FREE MARKET FOUNDATION

PRIORITIES FOR GROWTH

5 POLICY AREAS TO MAXIMISE GDP PER CAPITA GROWTH IN A YEAR

Prepared by:

MPIYAKHE DHLAMINI
DATA SCIENTIST

Priorities for growth:
5 policy areas to maximise GDP per capita growth in a year

Mpiyakhe Dhlamini

5 February 2019

Executive summary

The Fraser Institute's Economic Freedom of the World (EFW) index is made up of five areas that are in turn made up of 43 component variables representing various data sources produced by organisations like the World Economic Forum, the World Bank, etc. The object of this paper was to find the five most relevant variables affecting GDP per capita growth over a one-year period.

The most important variables affecting GDP per capita growth over the following year in a country were found to be the following using regression analysis and in descending order of importance:

1. Inflation: Most recent year.
2. Gender adjustment.
3. Foreign ownership/investment restrictions.
4. Impartial courts.
5. Administrative requirements.

Prepared by:

Mpiyakhe Dhlamini
Data Scientist
Free Market Foundation

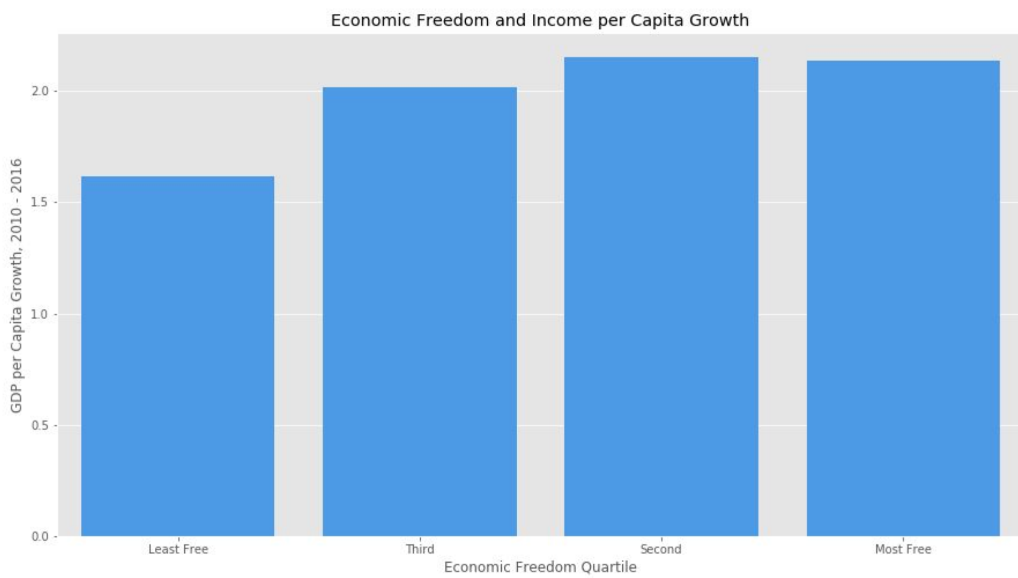
5 February 2019

Ⅰ. Introduction

"The index published in Economic Freedom of the World measures the degree to which the policies and institutions of countries are supportive of economic freedom. The cornerstones of economic freedom are personal choice, voluntary exchange, freedom to enter markets and compete, and security of the person and privately owned property. Forty-two data points are used to construct a summary index and to measure the degree of economic freedom in five broad areas."

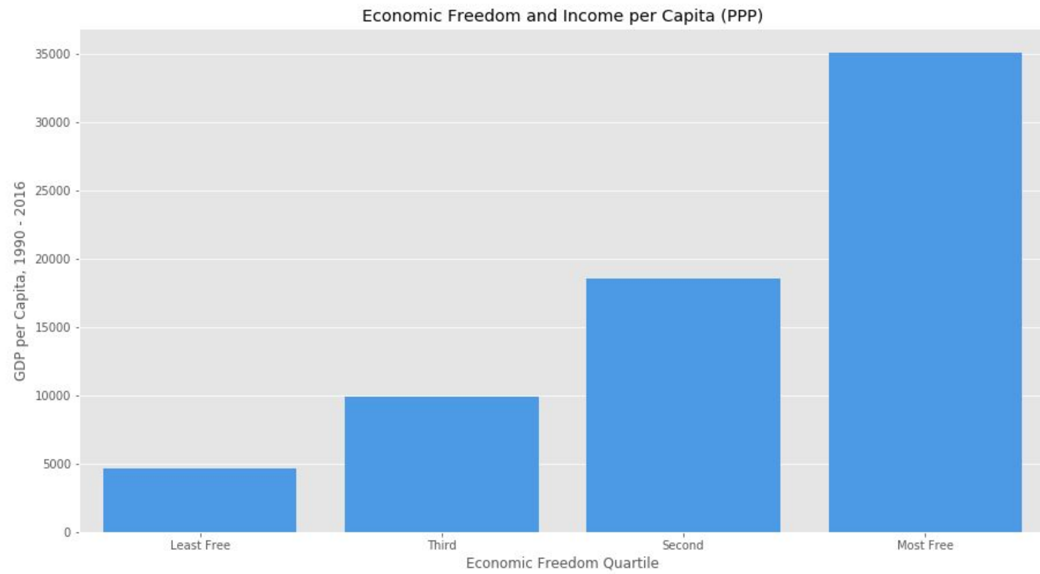
- James Gwartney, Robert Lawson, Joshua Hall, and Ryan Murphy (2018). *Economic Freedom of the World: 2018 Annual Report*. Fraser Institute.

The index seems to correlate positively with GDP per capita growth. This can be seen from the following figure:



We see from the above that the countries at the highest quartiles of economic freedom enjoy a higher growth in their GDP per capita.

There seems to be an even stronger correlation with GDP per capita itself:



Data sources for these figures: Fraser Institute EFW dataset,[1] World Bank GDP per capita growth,[2] and World Bank GDP per capita in purchasing power parity terms using 2011 USD.[3]

This apparent correlation motivates the question at the heart of this study: Can we produce a set of five policy recommendations that government can immediately undertake to maximise the chances of growing the GDP per capita in the shortest time possible?

We believe it is possible to do this if we focus on extracting the most relevant variables in the EFW for GDP per capita growth.

Overview of methodology

The approach taken in this paper is to first select the independent variables (from the EFW dataset) that explain as much of the variance in the dependent variable (GDP per capita growth) as possible, and then use the chosen variables to perform a linear regression.

The top five variables in terms of the value of their coefficients (positive) in the resulting regression equation are then taken to be the most relevant, and from these we produce a set of recommendations as to what policy-makers in South Africa should be doing if their priority is increasing GDP per capita growth. We do this using GDP per capita growth instead of GDP growth, as this has the useful property that population effects are already built-in.

1 https://www.fraserinstitute.org/studies/economic-freedom-of-the-world-2018-annual-report
2 https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG
3 https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD

The primary tool is the python programming language and elements of its data science stack (libraries for processing and analysing data), the language version is 3.6 and the code was run on a Jupyter Notebook (Appendix I). All of the imported python libraries are available as part of the Anaconda distribution from Continuum Analytics (the exception being a function for cleaning World Bank data, which I wrote myself and which is attached as Appendix II).

In the following sections we describe: Data preparation, scoring metrics, feature selection, building the regression model and recommendations for South African policy makers. The last section is the conclusion.

Ⅱ. Data preparation

(a) Nulls

The two data sources used to prepare this report are the *Economic Freedom of the World* dataset (1970-2016) and the World Bank GDP per capita growth dataset (1961-2017). The data was imported using the pandas library as dataframes. The EFW data initially had 3726 rows but a significant portion of these had nulls. To give one example, the reliability of police column had only 1655 out of 3726 rows being non-null. After merging the processed GDP per capita growth and EFW data (based on country names and years) as well as dropping any rows with nulls, we were left with only 1047 rows in the combined dataset.

(b) Processing World Bank data

A custom function was used to clean the World Bank data. This function works for most World Bank country data and consists of extracting years, countries and creating a data column using a passed in label. The function returns a pandas dataframe.

(c) Final processing

After removing nulls and completing all necessary processing steps including merging the two dataframes. The merge was based on year and country, the year was offset by one between EFW and GDP per capita growth data in order to match EFW data for a particular year to the following year's GDP per capita growth. We had one table with both EFW data and GDP per capita growth, we then dropped all variables not needed for the regression analysis including the year variables, countries, Summary Index from the EFW data, etc. See Part 0 of Appendix I, the jupyter notebook.

We then randomly shuffle the dataframe.The range of years covered by the combined dataset after processing is 2005-2016 inclusive.

Also note that the gender adjustment variable was multiplied by 10 in order to put it on the same scale as all the other variables.

Ⅲ. Scoring metrics

Metrics are vital in performing a regression analysis or any task that falls under the category of machine/statistical learning. In the course of building a statistical model – even if the purpose is determining variable importance as in our case -- choices around which variables to include, which model would be best, etc., have to be made and therefore we need some robust metric(s) to guide us as we build the model.

We have decided to use two metrics: The r^2 score for finding the optimal set of variables and the MSE (mean squared error) both for choosing the best variable selection procedures (discussed in more detail in the feature selection section) as well as choosing the best linear regression model (discussed in the building the regression model section).

A brief discussion of each metric follows:

(a) R^2 score

Also known as the coefficient of determination, it denotes the percentage of the variance in the dependent variable that is predictable from the independent variables. It is calculated as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where $SS_{res} = \sum_i (y_i - f_i)^2$, the residual sum of squares which sums the squared differences between the values of the dependent variable in the data (y_i) and the values predicted by the regression equation (f_i). $SS_{tot} = \sum_i (y_i - \underline{y})^2$ the total sum of squares which is proportional to the variance of y and is the sum of the squared differences between the values of the dependent variable in the data and the mean of that variable.

(b) Mean squared error (MSE)

The MSE is calculated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - f_i)^2$$

It is the sum of the squared differences between the actual sampled values and predicted values. For our purposes, we used a cross-validated MSE implemented by means of a custom python function, more on that in a later section.

Ⅳ. Feature selection

In order to get the set of EFW variables with the most predictive value on GDP per capita growth, we chose to employ three separate variable selection methods. Each of the chosen methods used the R^2 score as a criterion for choosing which variables to select/reject.

The three variable selection methods include forward and backward selection (both under the step-wise regression family of variable selection methods) and recursive feature elimination. All of the methods were implemented using cross-validation, i.e. we used k-fold cross-validation with five folds in each case, we then selected the variables with the best mean performance over all five folds.

A brief description of each method follows:

(a) Forward selection

Part of the step-wise regression family of variable selection methods, forward selection involves adding variables to the regression model one-by-one based on which variable improves the chosen criteria the most (we use the R^2 score as criteria for all three methods). We do this until the cross-validated mean of the R^2 no longer improves when we add new variables. Our implementation of cross-validated forward selection chose 24 out of the 43 EFW variables.

These are the chosen variables:

1. Government consumption
2. Transfers and subsidies
3. Government enterprises and investment
4. Judicial independence
5. Impartial courts
6. Protection of property rights
7. Gender adjustment
8. Money growth
9. Standard deviation of inflation
10. Inflation: Most recent year
11. Freedom to own foreign currency bank accounts.
12. Standard deviation of tariff rates

13. Non-tariff trade barriers
14. Foreign ownership/investment restrictions
15. Freedom of foreigners to visit
16. Ownership of banks
17. Private sector credit
18. Hiring and firing regulations
19. Centralized collective bargaining
20. Hours regulations
21. Conscription
22. Administrative requirements
23. Bureaucracy costs
24. Licensing restrictions

(b) Backward selection

This is another step-wise regression method of variable selection. It involves starting out with all the model variables and testing which variable's deletion causes the smallest statistically insignificant deterioration of fit based on our criterion. This is done until no variable can be found for which deletion causes a statistically insignificant deterioration of fit in the linear regression model. Like with all the other methods, we used cross-validated backward selection.

Here are the chosen variables:

1. Government consumption
2. Transfers and subsidies
3. Top marginal income tax rate
4. Top marginal income and payroll tax rate
5. Impartial courts
6. Protection of property rights
7. Integrity of the legal system
8. Gender adjustment
9. Standard deviation of inflation
10. Inflation: Most recent year
11. Non-tariff trade barriers
12. Foreign ownership/investment restrictions
13. Capital controls
14. Freedom of foreigners to visit
15. Controls of the movement of capital and people
16. Ownership of banks
17. Private sector credit
18. Hiring and firing regulations
19. Hours regulations
20. Administrative requirements

21. Bureaucracy costs
22. Licensing restrictions

(c) Recursive feature elimination

Recursive feature elimination involves training a model with an initial basket of all the features, the assigned coefficients of the features are then used to recursively eliminate the least important features. Using cross-validation with 5 folds, we can converge to an optimal set of features. The criteria used to select the optimal set is the R^2 score as in all the other cases.

These are the chosen variables:

1. Government consumption
2. Transfers and subsidies
3. Top marginal income tax rate
4. Top marginal income and payroll tax rate
5. Judicial independence
6. Impartial courts
7. Protection of property rights
8. Reliability of police
9. Business costs of crime
10. Gender adjustment
11. Legal system and property rights
12. Money growth
13. Standard deviation of inflation
14. Inflation: Most recent year
15. Mean tariff rate
16. Standard deviation of tariff rates
17. Non-tariff trade barriers
18. Black market exchange rates
19. Foreign ownership/investment restrictions
20. Controls of the movement of capital and people
21. Ownership of banks
22. Private sector credit
23. Interest rate controls/negative real interest rates)
24. Hiring and firing regulations
25. Centralized collective bargaining
26. Hours regulations
27. Administrative requirements
28. Bureaucracy costs
29. Starting a business
30. Licensing restrictions

(d) Choosing the best set of variables.

We then tested four different linear models to determine which set of variables produced the best results against our cross-validated MSE criteria. The four possibilities were all simple linear models, but the variables chosen differed in each case. The first model was a linear model built using all the variables. This was done to provide a baseline/control against which all the models could be compared as well as comparing them to each other.

Here are the results:

Model (#variables)	**Median cross-validated MSE**
All features (43)	11.37
Forward selection features (24)	10.80
Backward selection features (22)	11.20
Recursive feature elimination (30)	10.82

As we can see, forward selection provides the best combination of features using our selected criteria. Backwards elimination provides the worst combination. We move on to building the model.

V. Building the regression model

Among the last choices we have to make before we can have a final model, is the choice of whether to use regularisation or not. If we make use of regularisation we then have to choose which type of regularisation to use between l1 (lasso regularisation), l2 (ridge regularisation) and/or some combination of the two (elastic net).

First, a brief explanation of what regularisation is.

Regularisation is a means of preventing overfitting when building a regression model by penalising larger weight values during training. The ridge and lasso methods accomplish this goal differently.

(a) Lasso (l1) regularisation

From the scikit-learn website[4]:

"The Lasso is a linear model that estimates sparse coefficients. It is useful in some contexts due to its tendency to prefer solutions with fewer parameter values, effectively reducing the number of variables upon which the given solution is dependent. For this reason, the

[4] https://scikit-learn.org/stable/modules/linear_model.html#lasso

Lasso and its variants are fundamental to the field of compressed sensing. Under certain conditions, it can recover the exact set of non-zero weights.

Mathematically, it consists of a linear model trained with ℓ1 prior as regularizer. The objective function to minimize is:

$$\min_{w} \frac{1}{2n_{samples}} ||Xw - y||_2^2 + \alpha ||w||_1$$

The lasso estimate thus solves the minimization of the least-squares penalty with $\alpha||w||_1$ added, where α is a constant and $||w||_1$ is the ℓ1-norm of the parameter vector."

(b) Ridge (l2) regularisation

The scikit-learn website[5] again:

"Ridge regression addresses some of the problems of Ordinary Least Squares by imposing a penalty on the size of coefficients. The ridge coefficients minimize a penalized residual sum of squares,

$$\min_{w} ||Xw - y||_2{}^2 + \alpha ||w||_2{}^2$$

Here, $\alpha \geq 0$ is a complexity parameter that controls the amount of shrinkage: the larger the value of α, the greater the amount of shrinkage and thus the coefficients become more robust to collinearity."

(c) Elastic net regularisation

Elastic net offers a balance between ridge and lasso regularisation, this from the scikit-learn website[6]:

"ElasticNet is a linear regression model trained with L1 and L2 prior as regularizer. This combination allows for learning a sparse model where few of the weights are non-zero like Lasso, while still maintaining the regularization properties of Ridge. We control the convex combination of L1 and L2 using the l1_ratio parameter.

Elastic-net is useful when there are multiple features which are correlated with one another. Lasso is likely to pick one of these at random, while elastic-net is likely to pick both.

A practical advantage of trading-off between Lasso and Ridge is it allows Elastic-Net to inherit some of Ridge's stability under rotation.

[5] https://scikit-learn.org/stable/modules/linear_model.html#ridge-regression
[6] https://scikit-learn.org/stable/modules/linear_model.html#elastic-net

The objective function to minimize is in this case:

$$\min_{w} \frac{1}{2n_{samples}} ||Xw - y||_2^2 + \alpha\rho||w||_1 + \frac{\alpha(1-\rho)}{2} ||w||_2^2,,$$

We compared the cross-validated MSE scores for four models: no regularisation, lasso, ridge and finally elastic net regularisation:

Model	**Cross-validated MSE**
No regularisation	10.80
Lasso (l1) regularisation	10.78
Ridge (l2) regularisation	10.74
Elastic net regularisation	10.74

We conclude that elastic net offers the best performance. The chosen l1 ratio is 0.1.

(d) Final model

We can now build the final model. This is the equation with only the top five variables (rounded off to two decimal places):

$$GDP\ per\ capita\ growth\ =\ 0.87\ (Inflation:\ Most\ recent\ year)\ +\ 0.48\ (Gender\ djustment)$$
$$+\ 0.45\ (Foreign\ ownership/investment\ restrictions)\ +\ 0.37\ (Impartial\ courts)$$
$$+\ 0.21\ (Administrative\ requirements)\ \ldots$$

Interestingly, **the most recent value for inflation** is almost twice as important as the next variable, meaning that a 5-percentage point drop in the inflation (CPI) rate produces, on average, a 0.87 percentage point increase in the GDP per capita rate of growth.

What about **the impact of treating women as equal citizens**? The difference between discriminating on the basis of sex in the law and not doing so can be as much 4.8 percentage points in GDP per capita growth on average (the gender adjustment variable was scaled to the range 0-10 like all the other variables). While these are only averages across many countries, it is clear that restricting the productive capacity of half your population is not a good idea.

The **foreign ownership/investment restrictions** EFW variable comes from the World Economic Forum’s *Global Competitiveness Report*, in particular two survey questions that are used to 1) gauge the prevalence of foreign ownership and 2) a country’s restriction of international capital flows. Treating foreign investors as favourably as domestic investors is therefore of the utmost importance. This also includes allowing investors and individuals in their personal capacity to be able to take money overseas without any onerous regulations.

The **impartial courts** EFW variable comes from the WEF's *Global Competitiveness Report* survey question: "The legal framework in your country for private business to settle disputes and challenge the legality of govt actions and/or regulations is inefficient and subject to manipulation?" It is clear therefore that improving the perception of all investors with regards to the legal system is of the utmost importance. It is a necessity if living standards are to improve.

The **administrative requirements** sub-component is based on the *Global Competitiveness Report* question: "Complying with administrative requirements (permits, regulations, reporting) issued by the government in your country is (1 = burdensome, 7 = not burdensome)?" Removing unnecessary reporting requirements could, on average, add as much as 2.1 percentage points to a country's GDP per capita all else being equal.

Ⅵ. Policy recommendations

This section makes recommendations on specific policy steps that can be taken by policy-makers to maximise the probability of increasing GDP per capita growth over the short term. We make five specific recommendations that members of the executive, legislature and judiciary can follow:

(a) Remove all unnecessary administrative requirements

South Africa's score on this sub-component is 3.64, making us 98th in the world. The top 5 countries are Singapore (7.69), the United Arab Emirates (7.25), Rwanda (7.14), Hong Kong (7.13) and Malaysia (6.39). Since this is based on the opinions of business executives, government should act urgently to remove any and all unnecessary compliance costs, reporting requirements, etc. Just as an indicator of the difference between Rwanda and SA, the GDP per capita growth in Rwanda was 3.57% in 2017 while South African GDP per capita growth was 0.07% during the same year.

Position	**Country**	**GDP per capita growth (%)**
1	Singapore	3.53
2	UAE	-0.61
3	Rwanda	3.57
4	Hong Kong	3.02
5	Malaysia	4.44

98	South Africa	0.07

Only the UAE experienced GDP per capita growth of less than 3% out of all the countries in the top 5 of the world in terms of this variable.

(b) Impartial courts

Investors need to know that they are protected from property expropriations, contract disputes as well as arbitrary abuse by government officials. In other words, investors need the rule of law (using the most meaningful sense of the term) in order to risk their money in any jurisdiction. South Africa is currently 34th in the world on this matric at 5.45. A table of the top 5 countries as well as South Africa, including their GDP per capita growth in 2017 is given below:

Position	**Country**	**GDP per capita growth (%)**
1	Finland	2.34
2	Switzerland	-0.02
3	Hong Kong	3.02
4	New Zealand	0.86
5	Singapore	3.53
34	South Africa	0.07

Keeping in mind that all these territories are more economically developed than South Africa, it is quite remarkable that only Switzerland had a lesser GDP per capita growth rate than we had.

(c) Foreign ownership/investment restrictions

Investors need the assurance that they will be able to take their money out of a country whenever they want. This is because they usually invest for profit and one should be able to enjoy the fruits of their labour. South Africa currently has a score of 6.00, ranked 74th in the world in terms of this variable.

Position	**Country**	**GDP per capita**
1	Singapore	3.53
2	Hong Kong	3.02
3	United Kingdom	1.13

4	Luxembourg	-0.68
5	Ireland	6.50
74	South Africa	0.07

It is of the utmost imperative to abolish exchange controls and make it as easy as possible for investors to put their money in South Africa. This includes reactivating the bilateral investor protection treaties and signing them with as many countries as possible. Open and free economies are rewarded with prosperity.

(d) Gender adjustment

It is pleasing that SA (score: 1.00, the maximum) is one of only 48 countries in which women and men are treated equally before the law. This is not to say that there are no challenges facing women, only that government does not pass laws oppressing women. The South African government has done well in this area and nothing should be changed. Of course, law enforcement could do much better in dealing with violence against women as well as protecting their property rights in rural areas where unwritten laws of male primogeniture still apply. If the property rights of rural women were sufficiently protected, it is likely that rural economies would benefit immensely.

(e) Inflation: Most recent year

This is based on the most recent value for the CPI. If we take SA's latest available CPI number (4.5%) and apply the necessary formula: $(\frac{50-4.5}{50-0})\times 10$, we get 9.1 for our ranking, this shows that the South African Reserve Bank has done a decent job of controlling inflation but it still only puts us in the 114th position in terms of the rest of the world.

Economist Dawie Roodt recently suggested[7] strengthening the Reserve Bank's independence and lowering the inflation target from the current 3-6% to 0-2%. If this were done, and assuming that CPI was exactly 2%, South Africa's score for this variable would be 9.6. This would put us in the 84th position in the world in terms of this variable.

Position	**Country**	**GDP per capita growth**
1	Ireland	6.50
2	Costa Rica	2.17
3	Slovenia	4.91

[7] https://businesstech.co.za/news/finance/295632/28-things-dawie-roodt-would-change-in-south-africa-if-he-were-president/

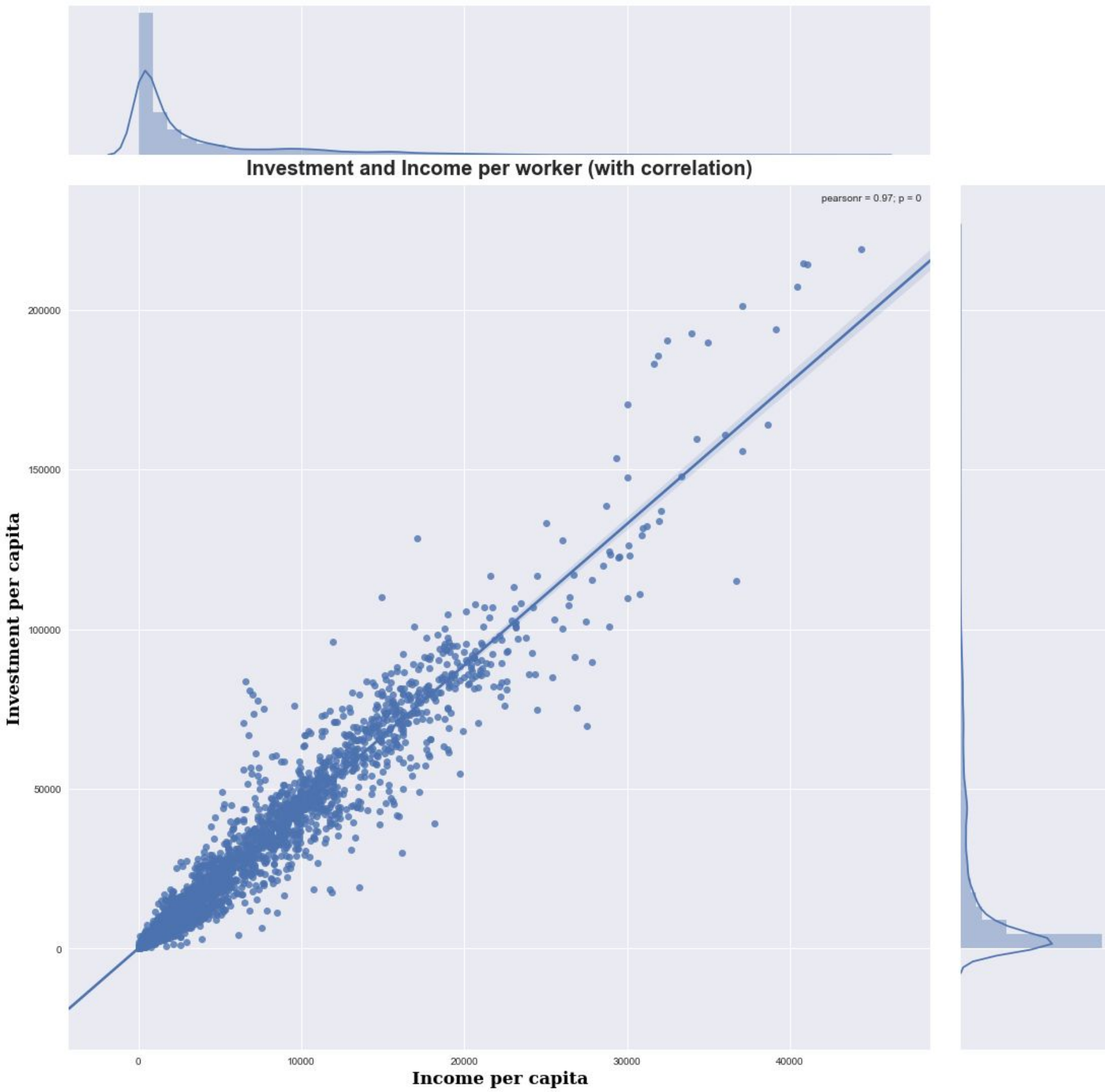
4	Barbados	1.41
5	Japan	1.88
124	South Africa	0.07

Inflation erodes the capital available to drive growth by devaluing savings that have been, in some cases, stored up over decades. Without capital formation, there can be no growth and there can be no capital formation without savings. Money that loses its value is money that is not attractive for denominating your savings in. Inflation incentivises people to spend instead of saving. This has a devastating effect on economic growth.

It is therefore not surprising that the variable quantifying price inflation would be the strongest predictor for GDP per capita growth. Inflation erodes away standards of living. Proposals to nationalise the Reserve Bank, while having little practical effect, would signal to currency traders and South African government bondholders that the government has an intention to interfere with the SARB's independence. This would have an inflationary effect on pensioners' savings denominated in rands as well as affecting the value of the bonds those same pensioners invested in, thinking they were safe investments.

This is not to mention the other proposal to change the SARB's mandate to be "pro-growth". Being pro-growth means preserving the value of the money people use for savings, and this is already a part of SARB's mandate. Anything extra will just be an invitation for more inflation and the subsequent erosion of savings and thus decreased economic growth.

Savers would be in for a particularly dismal time when you also consider the possibility of prescribed assets as mooted by some in the ruling party. There's a triple entendre of bad ideas that directly threaten South Africa's savings base and therefore the capital formation needed for economic growth.



The figure above shows the correlation between capital formation and employee incomes, the Pearson correlation coefficient is 0.97 (maximum possible is 1.00) at the 1% significance level. This means that when inflation threatens savings and therefore capital formation, worker productivity suffers as the necessary capital investments are not made leading to lower incomes for those workers. Members of COSATU should realise that a low-inflation policy is in their best interests and prevail upon the union leaders to abandon talk of a loose monetary policy.

Ⅶ. Conclusion

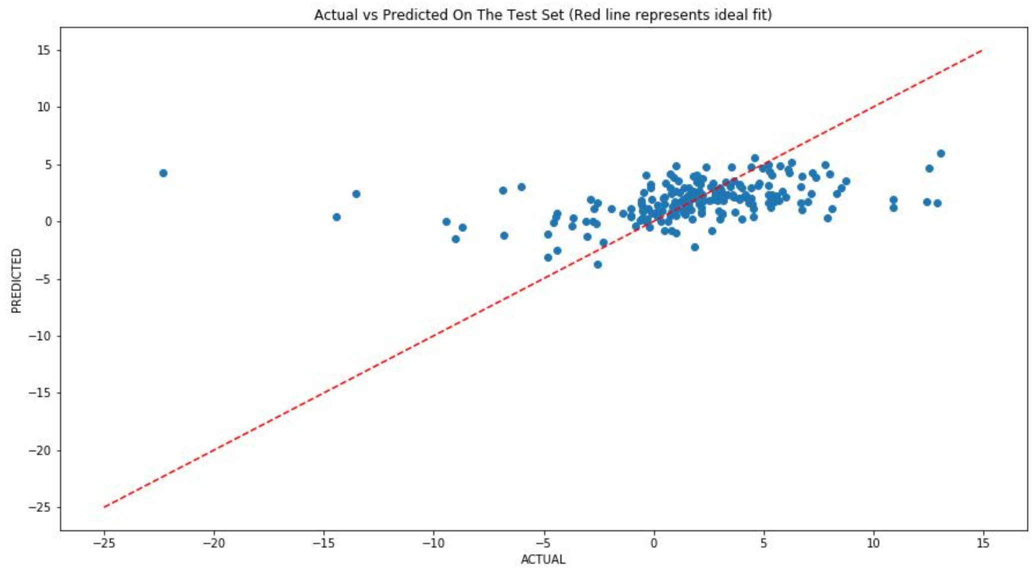
South Africa has been in an extended period of GDP per capita decline since at least the 2008 global financial crisis. Policy-makers have tried various measures to get the economy back on track; most have been dismal failures. Economic growth and job creation targets have been missed, lowered and then missed again. We were supposed to have GDP growth of 5%[8] in 2019 and a 14% unemployment rate by the year 2020. Instead, the economy only managed a weak 1.3% in 2017 and the official unemployment rate now sits at 27.5%.

Politicians seem to understand that they will pay no price for being wrong and do not seem overly concerned with solving very urgent economic problems, many of which are inflicted on the economy by the same government. This paper is an attempt to help policy-makers by identifying the five priorities for reform if the government wants the standard of living of South Africans to improve and do so as quickly as possible.

It might be worthwhile in future to perform the same type of research in attempting to find what the priorities should be for the quickest medium to long-term improvements. The fact that the current project is limited only to a year necessarily reduces the predictive accuracy of the regression and it should not be used for predictions; rather, the goal is to get five measures that the South African government could implement immediately for the greatest chance of short-term economic progress.

To reinforce the point made in the last paragraph, the following figure is a plot of predicted GDP per capita growth rates and the actual growth rates on a held-out set (the 20% of the data that was set aside and not used to train the model). The dashed red line shows where the points would lie if the regression was 100% accurate:

[8] https://www.fin24.com/Economy/South-Africa/Radebe-5-economic-growth-within-5-years-20140807



Therefore, and to conclude, the measures proposed in this paper would likely be helpful to government in the short-term, but if government wants sustained improvement and progress, it would also need to invest more effort in improving South Africa's overall economic freedom.

Prepared by:

Mpiyakhe Dhlamini
Data Scientist

With assistance from:

Martin van Staden
Legal Researcher

Date: 5 February 2019