

# Analysis of The World Happiness

Group 21-Ananth Padakannaya,Nivedita Patil,Li Wang,Wanqing Yang,Boyao Ma

2021/7/02

## Introduction

Happiness of an individual depends on multiple factors and the surrounding environment plays a key role in determining it. This survey on world happiness gives an insight into how happiness changes in our lives. Covid-19 has an adverse impact on our lives and it brings to attention the people's subjective well-being in the world at the moment and the factors affecting it. "As the pandemic struck, there was a large and immediate decline in mental health in many countries worldwide"(Helliwell et al. 2021). Therefore, it seems necessary to give more importance to happiness and well-being, thereby achieving better social and economic development (Greyling, Rossouw, and Adhikari 2021). Our study mainly aims to serve two purposes. Firstly ,we aim to find and evaluate the influential factors of happiness. Secondly, we aim to trace the size and distribution of happiness globally. We adopt the multiple regression analysis based on the data from the Gallup World Poll. There are six variables to measure happiness scores of 149 countries or territories. After processing the data, we finally get three influential factors.

## Exploratory Data Analysis

At first, we choose seven variables with one outcome variable, Happy score and six explanatory variables. They are as follows: Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity and Perceptions of corruption. The following is the interpretation of each of these variables in the data observed.

Happy score (score): respondents were asked to rate their possible lives from one to ten, with 1 being the worst and 10 being the best.

Logged GDP per capita (LoggedGDP): it denotes the logged GDP index from 149 countries.

Social support (Social): it denotes if the participants could find someone's help when they are in trouble.

Healthy life expectancy (expectancy): it denotes the expected life expectancy.

Freedom to make life choices (Freedom): it denotes if the participants are satisfied with their freedom to choose what they do in life.

Generosity: it denotes if the participants are generous after asking if they have donated money to charity organizations.

Perceptions of corruption (corruption): it denotes if the participants are conscious of the corruption within the government and business.

We visualize the values as boxplots:

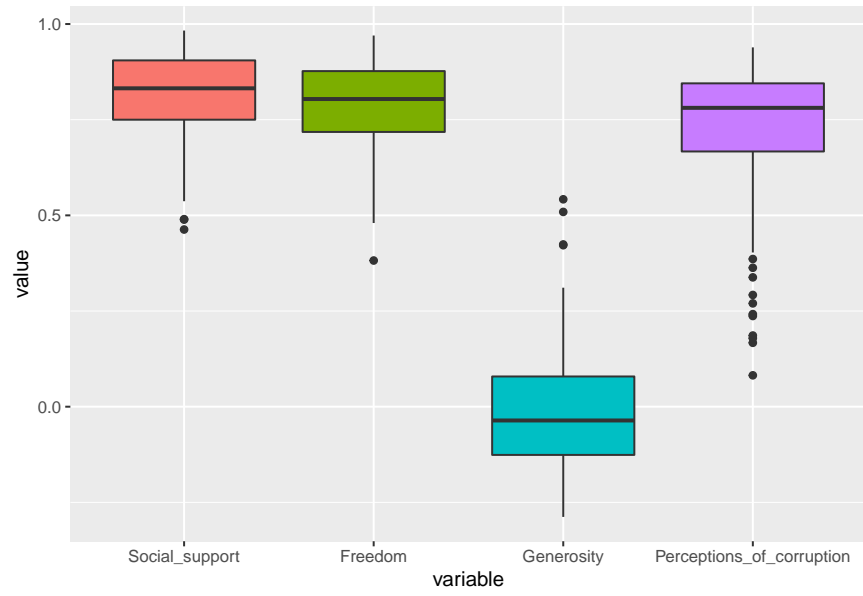


Figure 1: Boxplots of six variables.

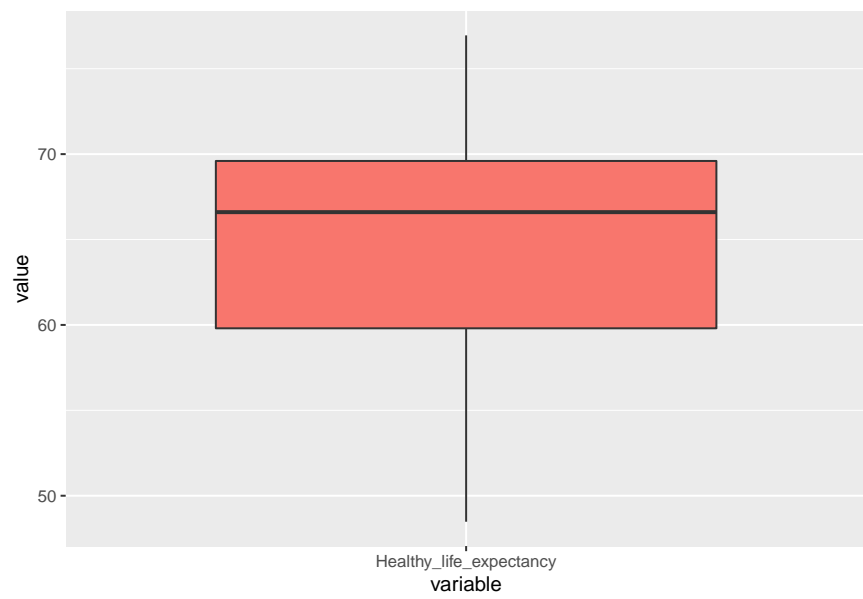


Figure 2: Boxplots of six variables.

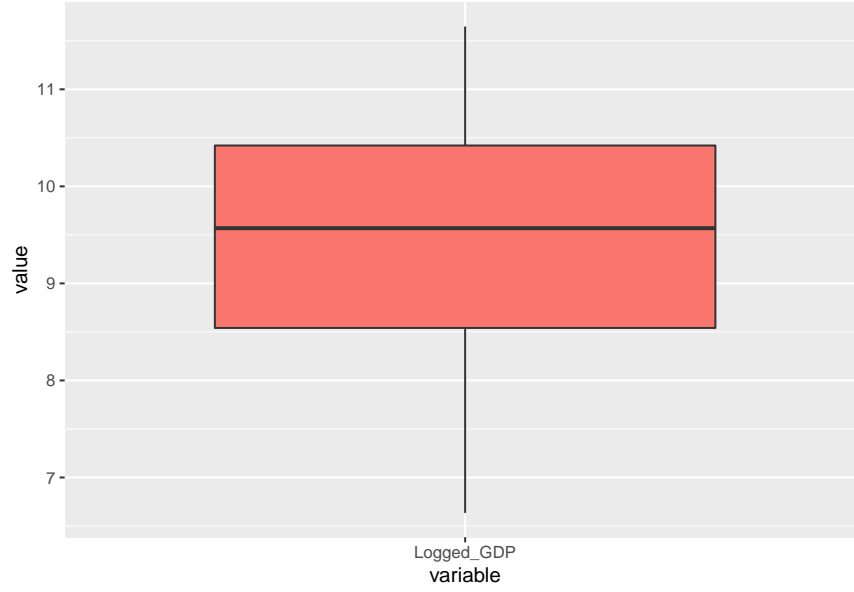


Figure 3: Boxplots of six variables.

Figure 1 shows the descriptive statistics of four variables - Social support, Freedom, Generosity and Perceptions of corruption. It is obvious that apart from Generosity, people in more than half of the countries grade more than 0.6 on Social support, Freedom and Perceptions of corruption although there are some outliers lower than 0.5. And figure 2 shows half of the population has a life span of 60 to 70 years. The figure 3 shows the median of Logged GDP is near 9.5.

Next we use table to illustrate the details of all seven variables.

Table 1: Summary statistics for observations with chosen variables.

	score	LoggedGDP	Social	expectancy	Freedom	Generosity	corruption
	Min. :2.523	Min. : 6.635	Min. :0.4630	Min. :48.48	Min. :0.3820	Min. :-0.28800	Min. :0.0820
	1st Qu.:4.852	1st Qu.: 8.541	1st Qu.:0.7500	1st Qu.:59.80	1st Qu.:0.7180	1st Qu.:0.12600	1st Qu.:0.6670
	Median :5.534	Median : 9.569	Median :0.8320	Median :66.60	Median :0.8040	Median :-0.03600	Median :0.7810
	Mean :5.533	Mean : 9.432	Mean :0.8147	Mean :64.99	Mean :0.7916	Mean :-0.01513	Mean :0.7274
	3rd Qu.:6.255	3rd Qu.:10.421	3rd Qu.:0.9050	3rd Qu.:69.60	3rd Qu.:0.8770	3rd Qu.: 0.07900	3rd Qu.:0.8450
	Max. :7.842	Max. :11.647	Max. :0.9830	Max. :76.95	Max. :0.9700	Max. : 0.54200	Max. :0.9390

Table 1 describes some basic features of those variables. For example, the average happiness score is 5.533 and the median is 5.534. It demonstrates that people's feelings of happiness or subjective well-beings are neither high nor low.

## Visualization of the data

We can visualize our data by producing histogram and radarplot, whereby taking note that we have several countries to compare, we shall plot the bars or lines using different colours for each country:

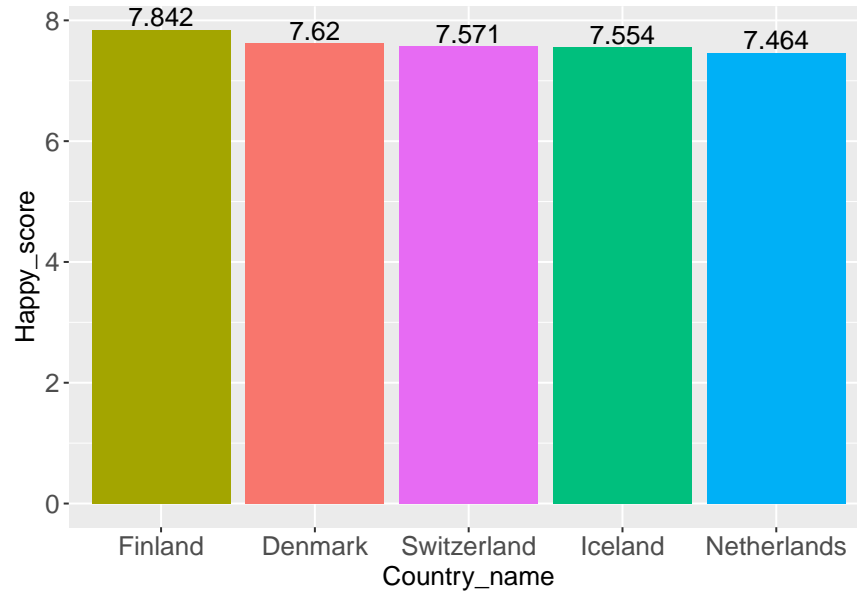


Figure 4: The 5 happiest countries.

Figure 4 shows the top 5 happiest countries. Finland gets the highest happy score with 7.842, followed by Denmark at 7.62. In addition, the differences between those five countries are small, less than 0.4, and all scores are greater than 7.4.

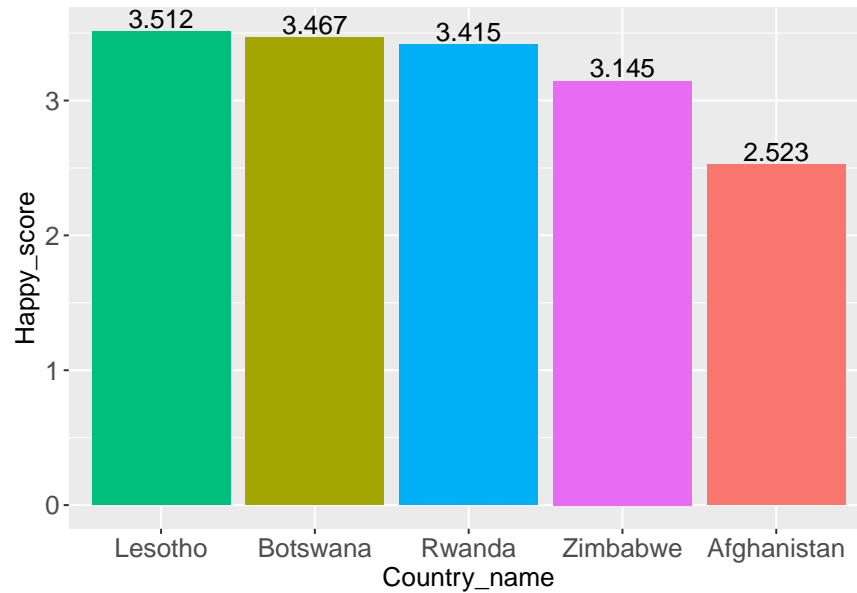


Figure 5: The 5 most unhappy countries.

Figure 5 shows the 5 most unhappy countries. Afghanistan takes the position as the most unhappy country with the score of 2.523. And the other four countries have similar scores lower than 3.6. Other than Afghanistan from South Asia, other four countries are in Africa.

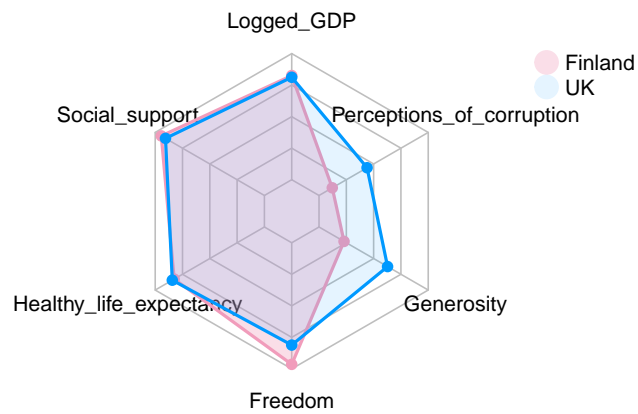


Figure 6: Comparisons between Finland and the UK from six indexes.

Figure 6 compares those six dimensions that Finland obtains with those that UK obtains. There are obvious differences between the perceptions of corruption and generosity, with the UK getting comparatively much higher values. Conversely, when referring to Freedom indicators, Finland has a slightly higher value. Besides, it owns similar grades among the other three aspects.

World Happiness Report 2021

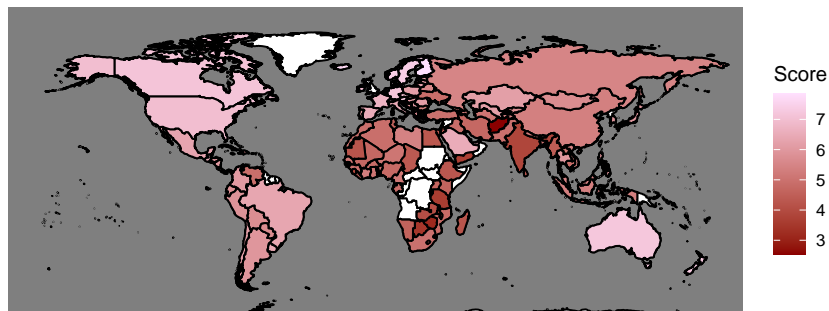


Figure 7: World happiness score map.

Figure 7 has displayed a happiness score map globally based on the World Happiness Report 2021. We can

see that most countries in the western area have high scores whereas the eastern countries have relatively low scores, which demonstrates that people in western countries lead happier and more fulfilling lives.

## Formal Data Analysis

To begin the analysis on the world happiness dataset, we need to check the correlation between the six explanatory variables to avoid the problem of multicollinearity.

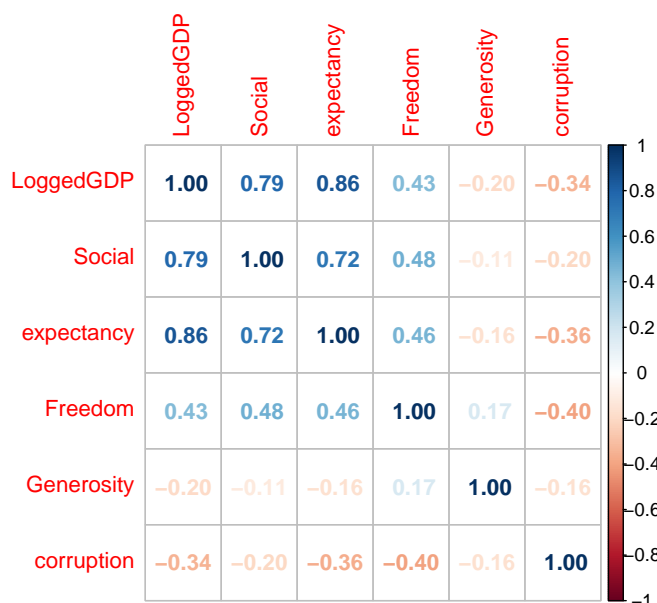


Figure 8: Correlation coefficient table.

From our correlation table we can see that the correlation between our Logged GDP and Healthy life expectancy is 0.859, which is a strong positive linear relationship. And the Logged GDP and Social support also have the high degree of collinearity. The correlation between this two variables is 0.785. So we remove Healthy life expectancy and Logged GDP. We then use the remaining four explanatory variables to perform stepwise regression, and observe whether the remaining variables need to be eliminated.

```
Start: AIC=-141.88
score ~ Social + Freedom + Generosity + corruption
```

	Df	Sum of Sq	RSS	AIC
- Generosity	1	0.091	53.856	-143.628
<none>			53.765	-141.879
- corruption	1	6.078	59.843	-127.920
- Freedom	1	6.364	60.129	-127.210
- Social	1	43.953	97.718	-54.857

```
Step: AIC=-143.63
score ~ Social + Freedom + corruption
```

	Df	Sum of Sq	RSS	AIC
<none>			53.856	-143.628
- corruption	1	5.988	59.844	-129.919

```
- Freedom      1      6.325  60.181 -129.082
- Social       1     47.398 101.254  -51.561
```

Call:

```
lm(formula = score ~ Social + Freedom + corruption, data = happiness)
```

Coefficients:

```
(Intercept)      Social      Freedom  corruption
      0.0779      5.6256      2.2271      -1.2254
```

According to the results of stepwise regression, we choose the model with the smallest AIC as the final model. Then, we fit the following linear model to the data.

$$\widehat{\text{score}}_i = \hat{\alpha} + \hat{\beta} * \text{Social}_i + \hat{\gamma} * \text{Freedom}_i + \hat{\delta} * \text{corruption}_i$$

where

- the  $\widehat{\text{score}}_i$ : the happiness score of the  $i$ th country.
- the  $\hat{\alpha}$ : the intercept of the regression line.
- the  $\hat{\beta}$ : the coefficient for the first explanatory variable Social.
- the  $\hat{\gamma}$ : the coefficient for the second explanatory variable Freedom.
- the  $\hat{\delta}$ : the coefficient for the second explanatory variable corruption.

When this model is fitted to the data, the following estimates of  $\alpha$  (intercept) and  $\beta, \gamma$  and  $\delta$  are returned:

Table 2: Estimates of the parameters from the fitted linear regression model.

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
intercept	0.078	0.559	0.139	0.889	-1.028	1.184
Social	5.626	0.498	11.297	0.000	4.641	6.610
Freedom	2.227	0.540	4.127	0.000	1.160	3.294
corruption	-1.225	0.305	-4.015	0.000	-1.829	-0.622

According to this table, the coefficient for social support tells us that, taking all other variables in the model into account and holding them constant, there is an associated increase. On average, every increase of 1 unit in the social support score increases the happiness index score by approximately 5.63 units. In the same way, when the freedom score of life choice increases by 1 unit, the happiness index score also increases by approximately 2.23 units. On the contrary, for every increase of 1 unit in the score for corruption, the total score of happiness index decreases by 1.23 units.

So the final model is:

$$\widehat{\text{score}}_i = 0.078 + 5.626 * \text{Social}_i + 2.227 * \text{Freedom}_i - 1.225 * \text{corruption}_i$$

Before we can continue to use the fitted model, we must check the model's assumptions. It is best to consider these according to the residual plot in Figure 9.

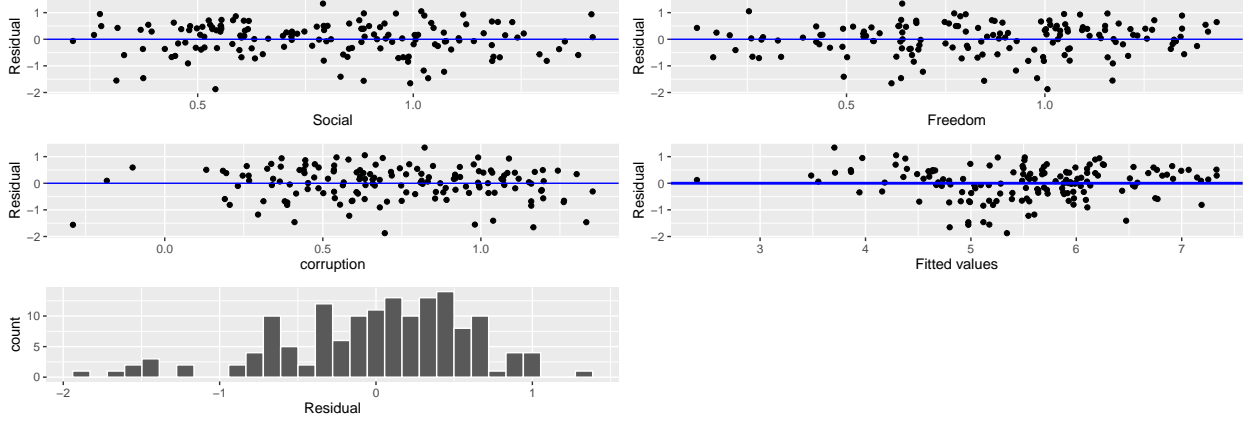


Figure 9: Scatterplots of the residuals by Social, Freedom, corruption, fitted value and the histogram of residuals.

The assumptions of the residuals having mean zero and constant variability across all values of the explanatory variable appear to be valid in this case. According to the three different explanatory variables scatter plots, it can be concluded that the residuals are uniformly distributed above and below the zero line, so the mean is 0. The residuals are randomly distributed around the zero line, and the distribution of the residuals is constant across all fitted values, so there is no obvious pattern or change in the variant. And also the histogram supports the assumption of normal distribution error.

After hypothesis testing, we will analyze the confidence interval and set the confidence level to 95%.

	2.5 %	97.5 %
(Intercept)	-1.027760	1.1835605
Social	4.641318	6.6098138
Freedom	1.160482	3.2937585
corruption	-1.828636	-0.6222416

The data in the table shows that the coefficient range of Social support affecting happiness score is 4.64 to 6.61, and the coefficient range of Freedom is 1.16 to 3.29. The most explanatory variable corruption affects the corresponding variable in the range of -1.83 to -0.62. The interval does not contain 0, indicating that the selected independent variables and their coefficients are valid, and as analyzed above, their given coefficients are the mean value of the interval.

## Conclusions

In our study, to explore what will influence people's evaluations of happiness, we have used six indicators to measure the happiness score. We have used the stepwise regression analysis to select variables. Then, we fit the multiple linear models with three factors as Social, Freedom, and corruption. It can be concluded that social support and freedom to make life choices are positively related to people's happiness and perceptions of corruption which can affect the happiness score. In other words, if a person could get help whenever required or if a person is free to make their own choices, they will perceive contented and joyful life. If we consider interpersonal relationships and emotional needs, the results appear to be reasonable and significant. Besides, if a person believes that their government or companies are corrupt, there will be a decline in the happiness score. Personal level provides better insights on the above all results. According to country-level statistics, residents of western countries lead a happier life than eastern, especially those from Northern European nations like Finland and Denmark. Furthermore, the size of happiness obtained by the top five countries is three times larger than that obtained by the bottom five nations. It shows there is such a huge gap in individual happiness and satisfaction globally.



## Future work

In the future we can consider more factors to have better insights on this happiness report such as the population of the country: as population increases, the diversity increases and providing the same facilities for everyone becomes difficult due to a larger pressure on limited resources, facilities available in a country: as European countries may have better facilities because they are able to allocate facilities efficiently due to a lower population, the effect of natural calamities: some countries may face extreme conditions such as flood, tsunami, etc. more frequently which affects human life, the effect of pollution: Population affects the environment and eventually affects human life.

We can use an alternative approach and fit different models by considering the topics mentioned above. Moreover, our study could be improved by adding the individual's daily happiness score, not just the average yearly happiness scores based on countries. Additionally, it will be useful to consider the death rate among every nation so that we could better understand the influence of covid-19 pandemic has on happiness score index.

## References

- Greyling, T, S Rossouw, and T Adhikari. 2021. "A Tale of Three Countries: What Is the Relationship Between COVID-19, Lockdown and Happiness?" 89: 25–43.
- Helliwell, John F, Richard Layard, Jeffrey Sach, and Jan-Emmanuel De Nev. 2021. *World Happiness Report 2021*. Vol. 1. The South African Journal of economics.