



Electronic Theses and Dissertations

2017

Determine the breaking point of Kenya debt an application of extreme value theory

Jacqueline Wachuka Mathenge
Strathmore Institute of Mathematical Sciences (SIMs)
Strathmore University

Follow this and additional works at <https://su-plus.strathmore.edu/handle/11071/5766>

Recommended Citation

Mathenge, J. W. (2017). *Determine the breaking point of Kenya debt an application of extreme value theory* (Thesis). Strathmore University. Retrieved from <http://su-plus.strathmore.edu/handle/11071/5766>

Determine the Breaking Point of Kenya Debt
An Application of Extreme Value Theory

Mathenge, Jacqueline Wachuka

Submitted in partial fulfilment of the requirements for the award of the degree of Master
of Science in Statistical Science at Strathmore University.

Institute of Mathematical Sciences
Strathmore University
Nairobi, Kenya

June, 2017

This thesis is available for Library use on the understanding that it is copyrighted
material and that no quotation from the thesis may be published without proper
acknowledgement.

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

© No part of this thesis may be reproduced without the permission of the author and Strathmore University

Mathenge, Jacqueline Wachuka

.....

June, 2017

Approval

The thesis of Mathenge, Jacqueline Wachuka was reviewed and approved by the following:

Prof. Samuel Mwalili

Lecturer- Institute of Mathematical Sciences

Strathmore University

Ferdinand Otieno

Dean- Institute of Mathematical Sciences

Strathmore University

Prof. Ruth Kiraka

Dean- School of Graduate Studies

Strathmore University

Abstract

The aim of the study is to determine the breaking point of Kenyan public debt through the use of Extreme Value Theorem (EVT). EVT focuses on the tail end of distributions to be able to identify maxima and minima points. With the rising debt levels since devolution, from Kenya Shillings (KES) 500 billion in 2013 to KES 2.5 trillion in 2015, and warnings from international bodies such as International Monetary Fund (IMF) and World Bank on rising debt levels, there is need to determine sustainability of debts beyond analyst speculations. The use of the special case of EVT known as Generalized Extreme Value (GEV) application looks at a degenerate distribution factor thus ensuring the tail end of the distribution, that is, the maxima converges to the GEV despite the distribution of the data set (no assumption on the distribution of the data set). From the study the Gumbel model was determined to be the most appropriate model and with a 95% threshold, the GEV projected total debt maxima to be KES 5 trillion. This is evidence that the current debt levels of KES 2.5 trillion is still sustainable but should however be monitored.

Table of Contents

Chapter 1: Introduction.....	1
1. 1: Background of the study	1
1. 2: Statement of the Problem	3
1. 3: Research questions	3
1.3.1: Primary Research Question.....	3
1.3.2: Secondary Research Questions	3
1. 4: Objectives of the study	4
1.4.1: General objective.....	4
1.4.2: Specific objectives	4
1. 5: Scope of the study	4
1. 6: Significance of the study	4
1. 7: Limitations of the study	5
Chapter 2: Literature Review	6
Chapter 3: Research Methodology	9
3. 1: The Research Design.....	9
3.1.1: Step 1 - Find the significant variables	9
3.1.2: Step 2 - Find the dependent variables influencing debt.....	10
3.1.3: Step 3 - Extreme Value Theory and Likelihood function.....	12
3. 2: Population and sampling.....	14
3. 3: Data Collection Methods.....	15
3. 4: Data analysis	15
Chapter 4: Results	16
4. 1: Exploratory Data Analysis	16
4. 2: Simulation: GEV distribution.....	18
4. 3: Results from GEV Analysis.....	23
Chapter 5: Conclusion	25

5. 1: Conclusion.....	25
5. 2: Further Research.....	26
5. 3: Limitations of the Study	26
Chapter 6: References	28
Chapter 7: Appendix	30
7. 1: Install R-package	30
7. 2: Exploratory Data Analysis.....	30
7. 3: GEV Output	31

Table of Figures

Figure 4.1: Time Series Plot of Total Kenya Public Debt.....	16
Figure 4.2: Time series plot of Domestic (blue) and External (red) debts.....	17
Figure 4.3: A multiple scatter plot to show the association between Total Debts and select predictor variables	18
Figure 4.4: Simulated GEV with $\xi = 0.5$ corresponding to Fréchet distribution.....	19
Figure 4.5: Simulated GEV with $\xi = 0.5$ corresponding to Gumbel distribution.....	20
Figure 4.6: Simulated GEV with $\xi = -0.5$ corresponding to Weibull distribution.....	21
Figure 4.7: Hill plot estimate of the tail index of heavy-tailed data	22
Figure 4.8: GEV fit to the the Kenya's Total Debt suggesting a Gumbel model.....	23
Figure 4.9: GEV projection of the Kenya's Total Debt showing a 95% threshold of 5 trillion shillings.....	24

List of Tables

Table 4.1: Estimated parameters from GEV model of the Total Debts	23
Table 4.2: Estimates of the regression coefficient from GEV model ($\xi=0$)	24

Abbreviations

AIC	Akaike Information Criteria
DFM	Debt and Finance Management
DSA	Debt Sustainability Analysis
EVT	Extreme Value Theorem
GDP	Gross Domestic Product
GEV	Generalized Extreme Value
iid	Independent and Identically Distributed
IMF	International Monetary Funds
KENBS	Kenya National Bureau of Statistics
KES	Kenya Shilling
Lasso	Least Absolute Shrinkage and Selection Operator
LICs	Low Income Countries
NIPALS	Nonlinear Iterative Partial Least Squares
NSE	Nairobi Securities Exchange
PLS	Partial Least Squares
PLS-BETA	Partial Least Squares Regression
PLS-VIP	Partial Least Squares- Variable Importance in the Projection
UNITAR	United Nations Institute of Training and Research
VAT	Value Added Tax
VIF	Variance Inflation Factor

Acknowledgement

I would like thank God for bringing me this far in life and education. I would also like show my deepest gratitude to my supervisor, Prof. Samuel Mwalili. Without his mentorship, constant feedback, critic and guidance, this dissertation would not have been possible. I would also like to thank my family for their continued support and encouragement throughout my academic years. Lastly, I would like to acknowledge my classmates for their assistants and support from formulation of the dissertation topic to the final dissertation submitted.

Dedication

I would like to dedicate this thesis to my family; Mr. John Mathenge, Mrs. Judy Mathenge, Miss Samantha Wandia, Miss Stephanie Kirigo and Miss Christine Karaimu.

Chapter 1: Introduction

1. 1: Background of the study

Economic theory suggests that reasonable levels of borrowing by a developing country are likely to enhance its economic growth (Patillo, Poirson and Ricci, 2002). Thugge, Heller and Kiringai (2008) stated that increasing levels of public debt has a worrying significant impact on economic growth to most developing countries. Were (2001) wrote that indebtedness does not necessarily lead to lack of economic growth. The main problem arises when a country is not able to meet its debt obligations.

Kenya debt levels have been on a rapid increase over the years. Since the implementation of the new constitution, the rise in debt limits has been almost annual. Ochieng (2013) noted that Kenya debt levels are governed by Treasury, under the External Loans and Credit Act, CAP. 422 of the Laws of Kenya which limits the total indebtedness in respect of principal amount to KES 500 billion. However, in 2013, the debt limit was raised to KES 800 billion and in 2014 to KES 1.2 trillion and finally to KES 2.5 trillion in 2015 (National Treasury, 2016).

Wafula (2010) studied debt sustainability and optimal debt for the government of Kenya that could enhance a 10% growth as projected in the Vision 2030. Cointegration testing of the present value of budget constraint was used to analyse the sustainability of the historical fiscal process and simulation was used to determine the optimal debt level. The findings were that public debt was sustainable, pointing to prudent public sector policies by the fiscal authorities. Further, the study found out the optimal debt level to be 35.2%. This study however focused on Vision 2030 as the primary forecasting factor but this is an assumption that Vision 2030 will be achieved with no variability in the outcomes or effect.

Nevertheless, looking at actual debt levels, the change in debt limit has caused an impact in borrowing levels and debt ratios. The actual debt levels went up from KES 2.8 trillion in 2015 to KES 2.3 trillion in June 2016. The current budget allocated KES 466 Billion towards debt repayment, accounting for a fifth of the yearly government budget (National Treasury, 2016). The increase in debt has caused the debt to GDP ratio to jump from 39.8% in 2013 to 52.85% in 2015. The main cause of this was due to un-proportional increase in

debt compared to GDP. This puts Kenya at position 58 out of 161 countries and 1 in East Africa (Trading Economics, 2016).

This shows that Wafula (2010) optimal debt levels cannot be applied to the real world scenario whereby debt level are used to meet the ever increasing fiscal deficit the country has been experiencing. The main reasons for the increase in debt are need for resources to meet public expenditure and finance past debts. Taxes largely provide the bulk of the revenue but public and private borrowings bridge the resource gap between receipt and expenditure (UNITAR-DFM E-Learning, 2008).

Use of debt for development projects has a long term economic growth outcome. However, Ngugi (2016) noted that with the increased cases of corruption in most developing countries such as Kenya, borrowings could soon begin to strain government finances as it does not have the intended growth effect (Ngugi,2016).

This forces more resources to be diverted towards debt service, leaving less money for development expenditure. A misuse of resources may easily lead to a build-up of debt to unsustainable levels which has been a major impediment to growth in emerging economies.

Stieglitz (2000) stated that government borrowing can crowd out investment, which in turn reduces future output and wages. When output and wages are affected the welfare of the citizens will be made vulnerable. These claims deserve serious attention in the context of the country's crying need to generate faster employment growth, meet the Millennium Development Goals and attain the Vision 2030 goals (Achieng, 2010).

A joint World Bank and IMF debt sustainability analysis (DSA) in April 2013 concluded that Kenya's debt is sustainable relative to the total size and productivity of the economy since the macroeconomic fundamentals are strong (World Bank, 2013). Macroeconomic variables nonetheless are ever changing with the increase in debt the country has been experiencing. In order to ensure Kenya's debt stays within manageable levels based on trends being experienced and forecasting from the trends, 'Extreme Value Theorem' (EVT) will be applied to try and find the sustainable level of debt for Kenya, given certain macroeconomic factors.

1. 2: Statement of the Problem

There have been many studies carried out to try and find the correlation between indebtedness and certain macroeconomics and microeconomic variables. Kabau (2009) showed that an increase in debt leads to an escalation in macroeconomic variables such as real interest rates, inflation, exchange rate, taxes and total debt service.

The findings of a study carried out by M' Amanja & Morrissey (2003) on fiscal policy on economic growth in Kenya revealed that external loans had a negative impact on long run growth and domestic debt. This was proved further by Achieng (2010) who sought to find out the effect of domestic debt on private investment. The findings showed that the debt service ratio domestic debt were significant at a confidence interval of 5%.

A study done by Njuru (2012) found that in Kenya, fiscal policy structures influenced private investment. However, Maana, Owino and Mutai (2008) proved that there was a positive and insignificant relationship between domestic debt and economic growth in Kenya.

As noted, research has been focusing on effect of debt on investment, growth and other microeconomic and macroeconomic variables. There is however need to find a debt cap to ensure that although debt levels are rising, they are still sustainable. Reinhart (2011) found that failure to monitor debt levels, especially external debt in a country can lead to external debt crisis. This study will not only find the relationship between debt and the microeconomic and macroeconomic variables but use these variables to try and derive a debt limit, beyond which, economic down turn is likely to occur.

1. 3: Research questions

1.3.1: Primary Research Question

Can 'Extreme Value Theorem' be used to estimate a breaking point in Kenya debt?

1.3.2: Secondary Research Questions

1. What macroeconomic and microeconomic variables are statistically significant when calculating sustainable debt levels?
2. What are the local and global minima and maxima for debt in Kenya?

1. 4: Objectives of the study

1.4.1: General objective

The aim of the study is to ascertain if there is an identifiable maximum debt levels given the predictor variable growth over time from historical data.

1.4.2: Specific objectives

1. To find the explanatory variables that influence debt levels.
2. To find the global and local maxima of debt from historical data and forecasted growth in explanatory variables.

1. 5: Scope of the study

The study will cover data from 2000-2016. The time series variable of interest are public debt, GDP, government revenue collection, real interest rates, exchange rates, investment rate, population growth, mortality rate and death rate. The dummy variables that will also constitute break points in the series are political changes (election periods), exchange rate regime changes, post-election violence, global market upheavals (Stock Market Crash and Housing market crash in America). The data will be analysed as panel data as it will have various variables analysed over time and space.

1. 6: Significance of the study

Autor (2012), stated that the inability of national government in European countries to tighten fiscal policies during the booming period led to companies in the private sector increasing their risks. During the booming season, revenues went up in terms of taxes, assets prices rising, increased investment and so on. However, most of the increased revenue collection by governments was being used to cater for extra public spending and tax cuts and low amounts actually went to fiscal improvements.

Reinhart (2011) found that failure to monitor debt levels, especially external debt in a country can lead to external debt crisis. External debt crisis as defined by Autor (2012) is the default on debt payments obligations incurred under foreign legal jurisdiction. Countries that fail to pay their debt are forced into two main actions repudiation or the

restructuring of debt. Restructuring is normally a last resort as it has less favourable terms than the original terms. Countries such as Russia and Greece took over 69 and 52 years respectively to recover.

Despite the effects and worries being raised by countries and international organizations, there is no concrete debt cap in place. The purpose of this research is to prevent a crisis by applying 'Extreme Value Theory', to be able to find the global maxima of public debt in Kenya.

1. 7: Limitations of the study

Due to pay-offs of most development projects being in the long term, the effect of the projects funded by debt cannot be factored in to the model and so the long term growth effect of the debt on GDP will not be a factor in coming up with a break point. Even if the effect of the projects can be calculated, the actual impact time cannot be accurately predicted due to delays and implementation problems.

Further study can be done based on past projects to determine effect of future projects. The EVT applied on to the variables will be based on forecasted data and thus has no certainty of occurrence, that is, probability will be applied.

Chapter 2: Literature Review

In line with Vision 2030, Kenya has been actively undertaking public development projects to improve the welfare of citizens and stimulate economic growth (Ngugi, 2016). Public debt has been used to finance most of the development projects and this has formed a heavy burden to the country (Were, Ngugi and Makau, 2006). Reinhurt (2011), found that there was a positive correlation between debt burdens and incidents of default. Currently, a fifth of Kenya's yearly budget is put towards re-servicing loans (Treasury, 2016).

Previous studies used to focus on government bonded debt. While studying bond market development determinant, Burger and Warnock (2006) used cross sectional data (22 emerging countries in a study with 49 countries) and found that there is a positive correlation between declining inflation, rule of law, and country size with domestic government bond market development. Fiscal balance and GDP growth were found to be negatively correlated with the size of the government bond market.

Claessens, Klingebiel and Schmuckler (2007) used panel data to study the determinants of the development of the market for local currency government bonds based on 36 countries, of which 12 are emerging markets. They found that country size, size of the banking system (as measured by total deposits over GDP), good institutions, low inflation, flexible exchange rates, and fiscal burden are positively correlated with the size of the domestic bond market. Eichengreen and Luengnaruemitchai (2004) however contracted the findings of Claessens et al (2007) as they found that lower exchange rate volatility is positively correlated with the size of the domestic bond market and argued that this might be due to the fact that a fixed exchange rate lowers currency risk and may encourage foreign participation.

Researchers built on to this by focusing on sovereign debt as a whole in recent years. Kumar and Woo (2010) found that controlling for other growth determinants and based on number of econometric techniques, there is an inverse relationship between initial debt and subsequent growth. They also found that there is proof of nonlinearity with higher levels of initial debt having a proportionately larger negative effect on subsequent growth. Breaking down growth, shows that the adverse effect chiefly reflects a slowdown in labour

productivity growth primarily due to reduced investment and slower growth of capital stock. Forslund, Lima and Panizza (2011) found there is no statistically significant effect of inflation history on public debt composition. Arezki, Candelon and Amadou (2011) found that that sovereign rating announcements have statistical and economical significant spill over effects on both across countries and financial markets. This implies that rating agencies announcements could spur financial instability.

With conflicting finding based on country study, year and assumptions laid out, the above variables (whether found to be historically significant or not) will be used in the initial model to be able to draw out the relevant variables for the Kenya case. Other variables such as population, political changes, post-election violence will also be factored into the model to ensure the output is as accurate as possible. The EVT will be used to come up with the maxima and minima of debt levels based on regressing and forecasting the statistically significant variables that affect debt.

Medford (2015) wrote that the application of Extreme Value Statistics allows us to investigate the behaviour of a stochastic process at very high or very low levels. EVT has been used in the field of modern science and engineering to model rare events that have significant consequence (Gilli and K llezi, 2003). However, EVT has found a place in Finance by being used to model risk in the financial sector (Embrechts, Resnick and Samorodnitsky, 1999).

EVT was derived by Bernard Bolzano (1830s) and Karl Weierstrass (1860). The theory states that:

'If $f(x)$ is a continuous function defined on a closed interval $[a, b]$, then the function attains its maximum value at some point c contained in the interval'.

The basic understanding of EVT is that if a function $f(x)$ is continuous on a closed interval $[a, b]$, then $f(x)$ has both a maximum and a minimum on $[a, b]$. If $f(x)$ has an extremum on an open interval (a, b) , then the extremum occurs at a critical point c . The proof of this is that if f is the continuous image of a compact set on the interval $[a, b]$, so it must itself be compact. Since $[a, b]$ is compact, it follows that the image $f([a, b])$ must also be compact.

To be able to apply EVT, a two-step process must be carried out. One is to determine that the data is continuous on a bounded set $[a, b]$ and two is next step is to determine all critical points in the given interval. The largest function value from the previous step is the maximum value, and the smallest function value is the minimum value of the function on the given interval. This process will be explained further in the research design.

The benefit of EVT is that it can be used to accurately estimate extreme quantiles and tail probabilities and hence used to model risk (Schuermann and Stroughair, 1998). Focusing on the tail end of distributions using limit laws (only the tail ends are important for extreme values) ensures that an actual distribution is not forced on the data. Another benefit is that the parametric modelling of the tails enables for the extrapolation of probability assignments to the quantiles even higher than the most extreme observation in the sample (Gençay and Selçuk, 2004).

Despite the popularity of the EVT, it has one major draw-back in that it assumes variables are independent and identically distributed (iid) (Bollerslev, Chou and Kraner, 1992). This is a severe pitfall as most real data contain an element of dependency. Schuermann and Stroughair, (1998) came up with two solutions to the dependency problem with specific focus on tail end (extreme values).

One is to fit a generalized extreme value distributions directly to series of per-period maxima and the second is to estimate the tail of the conditional rather than the unconditional distribution. The first solution reduces dependency when the periods taken are long enough and Maximum Likelihood Estimation. However, this method reduces efficiency and period length arises.

Chapter 3: Research Methodology

3. 1: The Research Design

The research will focus on quantitative data collected from secondary sources from 2000-2016. The dependent variable will be debt levels while the independent variables will include debt repayment, GDP, inflation, interest rates, exchange rate, NSE returns, population (total population, working age population, mortality rate, and fertility rate), global oil prices and unemployment rate. There will however be several dummy variables to account for random change in economic and political climate. They will include: change in exchange rate to a floating system, elections, post-election violence, implementation of new constitution, European crisis, and the global financial crisis.

The analysis will be a 3 step process:

3.1.1: Step 1 - Find the significant variables

The data to be analysed is time series data. As is the case with most time series data, there will be multicollinearity in the data set. This provides a need to test for multicollinearity between the explanatory variables. The Variance Inflation Factor (VIF) is used to test multicollinearity. In this case the each explanatory variable is regressed against the other explanatory variables. The VIF is such that:

$$VIF_i = \frac{1}{1-R_i^2}, \quad (3.1)$$

Where R_i^2 is the R^2 (statistical measure of how close the data are to the fitted regression line) regression for regressing x_i against x_j .

There is no clear cut off point for VIF but researchers revolve around 5 and 10. In line with prudence, in this paper, 5 will be the cut of point and $VIF_i \geq 5$ will be taken as highly correlated data. Due to expected high multicollinearity, the Partial Least Square Regression (PLS) will be used. The benefit of this model is that it reduces the number of explanatory variables to smaller uncorrelated sets. It is expounded on further in the next step.

3.1.2: Step 2 - Find the dependent variables influencing debt

The variable chosen will be all possible data sets that influence debt levels. However, there is a high level of multicollinearity in the data set. This limits the type of variable selection model that can be used. Chong and Jun (2004) tested various models using data containing multicollinearity. They designed 108 experiments from observations drawn from true models. They looked at four factors when analysing the output. They are the proportion of the number of relevant predictors, the magnitude of correlations between predictors, the structure of regression coefficients, and the magnitude of signal to noise. They evaluated the performance of the partial least squares (PLS), the Lasso (least absolute shrinkage and selection operator), and stepwise method.

They concluded that PLS was the best performing technique for highly correlated data. Comparing 2 types of PLS, that is PLS-VIP (Variable Importance in the Projection) method and the PLS-BETA (regression model) method. They found that the 2 methods are complementary. Wold, Johansson and Cocchi (1993), recommend a combination of PLS-VIP and PLS-Beta for variable selection, which states that both should be small for a variable to be excluded.

For the purpose of this study, both the PLS-VIP and PLS-Beta will be used to come up with a parsimonious method. The hypothesis will be:

$$H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \dots, \beta_k = 0.$$

$$H_1: \beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0, \dots, \beta_k \neq 0.$$

The formula used for PLS-VIP and Beta are derived from the partial linear regression (Wold, Johansson and Cocchi, 1993):

Suppose y is the response variables and there are p predictors. The PLS regression model with h latent variables such that $h \leq p$ can be expressed as:

$$X = TP^t + E \tag{3.2}$$

$$y = Tb + f \tag{3.3}$$

X , T , y and b are used for predictors X scores, X loadings, a response and regression coefficients of T . The k_{th} element of column vector b explains the relation between y and t_k , the k_{th} column vector of T . E and f are the random errors of X and y respectively. A weighted matrix W is obtained using the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm. This makes $\|f\|$ (Euclidian norm) as small as possible and derive a relationship between X and y .

The NIPALS algorithm works as follow:

Suppose that matrix X and column vector y have been standardized to have mean 0 and unit variance. The model parameters are determined by:

$$\mathbf{y}_{(k)} \leftarrow \mathbf{y}_{(k-1)} - \mathbf{b}_{k-1} \mathbf{t}_{k-1}; \mathbf{y}_{(1)} \leftarrow \mathbf{y} \text{ and } \mathbf{X}_{(k)} \leftarrow \mathbf{X}_{(k-1)} - \mathbf{t}_{k-1} \mathbf{p}_{k-1}; \mathbf{X}_{(1)} \leftarrow \mathbf{X} \quad (3.4)$$

$$\mathbf{w}_k^t = \mathbf{y}_{(k)}^t \mathbf{X}_{(k)} / \mathbf{y}_{(k)}^t \mathbf{y}_{(k)} \quad (3.5)$$

$$\mathbf{w}_k^t \leftarrow \mathbf{w}_k / \|\mathbf{w}_k\| \quad (3.6)$$

$$\mathbf{t}_k = \mathbf{X}_{(k)} \mathbf{w}_k / \mathbf{w}_k^t \mathbf{w}_k \quad (3.7)$$

$$\mathbf{p}_k^t = \mathbf{t}_k^t \mathbf{X}_{(k)} / \mathbf{t}_k^t \mathbf{t}_k \quad (3.8)$$

$$\mathbf{t}_k \leftarrow \mathbf{t}_k \cdot \|\mathbf{p}_k\| \quad (3.9)$$

$$\mathbf{w}_k \leftarrow \mathbf{w}_k \cdot \|\mathbf{p}_k\| \quad (3.10)$$

$$\mathbf{p}_k \leftarrow \mathbf{p}_k \|\mathbf{p}_k\| \quad (3.11)$$

$$\mathbf{b}_k = \mathbf{y}_{(k)}^t \mathbf{t}_k / \mathbf{t}_k^t \mathbf{t}_k \quad (3.12)$$

From here, the PLS-VIP and PLS-Beta are used to get regression co-efficient estimates.

- **PLS-VIP**

The VIP score is a summary of the importance for the projections to find h latent variables.

The VIP score for the j_{th} variable can be calculated by the equation below:

$$VIP_j = \sqrt{P \sum_{k=1}^h (SS(\mathbf{b}_k \mathbf{t}_k) (\mathbf{w}_{jk} / \|\mathbf{w}_k\|)^2 / \sum_{k=1}^h SS(\mathbf{b}_k \mathbf{t}_k))} \quad (3.13)$$

Where:

$$SS(\mathbf{b}_k \mathbf{t}_k) = \mathbf{b}_k^2 \mathbf{t}_k^t \mathbf{t}_k$$

Since the average of squared VIP scores equals 1, 'greater than one rule' is generally used as a criterion for variable selection.

- **PLS-Beta**

The relation of T and W obtained by the NIPALS algorithm is given by:

$$T = XW^* \quad (3.14)$$

Where:

$$W^* = (P^t W)^{-1}$$

The predicted values can be directly calculated by:

$$\hat{Y} = T(T^t T)^{-1} T^t y = Xb_{pls} \quad (3.15)$$

Where:

$$b_{pls} = W(P^t W)^{-1} (T^t T)^{-1} T^t y$$

The relevant predictors could be selected according to the magnitude of the absolute values of regression coefficients.

3.1.3: Step 3 - Extreme Value Theory and Likelihood function

Based on the theoretical research by Cole (2001), suppose there are a sequence of random variables with a common distribution function such that:

$$F(x) \sim (X_1, X_2, \dots, X_n) \quad (3.16)$$

The maximum of the sequence containing n variables will be M_n . EVT focuses on the maxima and minima distribution as n increases such that:

$$\begin{aligned} P(M_n \leq z) &= P(X_1 \leq z, X_2 \leq z, \dots, X_n \leq z) \\ &= P(X_1 \leq z)P(X_2 \leq z) \dots P(X_n \leq z) \\ &= F^n(z) \end{aligned} \quad (3.17)$$

The distribution of $F(x)$ is unknown. This is not a problem as EVT is used to focus on the tail end of the distribution and not the actual data distribution. Thus the function will be

with regards to the tails and not the actual data set. Let M_n have a distribution G that is not influenced by F .

The distribution of M_n is degenerate since as $n \rightarrow \infty$, the distribution function F converges with certainty to a single point. However, linear rescaling of M_n is used to tackle the degenerate limit (Coles, 2001). This is known as Extreme Type Theorem (Fisher and Tippett, 1928; Cole, 2001).

This theorem states that if there exists sequences of constant $\{a_n > 0\}$ and $\{b_n\}$ such that as $n \rightarrow \infty$:

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \rightarrow G(z) \quad (3.18)$$

Where $G(z)$ is a non-degenerate distribution function, then G must be a member of the Generalized Extreme Value (GEV) family of distribution. This ensure despite the underlying distribution, the maxima (tail end) converges to a GEV.

The GEV distribution is given by:

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right\} \quad (3.19)$$

defined on $\{z: 1 + \xi(z - \mu)/\sigma > 0\}$.

Where:

ξ determines the heaviness of the right tail.

The model focuses on 3 parameters:

- Location parameter = μ ($-\infty < \mu < \infty$)

This is the center of the distribution.

- Scale parameter = σ ($\sigma > 0$)

This is the size of the deviation around the center of the distribution (location).

- Shape parameter = ξ ($-\infty < \mu < \infty$)

This forms the behaviour of the tail.

The value of ξ determines the type of right tail distribution. When $\xi < 0$, this is a short tailed Weibull distribution with a bounded upper finite end point. When $\xi > 0$, it forms a Frechet distribution with a decaying polynomial heavy tail. $\xi = 0$, it is a Gumbel type distribution and occurs when the limit of the $G(z)$ equation as $\xi \rightarrow \infty$ and there is an exponential tail decay leading to light tails:

$$G(z) = \exp \left\{ -\exp \left[-\left(\frac{z-\mu}{\sigma} \right) \right] \right\} \quad (3.20)$$

The next step is to attain the log-likelihood function of the GEV. The log-likelihood for the GEV parameters for when $\xi \neq 0$ is:

$$\ell(\mu, \sigma, \xi) = -m \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^m \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \quad (3.21)$$

provided that:

$$1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) > 0 \text{ for } i = 1, \dots, m \quad (3.22)$$

For the case where $\xi = 0$, the Gumbel limit of the GEV distribution is used. This leads to the log-likelihood:

$$\ell(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m \left[\left(\frac{z_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^m \exp \left\{ -\left(\frac{z_i - \mu}{\sigma} \right) \right\} \quad (3.23)$$

Maximization of equations (3.21) and (3.23) with respect to the parameter vectors (μ, σ, ξ) leads to the maximum likelihood estimate with respect to the entire GEV family.

Note: For ξ falling within a small window of 0, equation (3.23) is used.

This will be the applied approach.

3. 2: Population and sampling

Due to the nature of the data, it will be population data. The macroeconomic variables to be used are collected at a national level and so contain information on the total population of Kenya. However, there will be gaps in past data due to poor data storage or a general lack of information. In this case, data will be egressed to fill in gap. This will be clearly identified and pointed out as a limitation in actual output.

3. 3: Data Collection Methods

The data will be collected fully from secondary sources such as institute websites, data collection companies, publications and data documents. The main sources will be Kenya Treasury, Central Bank of Kenya, Kenya Revenue Authority, Nairobi Securities Exchange, Kenya National Bureau of Statistics, World Bank Data Bank, International Monetary Fund, Standard and Poors, Moody and Finch, and Trade Economics.

3. 4: Data analysis

The software of choice will be R-Studio as it contains in built tested out functions required to carry out EVT. The outputs for variable selection will be presented in a cross-box for comparison purposes. Charts and graphs will be used to show outputs of the exploratory analysis and log likelihood EVT.

Chapter 4: Results

4.1: Exploratory Data Analysis

Explanatory data analysis employs a visual of representation in order to maximize insight into a data set.

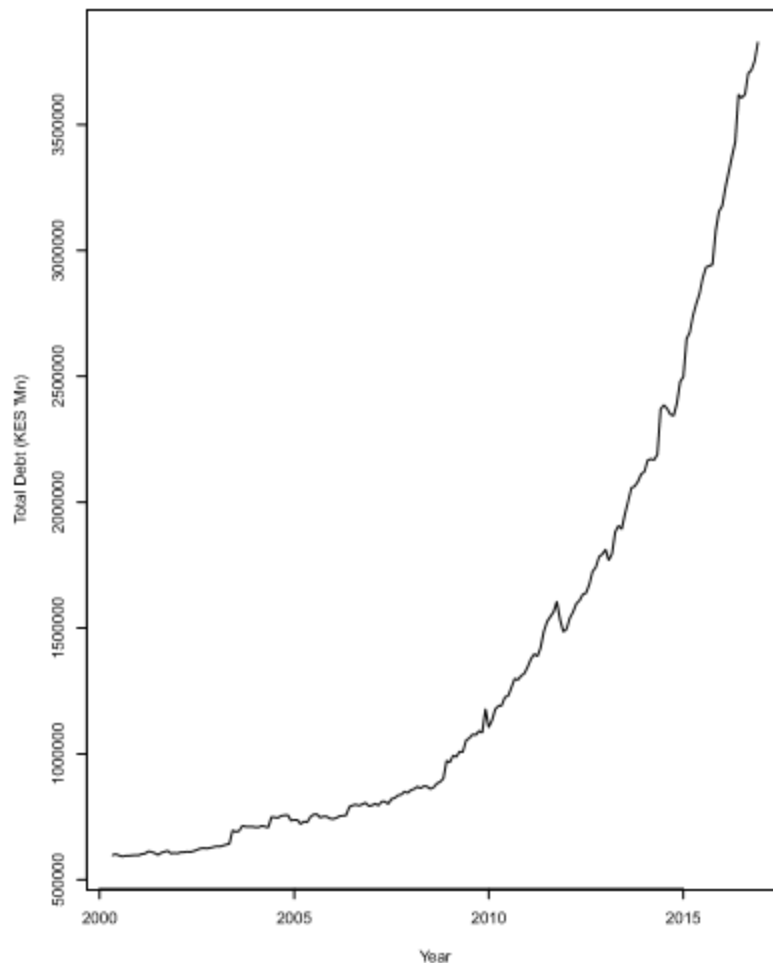


Figure 4.1: Time Series Plot of Total Kenya Public Debt

Figure 4.1 show the profile of the Total Debts in Kenya from 2000 to 2016. Total debt has been rising exponentially since 2010 with the inception of the new constitution that devolved services from the national government to the county government and created new positions in government. This coupled with high cost development plans such as SGR, highways and port upgrading have contributed to the rapid debt levels. Next step is

to break down the debt into its 2 distinct categories, that is, external debt and domestic debt.

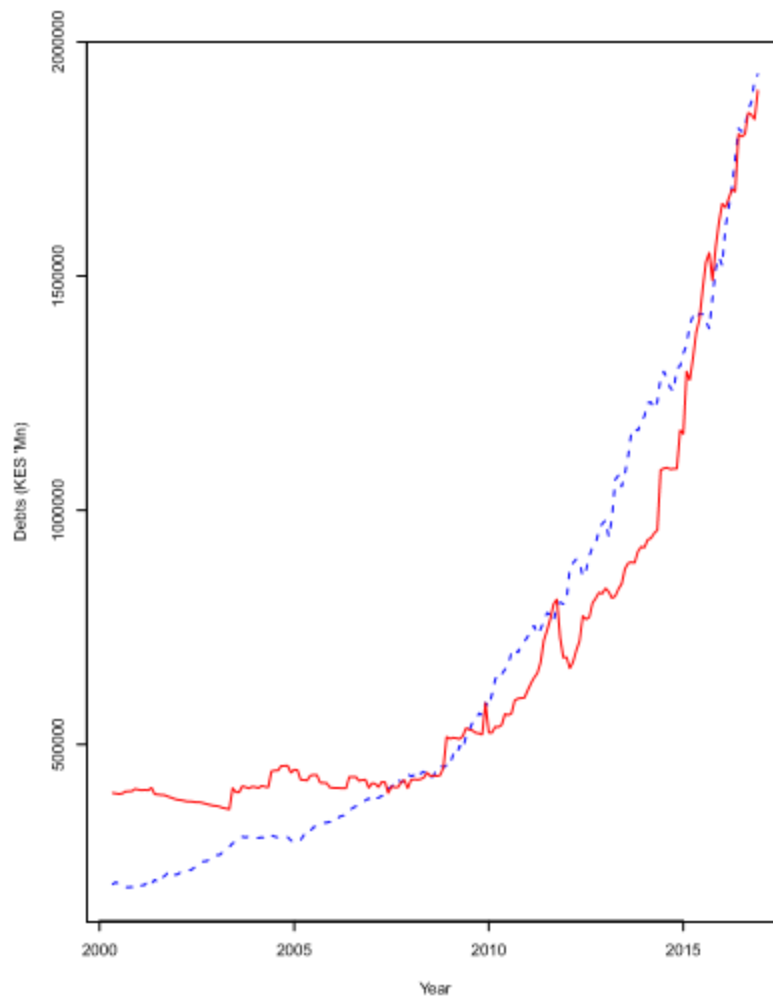


Figure 4.2: Time series plot of Domestic (blue) and External (red) debts

As shown in Figure 4.2 both the domestic and external debts follow the same trajectory, as expected. This shows the government relies on both domestic and external debt to cover revenue deficit. Although reliance on domestic debt was less than that on external debt, over time, they have tended towards equal reliance. The increase on reliance on domestic debt is mainly due to the fact the principle is not paid up on expiration but rather, the debt is restructured to a new interest rate upon maturity.

The next step is to look at the visual relationship between debt and various explanatory variables to check for pattern.

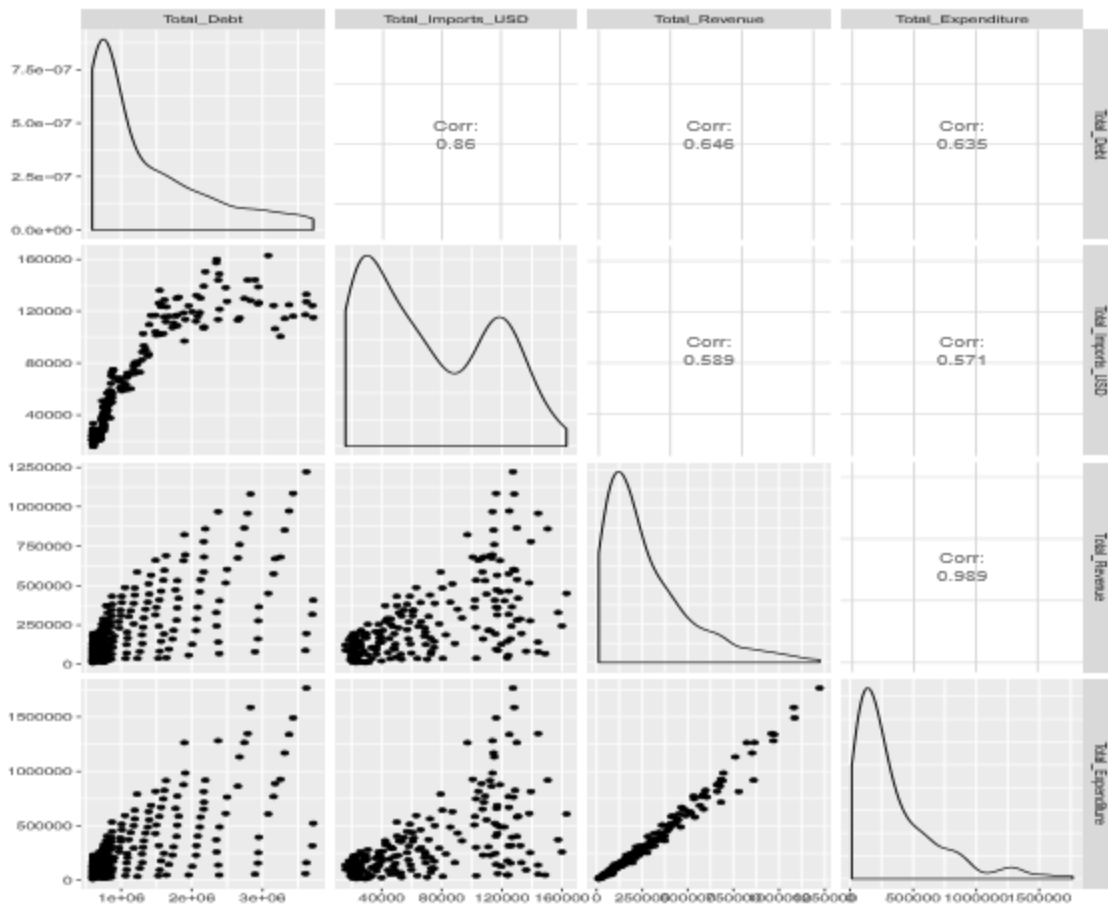


Figure 4.3: A multiple scatter plot to show the association between Total Debts and select predictor variables

As expected, there is a positive relationship between total debt, total imports, total revenue and total expenditure. With a rise in expenditure, revenues (imports are involve) go up as do debt as the government seeks to bridge the gap between the revenue they collect and the expenditure they incur.

4. 2: Simulation: GEV distribution

In this section a simple simulation from GEV distribution of a flexible three parameter model that combines the Gumbel, Fréchet, and Weibull maximum extreme value distributions is performed. Various values of the shape parameter yield the extreme value

type I, II, and III distributions. Specifically, the three cases $\xi = 0, \xi > 0$ and $\xi < 0$ correspond to the Gumbel, Fréchet, and “reversed” Weibull distributions.

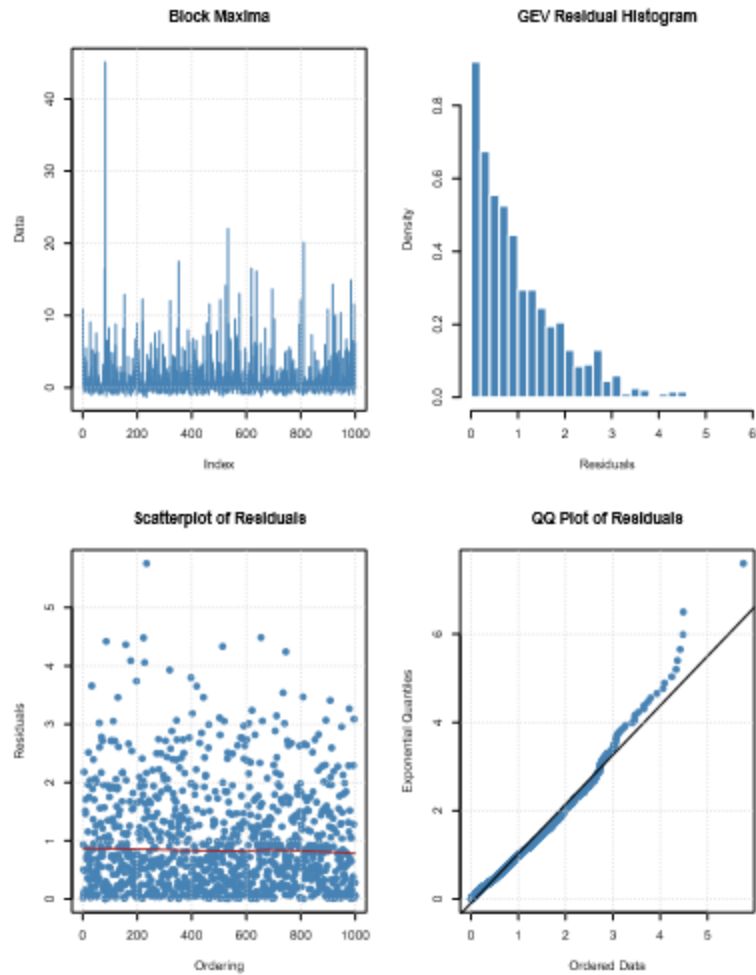


Figure 4.4: Simulated GEV with $\xi = 0.5$ corresponding to Fréchet distribution

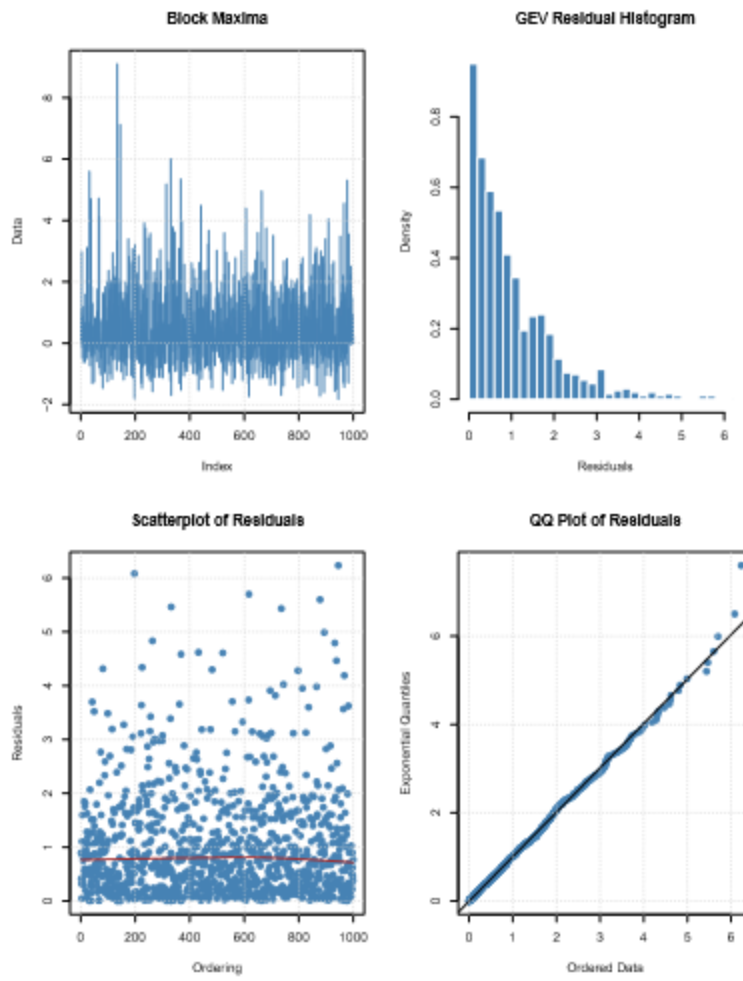


Figure 4.5: Simulated GEV with $\xi = 0.5$ corresponding to Gumbel distribution

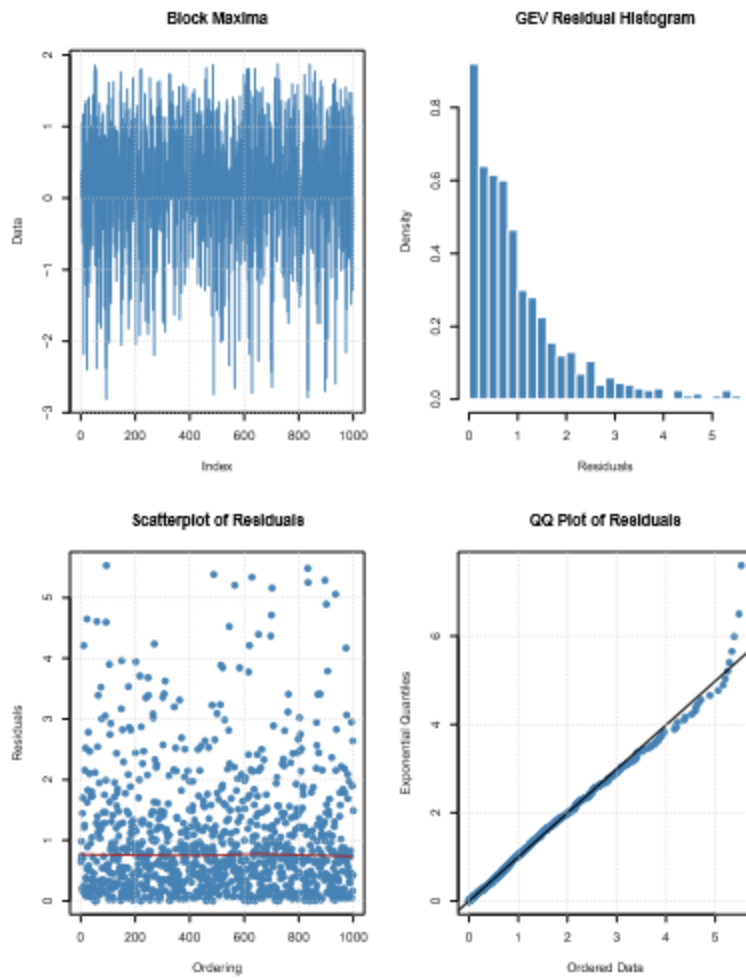


Figure 4.6: Simulated GEV with $\xi = -0.5$ corresponding to Weibull distribution

Looking at the descriptive graphs with regards to linearity and tail deviations, it is evident that the Gumbel model is the best fit for the data.

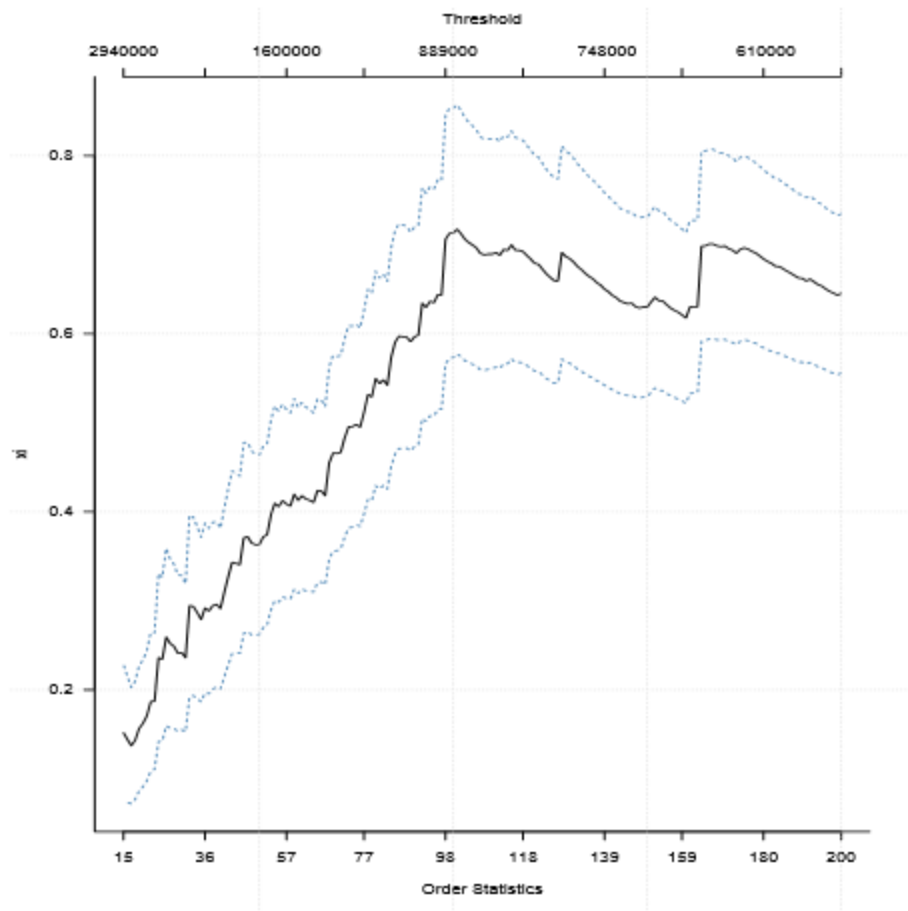


Figure 4.7: Hill plot estimate of the tail index of heavy-tailed data

Figure 4.7 show the hill-plot of the Total Debts. The Hill estimator looks at a choice of the number of order statistics usually referred to as ξ utilized in estimating the tail index. A trade off between the bias and variance of the estimator occurs in the estimation of ξ . The aim is to minimize the asymptotic mean square error. This means that ξ is a function of the heaviness of the tail in the data set. This plot was calculated from the ξ perspective.

4. 3: Results from GEV Analysis

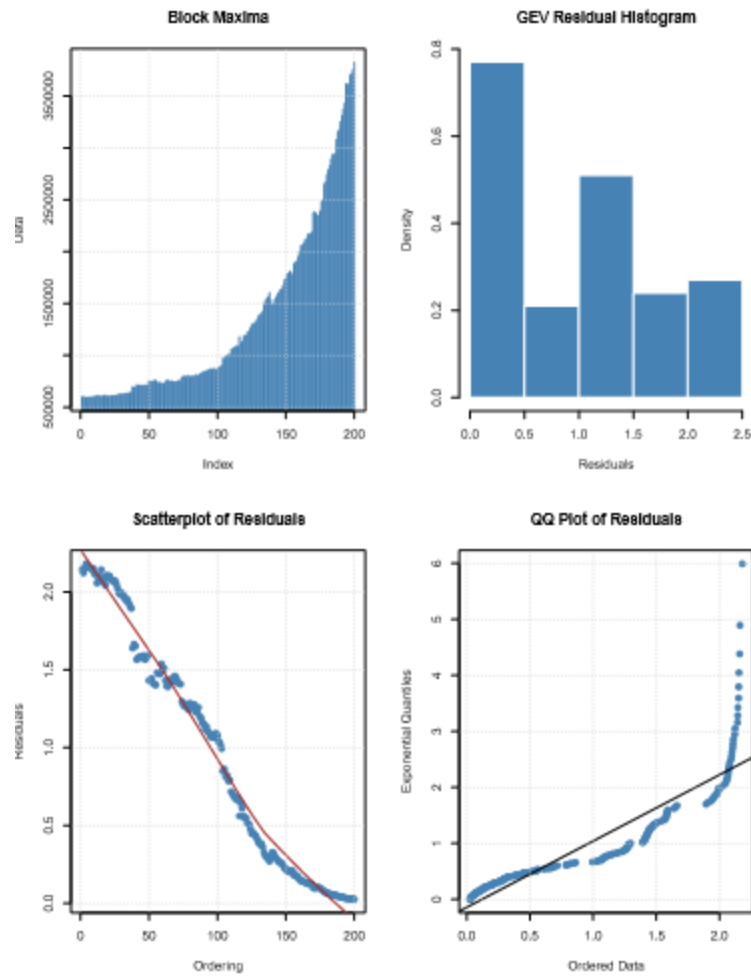


Figure 4.8: GEV fit to the the Kenya’s Total Debt suggesting a Gumbel model

Figure 4.8 shows that a Gumbel model would be a good GEV choice.

Table 4.1: Estimated parameters from GEV model of the Total Debts

ξ	$\hat{\mu}$	$\hat{\sigma}$
2.981141e-01	8.992533e+05	4.411188e+05

Table 4.1 show that Gumbel model is the best candidate for the Total Debts $\xi = 0.03$ which tends to zero. At a 95% threshold, as can be seen in Figure 4.8 the extreme debt level sustainable is KES 5 trillion. This shows that all factors growing at the predicted rate, based on past data. This projection explains why the despite the constant apprehension an the Kenyan debt levels, the economy has not yet experienced effects of unsustainable debt.

This provides room for more debt but the debt should be for development projects as it leads to higher growth and better sustainability figures. This projection is however dependent on the current debt structure. A panic by creditors will however reduce the maxima as debt restructuring and limited foreign reserves will have a limit repayment capabilities.

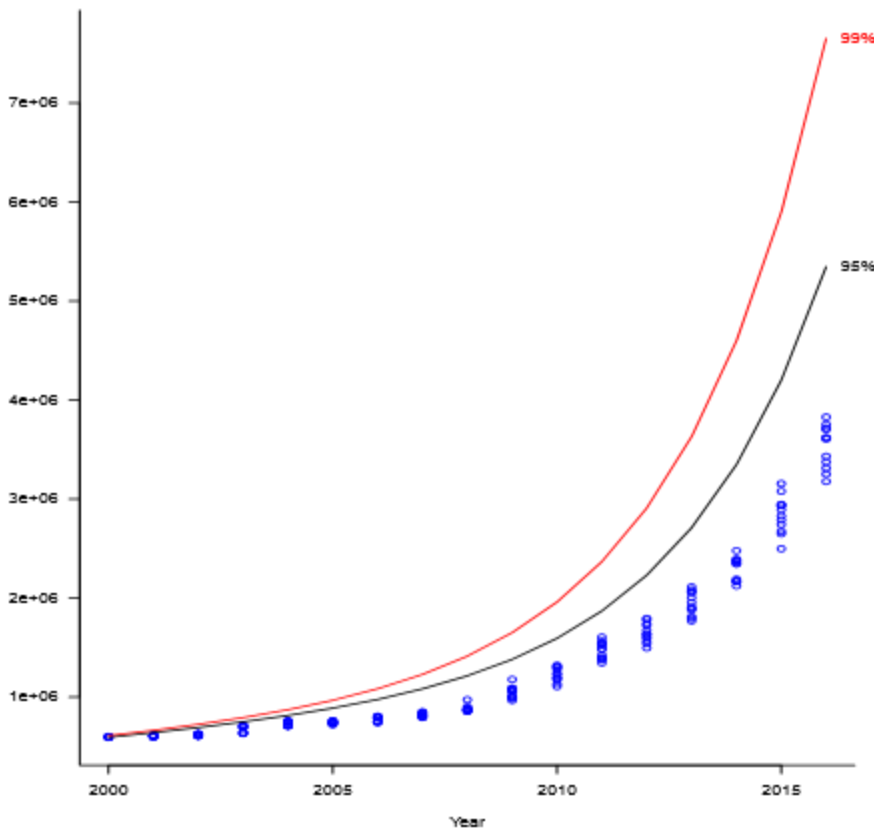


Figure 4.9: GEV projection of the Kenya's Total Debt showing a 95% threshold of 5 trillion shillings

As can be seen in Table 4.2 there is a positive relationship between total debt, total imports, total revenue and total expenditure

Table 4.2: Estimates of the regression coefficient from GEV model ($\xi = 0$)

Parameter	Estimate	Std. Error	Z value	$Pr(> z)$
(Intercept):1	1.180e+05	5.701e+04	2.069	0.0385
(Intercept):2	1.288e+01	5.138e-02	250.650	<2e-16
Total Imports USD	1.438e+01	8.261e-01	17.407	<2e-16
Total Revenue	2.514e-02	8.091e-01	0.031	0.9752
Total Expenditure	5.080e-01	5.788e-01	0.878	0.3802

Chapter 5: Conclusion

5. 1: Conclusion

It is evident that total debt increase is both at the domestic and external level as exhibited by their similar trend. Devolution caused by the new constitution in 2010 is the starting point for the rapid increase in total debt levels. Devolution caused development of several national services to county levels and raised the number of civil servant with new positions required to govern counties as shown by the positive correlation between total government debt and total government expenditure.

With concern on the debt levels coming from external bodies and raising food levels, there is a need to determine the sustainability of the rising levels. Using GEV, at a 95% threshold, the KES 5 trillion is the derived maxima of the debt levels. This shows that at the current level of KES 2.5 trillion, there is room to absorb more debt.

Despite evidence that Kenyan debt is still at sustainable levels, there are a couple of policy recommendations to ensure sustainability is possible. One is that, the use of the debt is a crucial factor as it influences the explanatory variables which have been proven to be positively correlated to each other. Thus debt for development expenditure is of greater benefits as it has an expected positive impact on explanatory variables. This falls in line with rise in recurrent expenditure. Rising recurrent expenditures have to positive economic impact as they do not lead to growth. Thus shifting expenditure from development expenditure to recurrent expenditure will negatively impact the maxima and lower it as growth variables decline.

Another policy factor is with regards to import tax rates. There is a positive correlation between imports and debt sustainability. Government has a past of raising taxes on imports to raise revenue, however, it tends to reduce import levels. Thus before raising tax levels, the government needs to carry out impact evaluation to ensure import levels do not go down to levels that reduce revenue.

In conclusion, although the maxima has been derived as KES 5 trillion, there are a lot assumptions used in the model and a violation of one of the variables due to unexpected macroeco

conomic or political upheaval and lack of adherence to the policies above could cause a change in the outcome.

5. 2: Further Research

There is a lot of research to be based from the findings of this paper. The first being on testing various models to find the best fitting model for this type of data. An example can be use of stress testing, GLM and so on to find the best fitting models or models with near similar outputs.

Secondly, researchers can also split the debt to their sources and interest charged. This will provide insight on debt that are more sustainable than others and influence each debt has on the sustainability levels.

The third can be to analyse the GEV factoring in the global economic climate. This study was limited to Kenya as an independent country with no external influence. However, it is known that the country economies are interlinked and so one country will impact the next and this will influence debt sustainability.

Finally, further research can look at sustainability with regards to the East Africa Community (EAC) union. With Kenya and other EAC planning on becoming a union with one currency and free trade, the impact of this on the debt sustainability in Kenya would be of interest to policy makers as the debt of each country has an impact on the other countries.

5. 3: Limitations of the Study

There are several limitations of the study. One of the main problems is limited time frame to do the analysis. The project was restricted to a limited time frame for submission thus there was a limitations on the analysis that could be carried out.

A second problem is data availability was limited for earlier years. The data available was restricted to the period between 2000 and 2016. This decreased the prediction power of variables as their pattern over time was restricted to the time period available.

A third limitation was interrelationship between countries was not factored into the model. In theory and practise, economies tend to be dependent on each other. However,

due to time and data constraints, this study was not able to factor the effect of various countries on Kenyan debt into the model.

Fourth is the use of the debt (such as development, loan servicing and so on) was not factored into the model. In practise, a debt used for development will have a positive impact on the country in the long run as opposed to loans taken to pay off other loans or for recurrent expenditure. The use of the loan will thus have an effect on the forecasted figures in the model.

The different maturity times of the debts was not factored into the model. The effect of debt maturity was not factored into the model although it influences the debt levels. The unavailability of the data was the main factor that led the information not to be included in the model.

Chapter 6: References

1. Burger, John, Warnock, Francis. 2006. Local currency bond markets. IMF Staff Papers, vol. 53, International Monetary Fund, Washington, D.C., pp. 133-146.
2. Chong Il-yo and Jun Chi-Hyuk. 2004. Performance of Variable Selection Methods when Multicollinearity is Present. *Chemometrics and Intelligent Laboratory Systems* 78 (2005) 103 - 112.
3. Diaz-Alejandro, Carlos. 1985. Good-Bye Financial Repression, Hello Financial Crash. *Journal of Development Economics*, 19(1-2): 1-24.
4. Diebold X. Francis, Schuermann Til and Stroughair D. John. 1998. Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management.
5. Eichengreen, Barry, Borensztein, Eduardo and Panizza, Ugo. 2006. A Tale of Two Markets: Bond Market Development in East Asia and Latin America. Hong Kong Institute for Monetary Research Occasional Paper No. 3.
6. Embrechts Paul, Resnick Sidney I . and Samorodnitsky Gennady . 1999. Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal*, Volume 3, Number 2.
7. Gençay Ramazan and Selçuk Faruk. 2004. Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of Forecasting* 20 (2004) 287 - 303.
8. Gilli Manfred and Këllezi Evis. 2003. An Application of Extreme Value Theory for Measuring Risk.
9. Kumar, Manmohan S. and Jaejoon Woo, 2010. Public Debt and Growth. IMF Working Paper 10/174 (Washington: International Monetary Fund).
10. Maana, I., Owino, R., and Mutai, N. 2008. Domestic Debt and its Impact on the Economy-The Case of Kenya. African Econometric Society Conference. Pretoria: Central Bank of Kenya.

11. Medford Anthony. 2015. Best Practice Life Expectancy: An Extreme Value Theory Approach. Max Planck Odense Centre on the Biodemography of Aging. University of Southern Denmark.
12. Pattillo, C., H. Poirson, and L.A. Ricci, 2002, "External Debt and Growth?" IMF Working Paper 02/69 (Washington: International Monetary Fund).
13. Reinhart, Carmen M., and Kenneth S. Rogoff. 2009a. This Time Is Different: Eight Centuries of Financial Folly. Princeton, NJ: Princeton University Press.
14. Renze, John and Weisstein, Eric W. Extreme Value Theorem. From MathWorld--A Wolfram Web Resource.
15. Republic of Kenya. 2007. Kenya Vision 2030-The Popular Version. Nairobi: Ministry of Planning and National Development.
16. Thugge, K., Heller, P.S., & Kiringai, J. 2012. Fiscal policy in Kenya: Looking toward the medium to Long-Term. Unpublished.
17. Sirengo, J. 2008. Determinants of Kenya fiscal performance. Kenya Institute for Public Policy Research and Analysis, Discussion Paper, No.91.
18. Stijn Claessens, Daniela Klingebiel, Sergio Schmuckler. 2007. Government Bonds In Domestic And Foreign Currency: The Role Of Macroeconomic And Institutional Factors. *Review of International Economics*, 15, pp. 370–413.
19. Velasco, Andres. 1987. "Financial Crises and Balance of Payments Crises: A Simple Model of the Southern Cone Experience." *Journal of Development Economics*, 27(1–2): 263–83.
20. Wafula, Martin. 2010. Debt Sustainability and the Optimal Debt in Kenya.
21. Were, Mercy. 2001. The Impact of External Debt on Economic Growth in Kenya. Nairobi: Kenya Institute for Public Policy and Research Analysis.
22. Wold S., Johansson E., Cocchi M. 1993. 3D QSAR in Drug Design; Theory, Methods, and Applications. ESCOM, Leiden, Holland, 1993, pp. 523 – 550.

Chapter 7: Appendix

The appendix contains the R-scripts used to come up with the output in Chapter 4.

7. 1: Install R-package

```
fitGEV <- gev.fit(Zmax)

gev.diag(fitGEV) #plot(fitGEV)

install.packages("fExtremes")

# read data

dat <- read.csv(file="C:/Work/tempo/JM/Proposal Data 2.csv")

names(dat)
```

7. 2: Exploratory Data Analysis

```
#exploratory data analysis

ts.obj <- ts(dat$Domestic.Debt..KES..Mn.,start = c(2000, 5), frequency = 12)

xx<-table(dat$Year)

rep(names(xx),xx)

# total debts

ts.plot(ts.obj,

        gpars=list(xlab="Year", ylab="Total Debt (KES 'Mn)"))

#overdraft

ts.plot(ts.obj,

        gpars=list(xlab="Year", ylab="CBWA Overdraft Rate" ) )
```

```
#debts
```

```
ts.obj.dom <- ts(dat$Domestic.Debt..KES..Mn.,start = c(2000, 5), frequency = 12)
```

```
ts.obj.ext <- ts(dat$External.Debt..KES..Mn.,start = c(2000, 5), frequency = 12)
```

```
ts.plot(ts.obj.dom,
```

```
    gpars=list(xlab="Year", ylab="Debts (KES 'Mn)'),col="blue")
```

```
    lines(ts.obj.ext,col="red")
```

```
#legend(2005,65,c("Domestic","External"),col=c("blue","red"),lty=1)
```

```
plot(dat[,c(4:5,9:11)])
```

```
#cor(dat[,c(4:5,9:11)])
```

```
#Scatter plot
```

```
cor(dat[, c("Total_Debt", "Total_Imports_USD", "Total_Revenue","Total_Expenditure")])
```

```
    Total_Debt Total_Imports_USD Total_Revenue Total_Expenditure
```

```
    library(fExtremes)
```

```
    head(dat[,c(4:5,9:11)])
```

7.3: GEV Output

```
#GEV parameter estimation
```

```
colnames(x) <- "Danish"
```

```
head(x)
```

```
## hillPlot -
```

```
# Hill plot of heavy-tailed Danish fire insurance data
```

```
lpar(mfrow = c(1, 1))
```

```
hillPlot(ts.obj, plottype = "xi")
```

```
grid()
```

```
#regression
```

```
vglm(formula = Total_Debt ~ Year, family = gumbel, data = dat,
```

```
trace = TRUE)
```

```
vglm(formula = Total_Debt ~ Total_Imports_USD + Total_Revenue +
```

```
Total_Expenditure, family = tobit, data = dat)
```