



Predicting Inflation Crisis in 13 African Countries


By Camille Kirongozi

Research Question and Hypothesis

Research question: Which factors are significantly associated with inflation crisis in the 13 African countries observed?

Hypothesis: Null hypothesis, there are no significant factors that are associated with inflation crisis at a .05 significance level. Alternative hypothesis, there are factors associated to inflation crisis at the .05 significance level.

The four assumptions of the logistic regression:

1. A binary outcome variable: The outcome variable was the inflation_crisis variable which had values 1 and 0.
 2. Independent observations: This means there should be no repeated observations in the dataset. The dataset did not include any repeated observations. Each observation was a yearly economic summary for a country.
 3. Absence of multicollinearity: I ran a variance inflation factor (VIF) command on the model I decided to use, and no variables had a value of more than ten. A value of ten would indicate multicollinearity.
 4. A linearity between the numeric independent variables and the log odds: To test this assumption, I created a plot for the only numeric variable `exch_usd` against the log odds of the dependent variable. Linearity was observed.
- 

Data Analysis Process

- Removed the following columns: case number, three letter country code, country name, independence, and

Warning message:

glm.fit: fitted probabilities numerically 0 or 1 occurred

- Altered the banking_crisis variable to have values 1/0 instead of crisis/no crisis.
- Removed some rows based on the outliers found in the exch_usd and currency_crises variables.
- The steps reduced the number of variables from 14 to 9 and the number of observations from 1059 to 1053.
- Produced a summary statistic report for the cleaned data set along with univariate and bivariate plots.
- Data set was split into Train/Test set using an 80/20 ratio.
- Using the training set, I created the first logistic regression model to include all variables and then a second with only significant variables.



Results

- Rejected the null hypothesis.
- The final logistic regression model provided predictions at an accuracy rate of 90%.

```
Call:
glm(formula = inflation_crisis ~ ., family = binomial(link = "logit"),
    data = train_set)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9226  -0.3683  -0.3629  -0.3377   2.8080

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.618467   8.282938  -0.195   0.84508
year           -0.000540   0.004225  -0.128   0.89830
systemic_crisis1 -0.203886   0.703459  -0.290   0.77194
exch_usd       -0.004747   0.001586  -2.993   0.00276 **
domestic_debt_in_default1  0.094502   0.576391   0.164   0.86977
sovereign_external_debt_default1  1.217979   0.400236   3.043   0.00234 **
gdp_weighted_default -1.785009   2.669839  -0.669   0.50376
currency_crisis1  2.081163   0.259506   8.020 1.06e-15 ***
banking_crisis1  1.368875   0.668353   2.048   0.04055 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 645.10  on 841  degrees of freedom
Residual deviance: 503.72  on 833  degrees of freedom
AIC: 521.72

Number of Fisher Scoring iterations: 5
```

```
Call:
glm(formula = inflation_crisis ~ exch_usd + sovereign_external_debt_default +
    currency_crisis + banking_crisis, family = binomial(), data = train_set)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9189  -0.3630  -0.3626  -0.3391   2.8059

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.68650    0.16171 -16.613 < 2e-16 ***
exch_usd       -0.00471    0.00151  -3.119 0.001817 **
sovereign_external_debt_default1  1.14020    0.29572   3.856 0.000115 ***
currency_crisis1  2.08868    0.24828   8.413 < 2e-16 ***
banking_crisis1  1.21861    0.34902   3.492 0.000480 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 645.10  on 841  degrees of freedom
Residual deviance: 504.46  on 837  degrees of freedom
AIC: 514.46

Number of Fisher Scoring iterations: 5
```

Limitations

- The data provided only included some basic economic information for the countries.
- The years included in the data set were 1860 to 2014.
- One country had data starting from 1860, and only eleven out of thirteen countries had data in 2014.
- I was limited in making the predictions at a one hundred percent accuracy rate.



Proposed Action/Expected Benefits

- Focus on economic development by creating more jobs and implementing monetary policies to prevent currency and banking crisis.
- Promote narrow banking to address banking crisis.
- adopting a floating exchange rate and raise interest rates to address currency crisis.
- Taking the following actions should alleviate the sovereign external debt in default and in turn the exchange rate to USD.
- The countries can expect to observe less occurrences of inflation crisis.



The End

Thank you for your time and attention.



Works Cited

- "A Box and Whiskers Plot (in the Style of Tukey) - geom_boxplot." - *geom_boxplot* • *ggplot2*, 2021, https://ggplot2.tidyverse.org/reference/geom_boxplot.html.
- Chiri. "Africa Economic, Banking and Systemic Crisis Data." *Kaggle*, 21 July 2019, <https://www.kaggle.com/chirin/africa-economic-banking-and-systemic-crisis-data>.
- Cloud Team, IBM. "Python vs. r: What's the Difference?" IBM, 23 Mar. 2021, www.ibm.com/cloud/blog/python-vs-r#:~:text=R%20is%20an%20open%20source,elegant%20tools%20for%20data%20reporting.&text=Creating%20visualizations,learning%20and%20deep%20learning%20algorithm.
- "Currency Crisis: Causes, Signs, Impacts and Possible Solutions." *Penpoin.*, 15 Sept. 2021, <https://penpoin.com/currency-crisis/>.
- Gelman, Andrew. "Statistical Modeling, Causal Inference, and Social Science." *Statistical Modeling Causal Inference and Social Science*, 10 Jan. 2020, <https://statmodeling.stat.columbia.edu/2020/01/10/linear-or-logistic-regression-with-binary-outcomes/>.
- Hartman, Mitchell. "Why Is Inflation Necessary?" *Marketplace*, 12 Sept. 2019, <https://www.marketplace.org/2019/09/12/why-is-inflation-necessary/>.
- Pennacchi, George. *Narrow Banking*. ANNUALREVIEWS; Journal: ARFE:Annual Review of Financial Economic, 2012, <https://gpennacc.web.illinois.edu/GPNarrowBankARFE.pdf>.
- Rulphus, Adam, and Roman Ahmed. "Logistic Regression Model: " Glm.fit: Fitted Probabilities Numerically 0 or 1 Occurred "." *RStudio Community*, 19 June 2018, <https://community.rstudio.com/t/logistic-regression-model-glm-fit-fitted-probabilities-numerically-0-or-1-occurred/9828/15>.
- Schreiber-Gregory, Deanna. *Logistic and Linear Regression Assumptions: Violation Recognition and Control*. Henry M Jackson Foundation for the Advancement of Military Medicine, 2018, https://www.lexjansen.com/wuss/2018/130_Final_Paper_PDF.pdf.
- Vettoretti, Martina, and Barbara Di Camillo. "A Variable Ranking Method for Machine Learning Models with Correlated Features: In-Silico Validation and Application for Diabetes Prediction." *Applied Sciences*, vol. 11, no. 16, 2021, p. 7740., <https://doi.org/10.3390/app11167740>.
- Vidhya, Analytics. "Logistic Regression R: Introduction to Logistic Regression." *Analytics Vidhya*, 1 Nov. 2015, <https://www.analyticsvidhya.com/blog/2015/11/beginners-guide-on-logistic-regression-in-r/>.