Interlinkages in the Malaysian Banking System
Principal Component Analysis on Sector-Wide Macrofinancial Indicators

Pulapa Sai Prasanth

11/04/2021

Abstract

This report documents a principal component analysis exercise to identify a cluster of indicators with potential predictive power to forecast deterioration in a Malaysia banking system's portfolio or costs, developed to detect potential warning signs in the wider banking system.

Introduction, Motivation and Data

This exercise is done based on the need to identify latent factors that will be used to forecast deterioration in Malaysia banking system's portfolio or costs, using measurements such as:

- The banking system's nonperforming loans (NPL) ratio;
- The banking system's cost of funds, relative to rates such as the Kuala Lumpur Interbank Offered Rate (KLIBOR) or Bank Negara Malaysia's (BNM's) Overnight Policy Rate (OPR);
- The banking systems's credit default swap (CDS) spreads, in terms of both absolute and relative to the sovereign CDS or an asset-weighted average of the CDS spread of competitor banks.

In this exercise, a principal component analysis (PCA) is done on economy-wide macrofinancial indicators to identify potential sector-wide linkages between these variables. The dataset used in this execise was sourced from BNM, which can be sourced from http://www.bnm.gov.my/index.php?ch=statistic and transformed to produce the following variables:

- Ratios to total deposits:

    - Demand deposits ratio (`dd.deposits.r`)
    - Foreign currency deposits ratio (`fx.deposits.r`)
    - Repurchase agreements ratio (`repo.deposits.r`)

- Ratios to total loan applied:

    - Passenger car loan application ratio (`loan.app.cars.r`)
    - Construction loan application ratio (`loan.app.construction.r`)
    - Non-residential property loan application ratio (`loan.app.nonresprop.r`)
    - Residential property loan application ratio (`loan.app.resprop.r`)
    - Working capital loan application ratio (`loan.app.workingcapital.r`)

- Total loans applied growth rate (`loan.yy`)
- Liquidity capital ratio (`lcr`)

PCA is useful in this endeavour as it involves reducing data dimensionality to provide a subspace that best represents the data in the sense of maximising the variability of the projected points.

Analysis Part I: Imputing Missing Values in the Dataset

In order to proceed with PCA, the dataset would first need to be "balanced" - i.e. there needs to be an equal amount of entries for each of the variable in the dataset. Dataset used here is inherently "unbalanced" due to the different starting points to which the data was first reported. As a result, there are clusters of variables with different number of observations; `npl.r`, which has data as far back as January 1997, has 270 observations, whereas `lcr`, which BNM only began reporting in June 2015, only has 40 recorded observations. If PCA is done on this raw dataset, the size of dataset included in the analysis will be constrained to the size of the variable with the least amount of entries and the entries which all variables concurrently recorded an observation. To illustrate this, the following is a graph of the available NPL ratio data (which BNM only began reporting in December 2008), which has recorded observations less than half of the variable with the longest reporting time frame in the dataset:
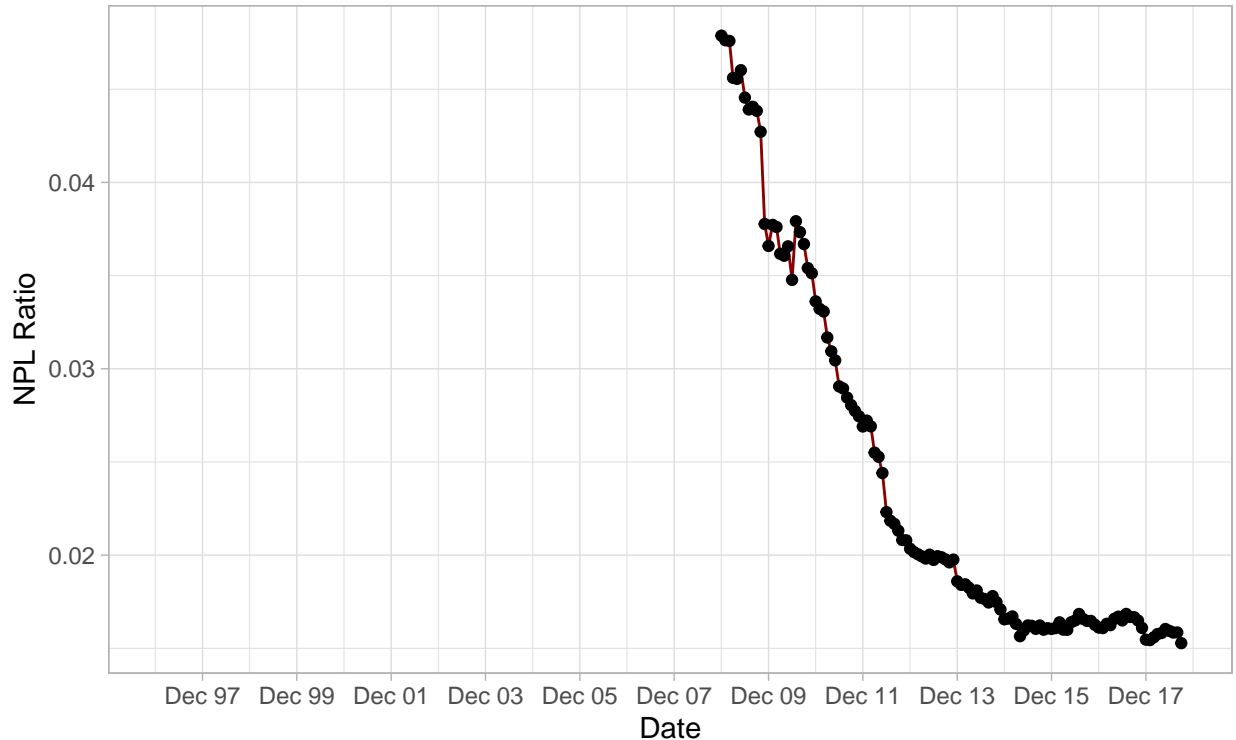

Figure 1: NPL ratio, as reported by BNM

To overcome this problem, a method formulated in Josse and Husson (2012) provided an algorithm, called the regularised iterative PCA algorithm, to impute the missing values using existing principal axes and components in the dataset whilst simultaneously overcoming the issue of overfitting. The R Package `missMDA` developed in Josse and Husson (2016) is used to perform principal component methods on incomplete data, aiming at estimating parameters and obtaining graphical representations despite mising values. Using these algorithms yields the following graph of imputed NPL ratio for entries before December 2008 and maintaining the original values December 2008 onwards:
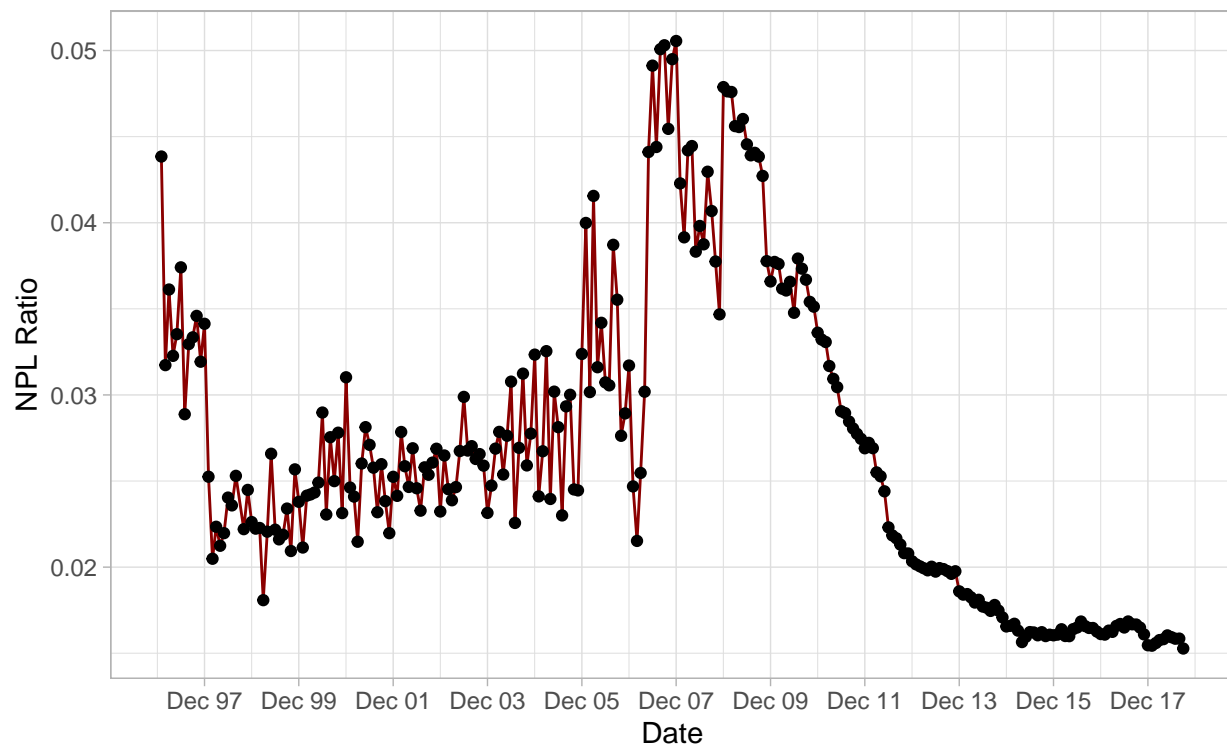
Figure 2: NPL ratio with imputed values for observations before December 2008

Note that in Figure 2, the algorithm imputed high NPL ratios in mid-1997 – reflecting onset of the Asian Financial Crisis – without explicit specification in the model. The imputed values does, however, affect the inter-variable relationships, by either strengthening existing correlations in the original dataset (see Figures 3 and 4). The full summary statistics of the datasets before and after the transformation are tabled in the Appendix section.
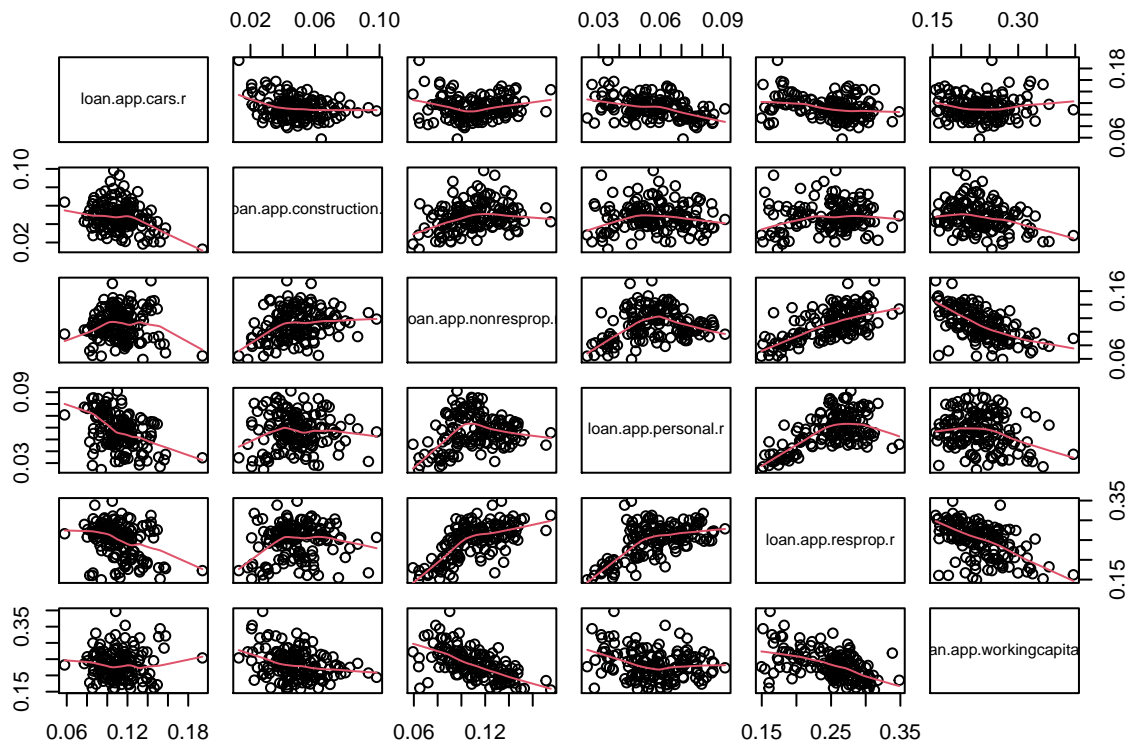
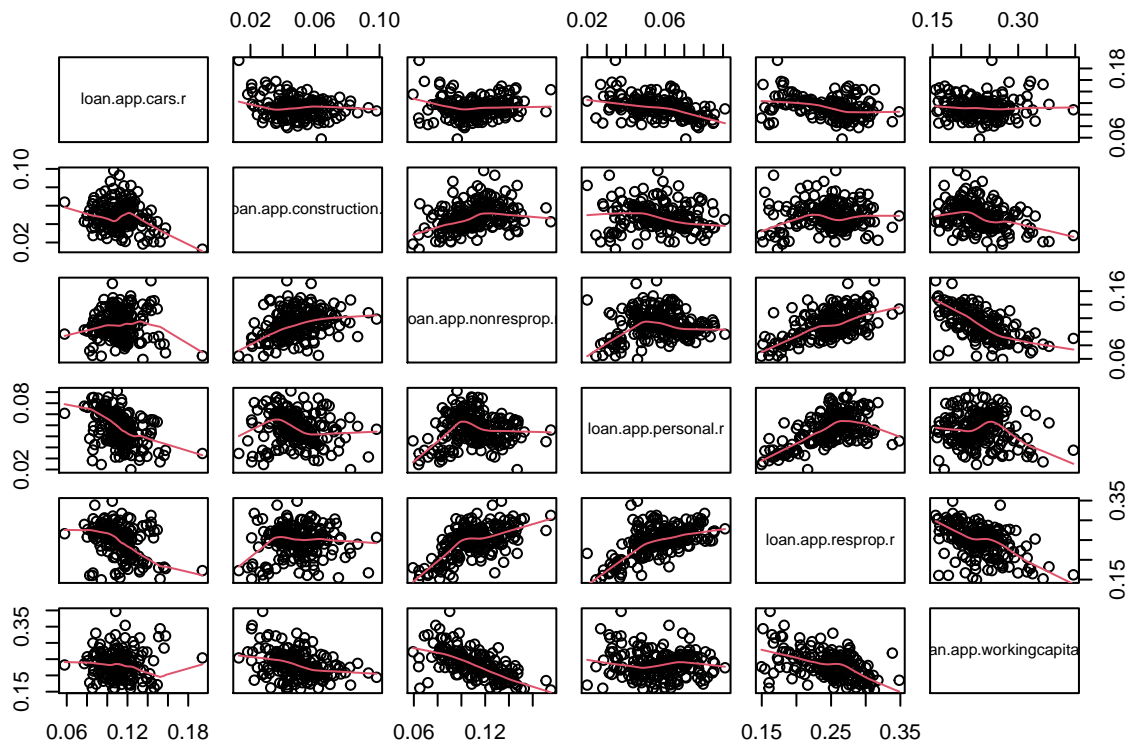Figure 3: Relationship between loan application ratios in the raw dataset

Figure 4: Relationship between loan application ratios in the imputed dataset

Analysis Part II: Running the PCA

How many principal components (PCs) to retain? Eigendecomposition – computing eigenvalues

The core idea behind PCA is to reduce the dimensionality, i.e. reduce the number of variables, of a dataset while retaining as much as possible of the variation present in the data. When running the PCA on the new transformed dataset with imputed and original values, the analysis yields eigenvalues, i.e. a vector of values that provide information about the amount of variability captured by each principal component (PC). Each eigenvalue covers a proportion of variation that exists in the dataset. The following shows a table and screeplot (Figure 5) of eigenvalues of each PC (listed as "comp") from the PCA:

Table 1: Table of eigenvalues

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	4.034	31.033	31.033
comp 2	3.450	26.536	57.569
comp 3	1.256	9.660	67.230
comp 4	1.104	8.494	75.724
comp 5	0.918	7.061	82.785
comp 6	0.586	4.507	87.292
comp 7	0.456	3.510	90.803
comp 8	0.380	2.922	93.725
comp 9	0.306	2.353	96.078
comp 10	0.173	1.328	97.406
comp 11	0.159	1.222	98.628
comp 12	0.136	1.044	99.673
comp 13	0.043	0.327	100.000

Since the exercise is to determine the number of dimensions to reduce, the amount of variance each PC covers is a useful metric to decide which PCs to retain. There is no universal criterion to determine the number of PCs to retain. From Table 1, the percentage of variance captured by the first two dimensions covers over half (57.6%) of the entire variation in the dataset. The percentage variance captured by the first three dimensions covers roughly two-thirds of the entire variation in the dataset (67.2%). In this exercise, the first two PCs will be given the most focus, while the third PC is included intermittently for comparative and illustrative purposes.
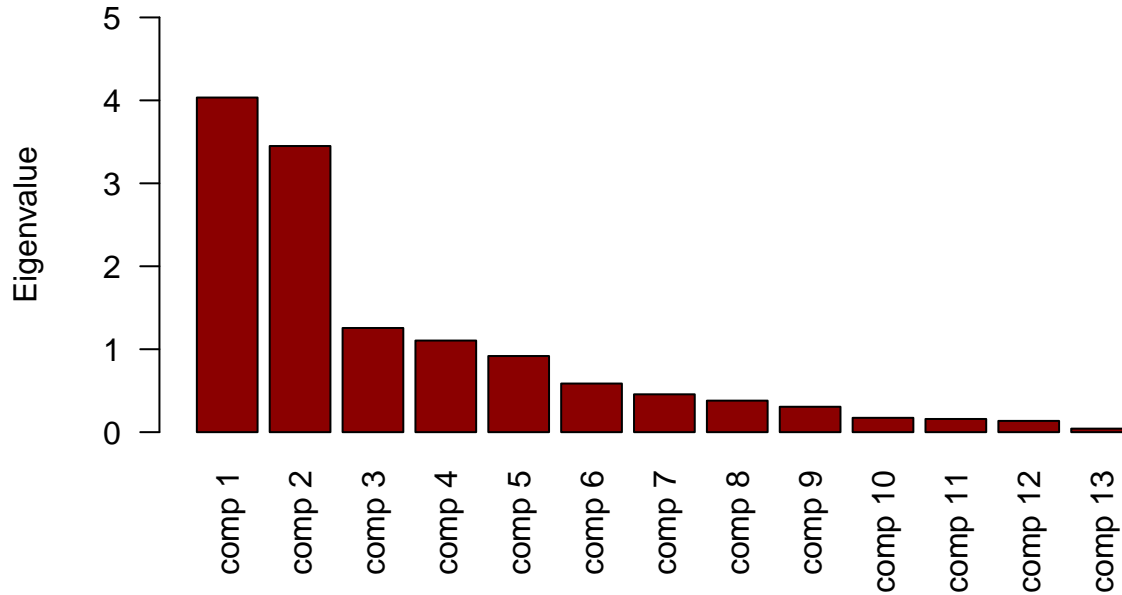
Figure 5: Screeplot of eigenvalues

Which variables characterise each PCs?

To see how each PC is characterised, the correlations between the variables and the PCs (listed in the Table 4 as "Dim") are evaluated:

Table 2: Correlations between variables and PCs

	Dim.1	Dim.2
dd.deposits.r	0.212	0.798
fx.deposits.r	0.656	0.674
repo.deposits.r	-0.685	0.382
loan.app.cars.r	-0.492	-0.074
loan.app.construction.r	0.111	-0.616
loan.app.nonresprop.r	0.507	-0.660
loan.app.personal.r	0.782	0.472
loan.app.resprop.r	0.885	-0.070
loan.app.workingcapital.r	-0.465	0.591
loan.yy	-0.138	-0.811
nfa.yy	-0.124	0.042
depo.yy	-0.388	-0.289
lcr	0.884	-0.265

Taking residential property loan applications ratio (`loan.app.resprop.r`) as an example, notice that PC1 has a strong positive correlation with `loan.app.resprop.r` whilst registering an insignificant, negative correlation with PC2. These statistics can be plotted as a radar or circle of correlations (see Figure 6). The closer an arrow is to the circumference of the circle, the better its representation on the given axes. Also note how the variables are grouped.
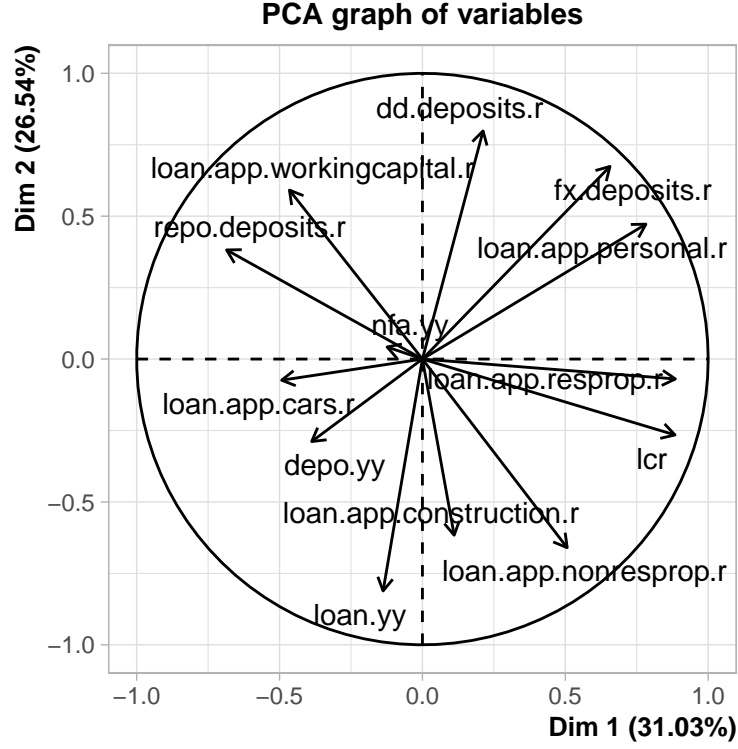
7

PCA graph of variables

Figure 6: Variables factor map

Influence of variables in each PC

Another perspective to evaluate how variables characterise each PCs is to examine the contributions of each variable to each PCs. The following table lists the proportion of each variable that make up a single PC (listed in Table 5 as "Dim"):

Table 3: Contributions of variables on each PC

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
dd.deposits.r	1.116	18.473	12.197	0.009	5.317
fx.deposits.r	10.680	13.169	0.656	0.705	1.870
repo.deposits.r	11.639	4.227	11.298	0.111	0.617
loan.app.cars.r	6.012	0.159	31.210	16.520	3.072
loan.app.construction.r	0.306	11.011	0.478	10.397	27.209
loan.app.nonresprop.r	6.359	12.632	8.668	0.834	1.077
loan.app.personal.r	15.163	6.458	0.134	1.096	0.041
loan.app.resprop.r	19.425	0.140	1.483	0.600	1.497
loan.app.workingcapital.r	5.365	10.134	13.370	0.271	1.663
loan.yy	0.470	19.087	1.974	0.597	0.124
nfa.yy	0.379	0.052	2.397	56.548	33.436
depo.yy	3.724	2.415	13.909	12.058	23.131
lcr	19.361	2.043	2.226	0.255	0.947
TOTAL	100.000	100.000	100.000	100.000	100.000

Results

Plotting PC scores

From the PCA exercise, the PC scores derived from each observation can be used as coordinates to plot the objects in a scatterplot. In Figure 7, each PC score is also applied a colour gradient to visualise the impact of moving towards certain regions in the PC1-PC2 plot on NPL ratios. Here we can observe that moving into the negative region of PC1 generally implies higher NPL ratios. This gives a the PC1 axis a strong predictive power on the general outlook of banks' loan portfolio performance.
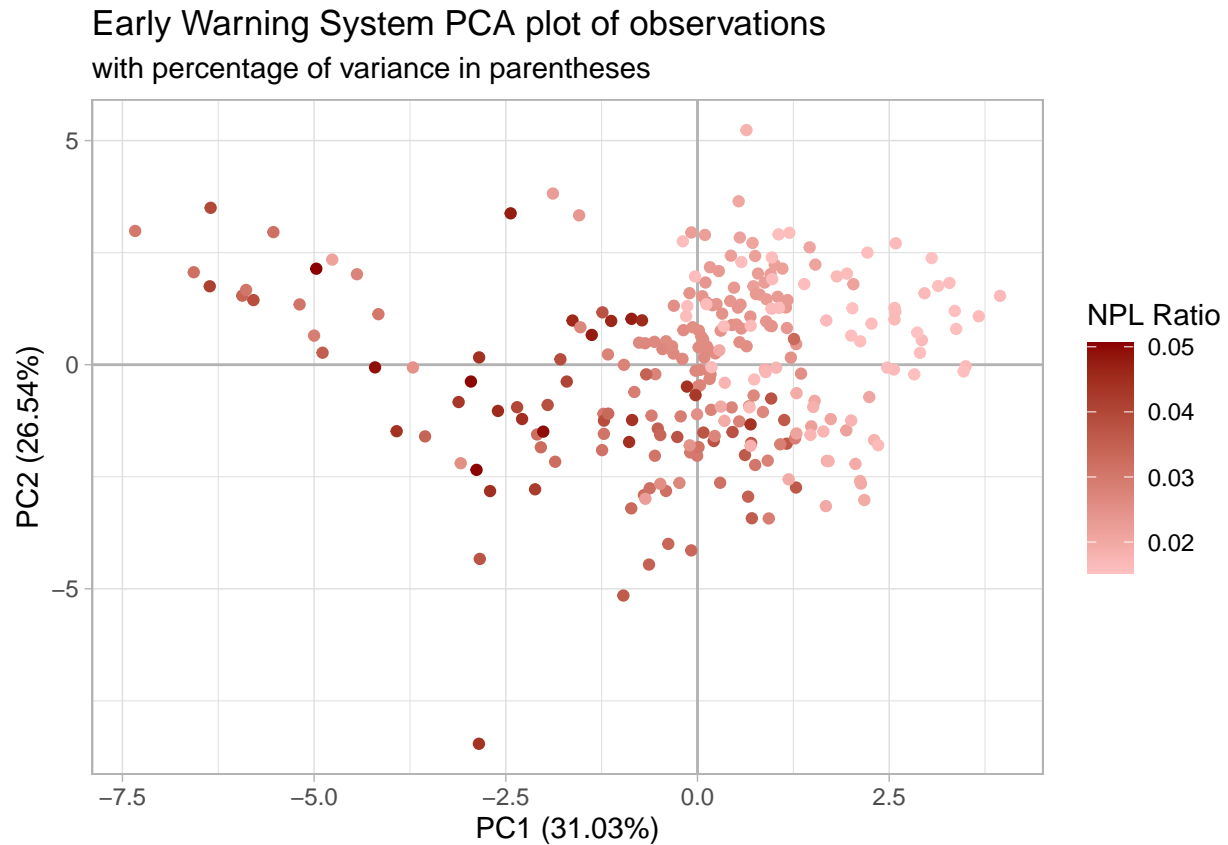
Early Warning System PCA plot of observations
with percentage of variance in parentheses

Figure 7: Scatterplot of PC scores on first two PC axes, applied colour gradient to each point to scale NPL ratio

Conclusion

In this paper, I have carried out a regularised imputed PCA algorithm to impute missing values in historical datasets and used the new transformed and uniform dataset for a second order PCA to extract the factors and sensitivities that drives the performance of the Malaysian banking system stemming from the performance of various macrofinancial indicators. This exercise shows how extremely complex, dynamic and interlinked the banking sector and how small gyrations in this sector can potentially impact the wider economy and vice versa. With this exercise, we can have a better understanding on how a cluster of factors can contribute to the overall macro effect, not simply through their individual mechanical effects, but also through their aggregated micro effects on each other.

Appendix

The regularised iterative PCA algorithm transforms the original workable dataset from the Table 4 to Table 5 (note the number of observations for each variable under the column "Count"):

Table 4: Before transformation

	Minimum	Maximum	Mean	Median	StDev	Count
dd.deposits.r	3.535	5.965	4.414	4.223	0.581	40
fx.deposits.r	1.392	2.608	1.787	1.778	0.263	40
repo.deposits.r	0.138	0.372	0.244	0.231	0.056	40
loan.app.cars.r	0.058	0.122	0.096	0.096	0.012	40
loan.app.construction.r	0.022	0.086	0.046	0.045	0.013	40
loan.app.nonresprop.r	0.087	0.118	0.103	0.102	0.008	40
loan.app.personal.r	0.055	0.091	0.075	0.075	0.007	40
loan.app.resprop.r	0.235	0.318	0.270	0.267	0.021	40
loan.app.workingcapital.r	0.173	0.330	0.238	0.237	0.034	40
loan.yy	-1.346	15.754	5.749	5.317	4.314	40
nfa.yy	-65.803	495.595	13.999	2.813	82.492	40
depo.yy	-27.889	26.595	3.553	3.872	9.183	40
lcr	116.000	145.000	130.325	128.500	8.380	40

Table 5: After transformation

	Minimum	Maximum	Mean	Median	StDev	Count
dd.deposits.r	1.545	5.980	3.926	3.972	0.681	260
fx.deposits.r	-0.255	2.608	1.235	1.250	0.415	260
repo.deposits.r	-0.262	3.342	0.449	0.221	0.669	260
loan.app.cars.r	0.058	0.194	0.112	0.111	0.015	260
loan.app.construction.r	0.013	0.098	0.047	0.046	0.013	260
loan.app.nonresprop.r	0.059	0.175	0.110	0.108	0.018	260
loan.app.personal.r	0.020	0.091	0.058	0.059	0.013	260
loan.app.resprop.r	0.149	0.348	0.248	0.252	0.034	260
loan.app.workingcapital.r	0.155	0.397	0.234	0.235	0.035	260
loan.yy	-20.843	50.468	7.995	7.120	8.592	260
nfa.yy	-67.445	495.595	16.072	4.951	54.420	260
depo.yy	-27.889	72.020	8.688	7.300	13.242	260
lcr	108.695	145.000	127.158	127.346	5.868	260

Focusing on the first three PCs, the contribution for each variable are plotted in the following figures:
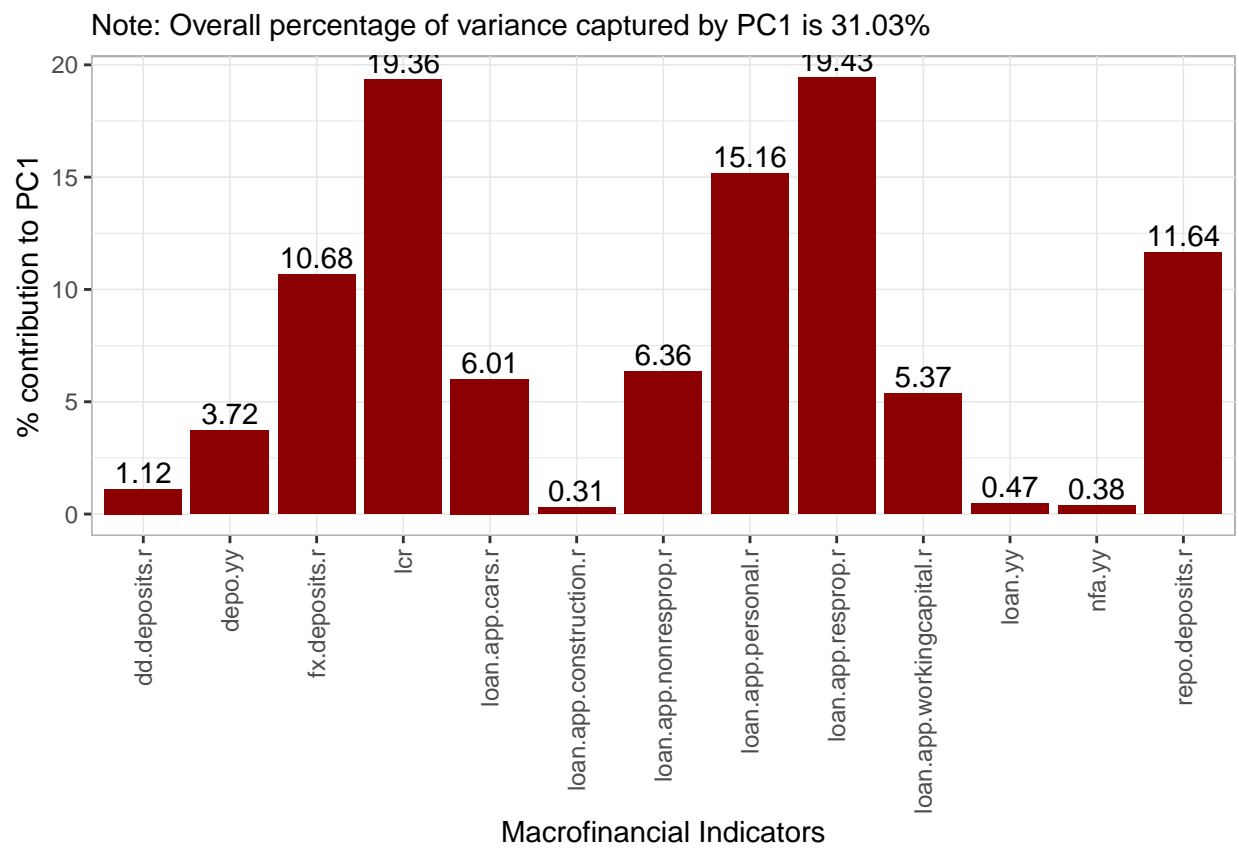
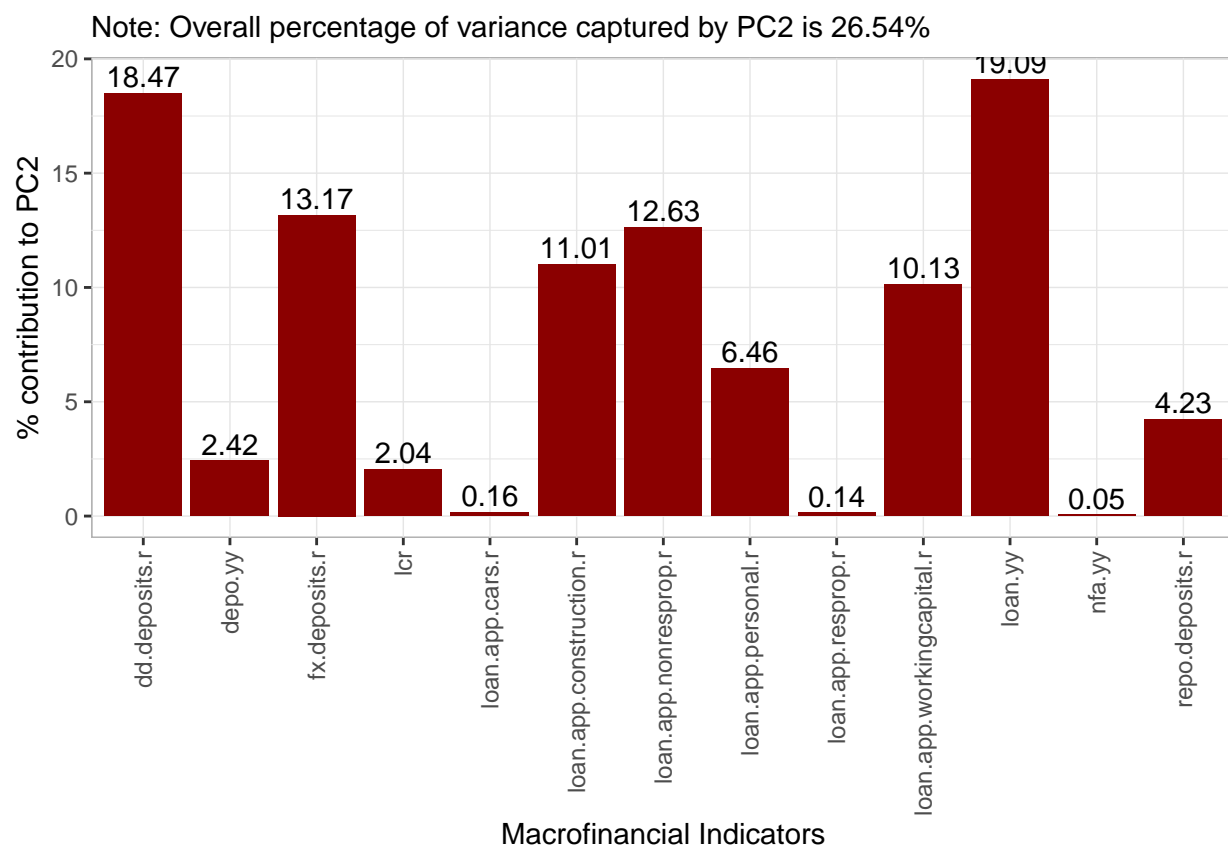Figure 8: Influence of each macrofinancial indicators on PC1

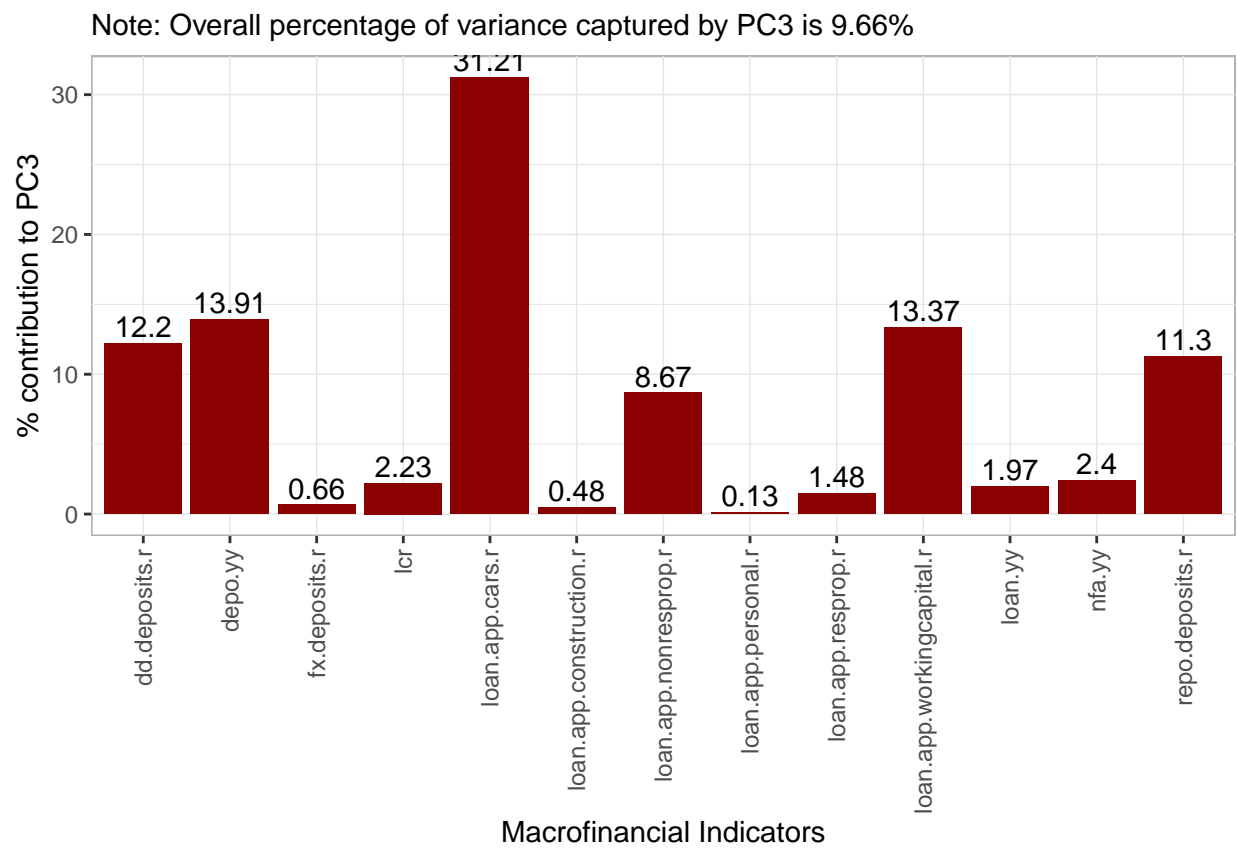Figure 9: Influence of each macrofinancial indicators on PC2

Figure 10: Influence of each macrofinancial indicators on PC3