economies and households and, by extension, the critical assets that financial institutions; the loans and advances. Asli Demirguc-Kunt et al. shows that the impact of the COVID-19 shock on banks was more pronounced than on all other corporates. The IMF notes that the economic risks of epidemics are not trivial. Fan, Jamison, and Summers recently estimated the expected yearly cost of pandemic influenza at roughly $500 billion,0.6 percent of global income, including both lost income and the intrinsic cost of elevated mortality. Even when the health impact of an outbreak is relatively limited, its economic consequences can quickly become magnified[2]. Bloom et al. note that the rapid decline in economic activity has a clear impact on the financial performance of businesses. This also impacts the households because of the job losses[3]. The consequences of outbreaks and epidemics are not distributed equally throughout the economy. Some sectors may even benefit financially, while others will suffer disproportionately., as they may have less access to health care and lower savings to protect against financial catastrophe[4]. COVID-19 resulted in an enormous shock to the corporate sector, which needed cash to cover operating costs from mitigation strategies (e.g., national quarantines, the shutdown of non-essential businesses). Banks are expected to supply much-needed funding (Acharya & Steffen, 2020). To do so, a wide range of financial sector interventions have been enacted. While these actions aim to increase the flow of credit in the short term, they may affect the banking sector's resilience in the longer term in terms of deterioration of asset quality.

2. Liberia, for example, saw GDP growth decline 8 percentage points from 2013 to 2014 during the recent Ebola outbreak in west Africa, even as the country's overall death rate fell over the same period.
3. In its annual Economic outlook for Subsaharan Africa released in March 2021, the international Monetary Fund (IMF), observed that Africa with limited purchasing power and few options, many countries will be struggling to simply vaccinate their essential frontline workers this year, and few will achieve widespread availability before 2023.
4. Pharmaceutical companies that produce vaccines, antibiotics, or other products needed for outbreak response are potential beneficiaries. Health and life insurance companies are likely to bear high costs, at least in the short term, as are livestock producers in the event of an outbreak linked to animals. Vulnerable populations, particularly the poor, are likely to suffer disproportionately.

The Central Bank regulatory capital framework places increased emphasis on risk management and banks are required to employ suitable procedures and systems to ensure their capital adequacy through Internal Capital Adequacy Assessment Process (ICAAP)[5]. This pandemic has posed a particularly daunting challenge to financial institutions because of its unprecedented impact on business activity and the rapid pace and impact of the measures taken to control including the lockdowns and a freeze of trading activities both at local and global level. One of the key aspects of the ICAAP process is the identification of material risks and this forms the basis of assessment of the adequacy of the capital held by the bank. The rapid pace and the depth of the adverse impact at which the pandemic is evolving pose a daunting challenge for risk managers of financial institution in terms of the ability to assess the quality of the loans and advances. The relief measures introduced by the government of Kenya at various times during the pandemic also imply that the conventional approaches may not reflect the full scale of the impact of the pandemic on the portfolio, these measures which include the

- Restructuring of the loans to give relief to customers
- Non reporting of defaulters to the bureaus
- Recent relief measures announced during the public holiday celebrations ,'mashujaa day' further underscore the need to enhance the risk assessment approaches.

This implies that the traditional approaches that focus on default as an indicator may not be appropriate for portfolio management may not be appropriate to assess the impact of the pandemic on the assets. Banks are expected by the regulator to operate a process that promotes accurate and timely risk identification and control, provides a complementary risk perspective to other risk management tools, improves capital management. Beyond the regulators requirements for capital adequacy, the policy makers would be interested in understanding not only the banking sector's resilience and the overall resilience of the business to craft the appropriate policy response to support the struggling businesses that form the backbone of the economy[6]. The main object of this study is to assess the appropriateness of using alternative approaches by using high frequency data , and web traffic data to assess the quality of the assets of a financial institution. Towards this objective we seek to:

1. Determine whether the high frequency cashflow data can be used as a predictor of the changes in asset quality.
2. Determine whether the web traffic data, alternative data sources including automation of cashflow tracking can be used to predict the changes in the asset quality.
3. Provide recommendations for risk managers with an objective methodology for the estimation of the changes in the asset quality based on the findings of the study.

5. These procedures are referred collectively as the Internal Capital Adequacy Assessment Process (ICAAP). ICAAP is the formal process through which a bank identifies, measures, aggregates, and monitors material risk, ultimately building a risk profile that becomes the basis for allocating capital.
6. A working paper released by metric stream dubbed organizations with agile integrated Risk management solution indicate that they responded better to pandemic-driven challenges; The paper suggest that agility, automation, and integrated approach have emerged as the key elements of an effective enterprise risk management (ERM) program

The rapid pace at which the pandemic has impacted the business environment has led the risk managers in financial institutions with little information to assess the quality of their assets. The extension of payment relief to borrowers by the regulators has also led to financial institutions having less information about the real behavior of their assets. This approach can help financial institutions to assess the extent of the impact of the pandemic on the borrowers cashflows and ability to sustain its operations both in the long ang short term and support banks process of determining the capital adequacy and plan its dividends and capital decisions. For the compliance officers, in banks it would aid the process of compliance with the Basel requirements in the internal capital adequacy assessment process as well as compliance with the new accounting standards regarding the estimation of impairment of loans and advances portfolio. The regulator could also use the approach to assess the quality of the assets held by financial institutions. This can aid in evaluating the short- and long-term impact of the pandemic on the stability of the financial system as a whole.

2.0 Background to the Banking Sector in Kenya

The Kenyan banking sector comprised of the Central Bank of Kenya (CBK), as the regulatory authority, 42 banking institutions. As at December 31, 2020, the Kenyan banking sector comprised of the Central Bank of Kenya (CBK), as the regulatory authority, 42 banking institutions (412 commercial banks and 1 mortgage finance company), 9 representative offices of foreign banks, 14 Microfinance Banks (MFBs), 3 Credit Reference Bureaus (CRBs), 17 Money Remittance Providers (MRPs), 8 non-operating bank holding companies, One Mortgage Refinance Company and 66 foreign exchange (forex) bureaus. Out of the 42 banking institutions, 40 were privately owned while the Kenya Government Includes Charterhouse Bank Limited and Chase Bank (K) Ltd which are In Liquidation and Imperial Bank Ltd, which is In Receivership. Government shareholding includes shares held by state corporations. had majority ownership in 2 institutions. Of the 40 privately owned banks, 23 were locally owned (the controlling shareholders are domiciled in Kenya) while 17 were foreign owned. The 23 locally owned institutions comprised 22 commercial banks and 1 mortgage finance company. Of the 17 foreign-owned institutions, all are commercial banks with 14 being local subsidiaries of foreign banks and 3 are branches of foreign banks. All licensed forex bureaus, microfinance banks, credit reference bureaus, money remittance providers, non-operating bank holding companies and mortgage refinance company were privately owned.

2.1 Government Response to the Pandemic

The government and central bank responded to the pandemic by putting up measures to support the businesses; The measures put up by the government include the reduction of the MSME turnover tax from 3 percent to 1 percent, as well as an increase in the cap for those liable to pay the turnover tax from Ksh.5 million to Ksh.50 million per year, reduction of Resident Corporate tax to 25 percent from 30 percent, reduction of Value Added Tax (VAT) from 16 percent to 14 percent. The Central Bank instituted measures to reduce the negative impact of the pandemic and these include the issuance of the guidance on restructuring of loans, personal ,business and MSMEs) which were performing by March 2, 2020, with the commercial banks bearing the cost of restructuring. This relief was to be granted on a case-to-case basis depending on an assessment of the circumstances facing the borrowers, six months suspension of listing of negative credit information for customers whose loans were previously performing and had become non-performing after April 1, 2020 and elimination of charges for transfers between the bank accounts and mobile wallets. The elimination of the bank transfer charges led to deepening of the adoption of mobile money transfers to an all-time high.

Other measures by the Government of Kenya includes the set up a Ksh.3 billion Credit Guarantee Scheme (CGS to cushion participating commercial banks, providing the commercial banks with confidence to extend loans to high-risk borrowers more efficiently and at flexible terms. The CGS is expected to support MSMEs' working capital, acquisition of assets and recovery from the COVID-19 pandemic impacts. These measures are key in measuring the risk of the participating MSMEs. The CGS was operationalized on October 13, 2020; as at June 2021, eleven commercial banks had signed agreements with the National Treasury to participate in the scheme, with more expected to join.

With more businesses and the general population embracing technology and the use of mobile phones for day-to-day activities, and a continuously growing demand for convenient financial services, fueled by the COVID-19 pandemic, the number of active mobile subscriptions has grown from 54.5 million in 2019 to 61.4 million in 2020. As a result, mobile penetration increased by 14 percent, from 114.8 percent to 129.1 percent. The number of active mobile money agents increased by 26 percent from 224,108 to 282,929. Consequently, this resulted in a 58 percent increase in the value of transactions from Ksh.382.9 billion in 2019 to Ksh.605.7 billion in 2020. The rapid pace of digitization offers the lenders a unique opportunity to enhance the risk management and portfolio evaluation approaches.

The measures to incentivize mobile money facilitated low value mobile money transactions and expanded usage of mobile money with higher limits. The monthly volume of person-to-person transactions increased by 87 percent between February and December 2020. Over the same period, the volume of transactions below Ksh.1,000 increased by 114 percent, while 2.8 million additional customers are using mobile money. Business related transactions also recorded significant growth over the same period. Ksh.32.62 billion (92.7 percent) of the Ksh.35.2 billion freed from the reduction in CRR had been disbursed by CBK to the banks as at December 31, 2020. Regarding the restructuring of the loans, a total of 401,498 loan accounts valued at Ksh.1.63 trillion were restructured in all economic sectors between March and December 2020 by all the Financial isntitutions. This accounted for 54.2 percent of banking sector gross loans of Ksh.3.0 trillion as at December 2020. The tenor of

202,373 Personal and household loans valued at Ksh.333. billion had been extended between March and December 2020. This accounted for 39.6 percent of the total Personal/Household loans as at December 2020 , Cumulatively, 199,125 loans valued at Ksh.1.29 trillion in the other ten sectors had been restructured between March and December 2020

2.2 Global impact of the pandemic

A paper authored by world bank analyses daily stock prices and other balance sheet information for a sample of banks in 53 countries to take a first look at this issue. Our contribution is twofold. It does an assessment of the impact of the pandemic on the banking sector and investigate whether the shock had a differential impact on banks versus corporates, as well as those banks with different characteristics. Second, using a global database of financial sector policy responses and an event study methodology, it investigates the role of different policy initiatives on addressing bank stress as perceived by markets, in the aggregate, as well as across different banks. Their results suggest that the adverse impact of the COVID-19 shock on banks was much more pronounced and long-lasting than on the corporates as well as other non-bank financial institutions, revealing the expectation that banks are to absorb at least part of the shock to the corporate sector. Furthermore, larger banks, public banks, and to some extent better capitalized banks suffered greater reductions in their stock returns, reflecting their greater anticipated role in dealing with the crisis. Banks with lower pre-crisis liquidity and oil sector exposure also suffered greater reduction in returns, consistent with their greater vulnerability to such a shock.

The International Monetary Fund, The IMF, projections released in 2021 show that sub-Saharan Africa is projected to grow by 3.4 percent in 2021 recovering from the worst on record contraction of 1.9 percent recorded in the year 2020. The that resulted in a large increase in poverty. In many countries, per capita income will not return to pre-crisis levels until 2025. The near-term outlook is subject to considerable uncertainty related to the course of the pandemic, access to vaccines, and the more challenging external financing environment. Economic and social dislocation caused by the crisis will further complicate policymaking. This undoubtedly affects purchasing power and consequently the ability for businesses to achieve their financial objectives and this has a direct impact on asset quality of financial institutions.

3.0 Literature Review

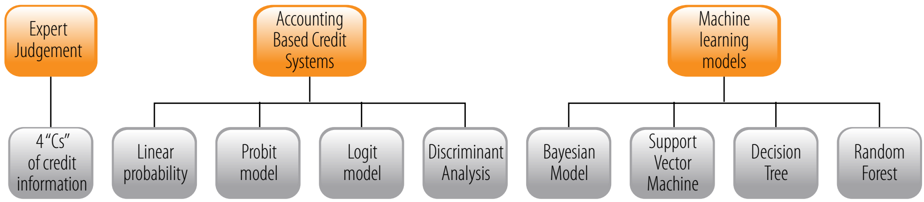
Credit scoring models have seen tremendous evolution over the years, partly due to the worldwide structural increase in the number of bankruptcies, the quest for more competitive margins on loans, and the shifts towards disintermediation by the highest quality and largest borrowers (Altman & Saunders, 1998). Initially, credit assessments by most financial institutions were pegged exclusively on subjective analysis (**Figure 1**). The 4 "Cs" of credit information[7] was essentially used on various borrowers to reach a determination on loans allocations (Altman and Saunders, 1998). Thereafter, accounting-based credit systems dominated risk assessments approaches (Altman, 1968; Altman, 1983; Ohlson, 1980; Zmijewski, 1984), and they rely on foundational assumptions and explicit models.

On the account of this approach, the financial institution decision-maker compares various key accounting ratios of potential borrowers with the industry or group benchmarks. The probit model is used during the assessment process when the assumption a standard normal distribution is made. In this regard, the cumulative distribution function in probit constitutes a transformation that puts the probability value in the [0, 1] interval while preserving the monotonic property (Ferenc, 2003). If the logistic distribution function is selected to express the probability of approval, it will lead to a logit model. In this case, the probability of a default is related to a set of potential explanatory variables, with less consideration being made to normality assumptions of the variables as in discriminant analysis (Liu and Schumann, 2005).

Discriminant analysis models assume that there are two populations of individuals, which are denoted '1' and '0', each of which is characterized by a multivariate normal distribution of the attributes. The analysis of the latter yields a linear function, akin to a regression equation, which can then be used to create an overall score (Altman, 1968 and Altman, et al. 1977).

7. The 4 "Cs" refers to the borrower character (reputation), capital (leverage), capacity (volatility of earnings) and collateral.

Figure 1: Credit Scoring Models



Source: Author

More recently, machine learning models have gained traction in building predictive models (Partalas & Tsoumakas, 2010 and Shrawan & Trivedi, 2018), as its being adopted by financial institutions for credit scoring (Wongnaa & Babu, 2020). Among these models is the Bayesian approach, which is used to predict class of membership samples (Friedman, et al. 1997), and the most used models are the Naive Bayes and Bayesian belief (Lewis, 1998). Support vector machine (SVM) optimally separates observations into two sets of data points. Vapnik (1995) and Chapelle & Vapnik (2000) argue that the latter is the best classifier developed for Pattern Classification, since it is mostly used in small samples and it does not limit distribution of data. Moreover, given that the model is anchored on structural risk, it also attains good robustness. The intuitive idea behind the SVM algorithm is to maximize the probability of making a correct prediction by determining the boundary that is the furthest away from all the observations. However, since real-world data sets are not perfectly linearly separable, the SVM algorithm is embedded with an adaptation called soft margin classification, which adds a penalty to the objective function for observations in the training set that are misclassified (DeRose and Lannou. 2020).

Decision tree classifier divides the dataset based on the attributes taken from training data. The idea of entropy is to train the data to develop the decision tree (DeRose and Lannou. 2020). The later neglects values that are missing and uses smart pruning to reduce size of the tree. At the initial stage, the latter grows the tree which over fits the data, and further, by using pruning method, it removes nodes and branches which are inefficient, and over fits the model on training (Rajeswari et al., 2017). Lastly, the Random Forest, which is an ensemble of classifiers approach (Pal, 2005; Trivedi & Dey, 2013 and Tripathi, et al., 2015), combines decisions of several decision dump classifiers to generate a suitable result. Bagging ensemble approach modifies samples of data and randomly selects feature input; further, each sample is taken to train the decision dump classifier. The aggregate result is computed by voting of all the classifiers taken for ensemble.

The traditional default prediction models bear a number of challenges. First, financial ratios based on accounting information are often difficult to verify, and it can be either incorrect or nonexistent for smaller firms. Moreover, these ratios report past business performance rather than future performance (Trujillo-Ponce et al., 2014). Secondly, less information is available to the banks to evaluate the borrowers risk profile (Hirata, 2005 and Klinger, et al., 2013). Thirdly, the models does not necessarily guarantee more efficiency since it often requires an additional process, such as in-person interviews, due to the imperfect information (Nemoto et al, 2018).

Lastly, the scoring model is prone to sample selection bias since it only covers the information of individuals and firms that applied for bank loans (Hirata, 2005; Chen, 2009 and Chen, 2012). This implies that the models won't be fine-tuned due to the restrictions deriving from existing statistical instruments (Tseng and Hu, 2010). Credit risk assessment based on machine learning is superior to the traditional approach on the account of its ability to reduce both computational times and biases regarding which parameters have to be considered to improve models (Fosso Wamba et al. 2015; Bukovina 2016 and Alaka et al. 2018).

4.0 Econometric Methodology

The credit scoring methodologies for scoring have been in use since the first model published by Altman in 1968. The first model used financial statements data to model default prediction scorecards. Since then, new models have developed, and these utilize other data sources including bank account data and other data sources including alternative data. Logistic regression is a common technique used to develop scorecards. We model multiple logistic regression to predict a binary outcome good/bad. Logistic regression uses a set of predictor characteristics to predict the likelihood of a specific outcome (the target). The equation for the logit transformation of a probability of an event is show as follows;

$$\text{Logit}\ (p_i) = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_k x_k$$

Where

p = posterior probability of "event," given inputs

x = input variables

β_0 = intercept of the regression line

β_k = parameters

Logit transformation is log of the odds, that is, log (p[event]/ p[nonevent]), and is used to linearize posterior probability and limit outcome of estimated probabilities in the model to between 0 and 1. Maximum likelihood is used to estimate parameters β_1 to β_k. These parameter estimates measure the rate of change of logit for one unit change in the input variable (adjusted for other inputs); that is, they are in fact the slopes of the regression line between the target and their respective input variables x_1 to x_k. The parameters are dependent on the unit of the input (e.g., a percentage number compared to total cash inflows into a customers account. and need to be standardized to ease analysis.

The strength of a can be measured many different ways. There is a long list of statistics that measure how well the model describes the data or how well it predicts the cases of goods and bads. Some of these include the usage of KS, chi-square, Hosmer-Lemeshow,15 AUC, AUCM Gini and the confusion matrix. The results of the modeling need to be validated. This is standard "out-of-sample" validation, where we confirm that the model developed is generally applicable to the subject population, and to ensure that the model has not been over-fitted. The industry norm is to use a random 70 percent or 80 percent of the development sample for building the model, while the remaining 30 percent or 20 percent "holdout sample" is kept for validation.

4.1 Data and measurement issues.

We review the daily transactions data forma sample of 294 MSMEs and assess the behavior using a logit model. The data used was collected over a period from October 2019 to October 2020. The conventional methodology to assessing the evolution of and changes to the portfolio behaviors is to observe the delinquency status in terms of the individual loans. This approach utilizes only one data point, the number of days that a loan is past due. The most used approach is to assess the quantum of the loans falling due by over 90 days. The methodology is now very defective since the covid19 pandemic started due to various reasons and chief among them are as follows;

1. The covid19 relief measures introduced by the government that allow banks to restructure loans make it harder to track and develop models using the days past due because the restructures wipe out the delinquency records.
2. The measures that lead to non-submission of customers with default history to the credit reference bureaus also imply that the data for non-performing debts is not visible to lenders for modelling and early warning signs
3. The wide adoption of digital banking meant that customers banking behavior shifted hence the need to adopt alternative methodology to collect data.

The other portfolio monitoring approaches that rely on the data submitted by the customers such as management accounts and audited accounts also failed due to unavailability of the financial statements. Most businesses were unable to provide the records at per expected timeframes and when submitted, the data was woefully unreliable due to the rapid pace of the changes in the economy. We adopt the use of high frequency data to assess the performance of the portfolio. The data used includes the following:

- The daily account deposit and withdrawal behavior
- The number of days an account remains in credit and funded within the month
- The number of days an account remains in debit and unfunded
- The number of days an account remains at zero balance
- The count of transactions
- Total exposure

We relaxed the approach towards defining a defaulter to use the restructured loans even where the customer may not have delinquency status. This analysis a sample of MSMEs only, and we mixed the sample to accommodate a high proportion of defaulters to facilitate the feature selection. 40.1% of the sample had defaulted. A paper by JPMorgan and a study by Bartolini et al(2009)notes that 'Cash is king'; cashflow behavior is a significant factor in determining whether or not a business succeeds or fails, it is a critical factor for businesses to get right. Even profitable businesses have gone bankrupt because of poor cash management. A business in distress tends to have their bank account operate in a manner that banks describe as hardcore. This implies that there is little movement in bank account balances from day to day. This is also implies that an account tend to be at zero balances often or overdrawn. The number of transactions also tend to be fewer as borrowers face declining business activity. The same can be postulated for the web traffic data, lower traffic and declining traffic imply lower business activity. The results of an annual survey of top senior finance executives, Chief Finance Officers, describe some other key concerns for the businesses. Cash flow management is the most cited key area of concern, tabbed by 41% of the respondents—not surprising, given the uncertainty caused by the economic downturn. Forecast accuracy is next on the list, at 31%—another expected leading concern because of the once-in-a-lifetime type of economic pressures exerted by the pandemic. Rounding out the top four key concerns are identifying new technologies to benefit the finance team and managing data security, at 24% and 22%, respectively.

Table 1: Description and measurement of variables

Feature	Description	Details
CountTxnx	Count of transactions that a customer makes. This is for both deposits and withdrawals across all channels. Checks, swift, mobile remittances	Consistent level of transactions indicate that the customer is active and les likely to default
Dayszero	Count of days that an account remains at zero balance	This is expected to be a strong predictor of default. A customer who is running normal operation is not expected to have this kind of account behaviour of having nil account balance.
daysdebit	Count of days that an account remains at is overdrawn	This is expected to be a strong predictor of default. A customer who is running normal operation is not expected to have this kind of account behaviour where the account remains with debit or overdrawn balances.
dayscredit	Count of days that an account remains at positive credit balance	This is expected to be a strong predictor of default. A customer who is running normal operation is not expected to have this kind of account behaviour

Feature	Description	Details
sumdebit	Total outflows from an account	Higher activity reflects a favorable outcome
sumcredit	Total inflows into an account	Higher activity reflects a favorable outcome
totalbal	Total Exposure	
exposuretoin-flows	Ratio of total exposure to cash inflows	Higher activity reflects a favorable outcome and ability to meet obligations as they fall due
default	The target feature based on request to restructure or days overdue exceeding 30 or both	

The summary statistics are displayed below.

Table 2: Description and measurement of variables

Feature	Min	1st Quartile	Median	Mean	3rd Quartile	Max
CountTxnx	0	43	145	1057.8	676.5	171395
Countrecords	2	572	598	545.7	598	1794
Dayszero	0	0	3	9.378	20	27
daysdebit	0	0	1	8.823	25	27
dayscredit	0	0	8	10.46	23	27
sumdebit	7.36E+10	-5.32E+07	-1.59E+07	-1.90E+08	-5.21E+06	0.00E+00
sumcredit	0.00E+00	4.90E+06	1.57E+07	1.91E+08	5.35E+07	8.05E+10
totalbal	1.00E+06	2.15E+06	3.94E+06	4.02E+07	8.52E+06	1.02E+10
exposuretoinflows	0	0	0.00429	0.1763	0.05	45.57
default	0	0	0	0.4022	1	1

4.2 Empirical estimation

We test the model on a sample of firms that defaulted and a sample of firms that did not default with a default definition being thirty or more days past due. A logit model was adopted. Lasso regression techniques was adopted to reduce model complexity and prevent over-fitting which may result from the regression. The following logit model was used in R studio, a data analytics tool[8].

8. The estimation being anchored on the following code in R-Studion:- mydata$rank <- factor(mydata$rank), mylogit <- glm(defau ~ CountTxnx+ Dayszero+ daysdebit+ dayscredit+ sumdebit+ sumcredit+ totalbal+exsurestoinflows, data = mydata, family = "binomial"), and summary(mylogit)

Table 3: Logistic Regression Results

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)       -2.869e+00  1.635e-01 -17.546  < 2e-16 ***
CountTxnx          3.893e-06  7.381e-06   0.527 0.597915
Dayszero           1.358e-01  5.619e-03  24.172  < 2e-16 ***
daysdebit          9.032e-02  5.435e-03  16.617  < 2e-16 ***
dayscredit         2.673e-02  7.254e-03   3.684 0.000229 ***
sumdebit           1.413e-09  4.622e-10   3.057 0.002239 **
sumcredit          1.316e-09  4.593e-10   2.866 0.004155 **
totalbal           2.114e-09  3.837e-10   5.508 3.63e-08 ***
exposuretoinflows -2.005e-01  6.646e-02  -3.017 0.002553 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10441  on 7746  degrees of freedom
Residual deviance:  7206  on 7738  degrees of freedom
```

The findings were that all features were significant save for the count of transactions. The low significance on this feature is attributed to the noise created by system generated entries of inflow and outflow for defaulting firms as the system recovers all available funds to service a loan. The AUC was quite high given the low number of features used. The gini coefficient retuned a figure of 71.2. This is extremely high by all standards ang d given the low count of the features used. The confusion matrix shows that the model was able to classify 85.7% of the defaulters accurately. Moreover, the ROC curve is very predictive as shown in **Figure 2** where the green line is far from the diagonal line. We also experimented with the web scraping to gather online statistic, and the results were not conclusive nor significant. The duration of web scraping was the major drawback, the need to track the traffic before covid and during would make the wen crawling data more useful.

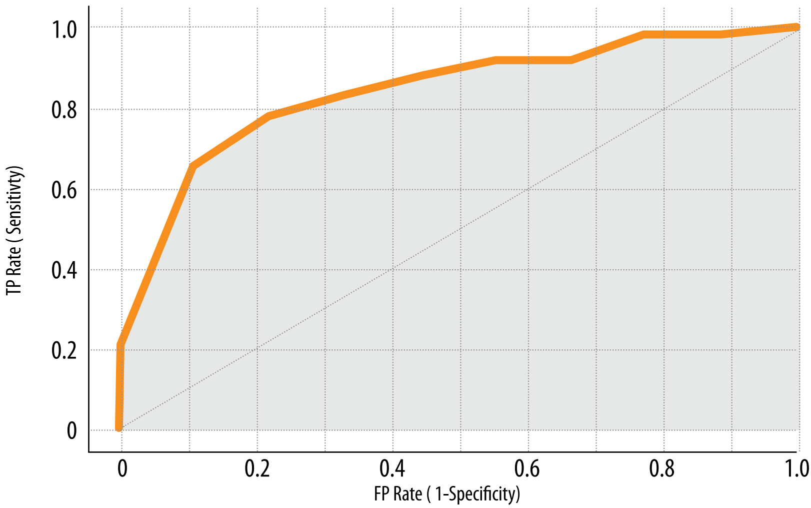
Table 4: Logistic Regression Model Predictive Power

Scores Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.856	0.790	0.790	0.790	0.790

Table 5: Confusion matrix for Logistic Regression (showing proportion of actual)

Actual \ Predicted	0	1	∑
0	87.8 %	12.2 %	178
1	32.3 %	67.7 %	116
∑	194	100	294

Figure 2: ROC curve



5.0 Conclusions and Policy implications

The use of high frequency account data can be adopted as an alternative when data is scarce and not reliable given the measures implemented to support borrowers. The has proven to have a higher predictive power than conventional approaches in the absence of default indicators. The use of external data from telco and mobile service providers which is similar the account activity data can substitute the bank's high frequency data and is also and similar to bank account data powerful and can be automated to help track the performance of customers on near real time basis. Further research is need to collect alternative data on a continuous basis to allow higher quality data over a long duration and this can be explored to help with risk management processes. The adoption of data that cold indicate the transaction behavior is has immense predictive power in the absence of good quality data from Credit reference bureaus and when default indicators are affected like in this case where covid relief measures have affected the quality and timeliness of data.

References:

1. Alaka, H., Oyedele, L., Owolabi, H., Akinade, O., Bilal, M., & Ajayi, S. (2018). *"A Big Data analytics approach for construction firms failure prediction models".* IEEE Transactions on Engineering Management.

2. Altman, E. (1968). *"Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy."* Journal of Finance, vol. 23: 589–609.

3. Altman, E. and Saunders, A. (1998). *"Credit risk measurement: Developments over the last 20 years".* Journal of Banking & Finance 21, pp. 1721 -1742. Salomon Brothers Center, Leonard Stern School of Business, New York University.

4. Altman, E. I. (1983). *"Corporate financial distress: a complete guide to predicting, avoiding, and dealing with bankruptcy".* John Wiley and Sons.

5. Altman, E., Halderman, R., and Narayanan, P. (1977). *"Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations."* Journal of Banking & Finance, vol. 1: 29–54.

6. Asli Demirguc-Kunt* Alvaro Pedraza† Claudia Ruiz-Ortega Deming, D., & Dynarski, S. (2021). *Banking Sector Performance During the COVID-19 Crisis* (IMF Working Paper No. 9363). International Monetary Fund https://documents1.worldbank.org/curated/en/209481597431651590/pdf/Banking-Sector-Performance-During-the-COVID-19-Crisis.pdf

7. Bukovina, J. (2016). *Social media big data and capital markets—An overview.* Journal of Behavioral and Experimental Finance, Vol 11, pp. 18–26.

8. Central Bank Of Kenya (2020) *Implementation of Emergency measures to mitigate the adverse impact of covid19 pandemic on loans and advances.* Retrieved from https://www.centralbank.go.ke/uploads/

banking_circulars/1400484618_Banking%20Circular%20No.%203%20of%202020%20-%20Implementation%20of%20Emergency%20Measures.pdf

9. Central Bank Of Kenya (2021) *Mobile Payments* Retrieved from https://www.centralbank.go.ke/national-payments-system/mobile-payments/

10. Central Bank of Kenya; 2020: *Emergency measures to mitigate the adverse economic effects on bank borrowers from the Coronavirus pandemic.* https://www.centralbank.go.ke/uploads/banking_circulars/1400484618_Banking%20Circular%20No.%203%20of%202020%20-%20Implementation%20of%20Emergency%20Measures.pdf

11. CFOs' *Appetite For Finance Technology Undiminished* retrieved from https://www.cfo.com/pressreleases/new-study-reveals-cash-is-still-king-amid-pandemic-fueling-investment-in-ap-and-ar-automation/

12. Chapelle, O. and Vapnik, V. (2000). *"Model selection for support vector machines. In Advances in neural information processing systems."* pp. 230–236.

13. Chen, H. L. (2009). *"Model for predicting financial performance of development and construction corporations".* Journal of Construction Engineering and Management.

14. Chen, J. H. (2012). *"Developing SFNN models to predict financial distress of construction companies".* Expert Systems with Applications.

15. Chrouhy,M ,Galai D and Mark R.(2009). *The Essentials of Risk Management.* New York. McGraw-Hill

16. David E. Bloom, Daniel Cadarette, and JP Sevilla; *2020; New and resurgent infectious diseases can have far-reaching economic repercussions Retrieved from* https://www.imf.org/external/pubs/ft/fandd/2018/06/economic-risks-and-impacts-of-epidemics/bloom.htm

17. DeRose, K. and Lannou, C. (2020). *"Machine Learning".* New York University, Stern School of Business (USA).

18. Ferenc, K. (2003). *"Credit scoring process from a knowledge management perspective".* Periodical polytechnica. VOL. 11, NO. 1, PP. 95–110, 2003.

19. Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). *"How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study".* International Journal of Production Economics, Vol 165, pp. 234–246.

20. Friedman, N., Geiger, D. and Goldszmidt, M. (1997). *"Bayesian network classifiers". Machine* Learning. Springer Vol. 29 pp 131–163.

21. Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics.* Boston, Mass: McGraw-Hill

22. Hirata, H. (2005). *"Recent Development in Japan's Small and Medium-sized Enterprise Finance"*. Journal of Business, 41(2): 31–51.

23. https://www.the-star.co.ke/business/kenya/2021-08-12-kenyans-transact-sh18-billion-daily-via-mobile-money-cbk/

24. International Monetary fund (April 2021) *Sub-Saharan Africa: Navigating a Long Pandemic retrieved from* https://www.imf.org/en/News/Articles/2021/04/14/pr21108-sub-saharan-africa-navigating-a-long-pandemic

25. JPMorgan Chase BankCash is King(2016). Flows, Balances and Buffer Days Evidence from 600,000 Small Businesses retrieved from *https://www.chase.com/content/dam/chasecom/ en/newsroom/multimedia/112316-report-cash-is-king.pdf*

26. Klinger, B., Castro, L., Szenkman, P., and Khwaja, A. (2013). *"Unlocking SME Finance in Argentina with Psychometrics".* Inter-American Development Bank. IDB Technical Note 532

27. Lewis, D. (1998). *"Naive (Bayes) at forty: the independence assumption in information retrieval".* European Conference on Machine Learning, Springer, Berlin, Heidelberg, 1998, pp. 4–15.

28. Liu, Y. and Schumann, M. (2005). *"Data mining feature selection for credit scoring Models"*. Journal of the Operational Research Society 56, pp 1099–1108.

29. Nemoto, N., Yoshino, N., Okubo, Y., Inaba, D. and Yanagisawa, K. (2018). *"Credit risk reduction effect on small and medium-sized enterprise finance through the use of bank account information".* Asian Development Bank Institute. ADBI Working Paper Series**.**

30. Ohlson, J. A. (1980). *"Financial ratios and the probabilistic prediction of bankruptcy".* Journal of Accounting Research, 18(1), 109–131.

31. Pal, M. (2005). *"Random forest classifier for remote sensing classification".* International Journal of Remote Sensing. Vol 26 (1), pp. 217–222.

32. Partalas, G. Tsoumakas, I. (2010). *"An ensemble uncertainty aware measure for directed hill climbing ensemble pruning".* Machine Learning. Springer Vol 81. PP. 257–282.

33. Siddiqi, N. (2006). *Intelligent Credit Scoring: Building and implementing intelligent credit scoring.* Hoboken, N.J: Wiley.

34. The guardian:2021: *Most poor nations will take until 2024 to achieve mass Covid-19 immunization* https://www.theguardian.com/society/2021/jan/27/most-poor-nations-will-take-until-2024-to-achieve-mass-covid-19-immunisation

35. The President of Kenya(2021); speech by his excellency retrieved from https://www.president.go.ke/2021/10/20/speech-by-his-excellency/

36. Tripathi, A., Sharma, R. D., & Trivedi, S. K. (2015). *"Identification of plant species using supervised machine learning".* International Journal of Computing. pp. 975.

37. Trivedi, S.K., Dey, S. (2013). *"An enhanced genetic programming approach for detecting unsolicited emails".* IEEE 16th International Conference on Computational Science and Engineering, IEEE, pp. 1153–1160.

38. Trivedi, S.K., Panigrahi, P.K. (2018). *"Spam classification: a comparative analysis of different boosted decision tree approaches"*. Journal of system and Information Technology.

39. Trujillo-Ponce, A., Samaniego-Medina, R., & Cardone-Riportella, C. (2014). *"Examining what best explains corporate credit risk: accounting-based versus market-based models".* Journal of Business Economics and Management, 15, 25– 176.

40. Tseng, F. M., & Hu, Y. C. (2010). *"Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks".* Expert Systems with Applications.

41. Vapnik, V. (1995). *"Support-vector networks".* Machine Learning. Springer. Vol 20 pp. 273–297.

42. Wongnaa, C.A. and Babu, S. (2020). *"Building resilience to shocks of climate change in Ghana's cocoa production and its effect on productivity and incomes".* Technology in Society. Vol 62

43. Zmijewski, M. E. (1984). *"Methodological issues related to the estimation of financial distress prediction models"*. Journal of Accounting Research, Vol 22, pp. 59–82.

Kenya Bankers Association
13th Floor, International House, Mama Ngina Street
P.O. Box 73100– 00200 NAIROBI
Telephone: 254 20 2221704/2217757/2224014/5
Cell: 0733 812770/0711 562910
Fax: 254 20 2221792
Email: research@kba.co.ke
Website: www.kba.co.ke

KENYA BANKERS
ASSOCIATION

One Industry. Transforming Kenya.