HACKATHON 2019 - QUALIFICATION ROUND REPORT

**Problem: Vietnam Macroeconomics Analysis**

**DASs Team**

|  |  |  |
| --- | --- | --- |
| Đinh Tiến Đạt  dinhtiendat229@gmail.com | Hà Thanh Luân  htluandc@gmail.com | Trần Thị Loan  loantran.uet@gmail.com |

Ha Noi, June 7th, 2019

1. Introduction

This report presents our approach and experiment with Vietnam Macroeconomic Analysis problem. We did exploratory data analysis and figured some useful insights out about the health of Vietnam economy. Based on some basic theory about data mining and economics, we applied the idea of Okun's Law for solving the problem.

1. Our Approach

II.i. Related knowledge:

1. Missing Data Imputation:

a. Mechanism of missingness: There are three main types of missing data:

* Missing completely at random (MCAR): MCAR exists when missing values are randomly distributed across all observations. Missingness in given variable does not depend on any other variable, whether observed or unobserved.
* Missing at random (MAR): The phrase “missing at random” is misleading since MAR data reflect a systematic rather than random pattern of missingness
* Not missing at random (NMAR): It exists when missing values are neither MCAR nor MAR. This happens when missingness depends at least in part on unobserved variables (which is why observed variables fail to predict missingness, making data not MAR).

b. Solution:

b.1. Traditional Approaches:

* Deletion Methods: Delete those with missing data. Bad idea: Unbiased if the cause of missingness is in the actual model. But standard errors will be off
* Single imputation methods:
  + Mean/Mode substitution: Only works if data is MCAR
  + Regression substitution: Predict missing value from observed values

*Pros:*

* Easy and fast.
* Works well with small data.

*Cos:*

* Doesn’t factor the correlations between features. It only works on the column level.
* Will give poor results on encoded categorical features (do NOT use it on categorical features).
* Not very accurate.
* It can introduce bias in the data

b.2: Advanced methods:

* Maximum Likelihood method (ML): A method of estimating the parameters by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. The point in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate. The logic of maximum likelihood is both intuitive and flexible, and as such the method has become a dominant means of inference within much of the quantitative research of the social and medical sciences.
* Multiple imputation method (MI):
* The model will do a "regression imputation"

It will predict the missing values from the available values

It will record the imputed values

* It will then estimate a new regression model using the imputed values
* The model will again impute the missing values from the new regression model

Again, it will record the imputed values

2. Okun's Law:

a. What's Okun's Law:

Okun's Law describes the observed relationship between changes in the unemployment rate and the growth rate of real gross domestic product (GDP). It determines that a GDP growth higher than 3% on average is needed to lower unemployment. This rate varies depending on the country, because it depends on two factors: growth of the labour force and labour productivity.

There are several reasons why GDP may increase or decrease more rapidly than unemployment

decreases or increases: As unemployment increases:

* A reduction in the multiplier effect created by the circulation of money from employees
* Unemployed persons may drop out of the labor force (stop seeking work), after which they are no longer counted in unemployment statistics
* Employed workers may work shorter hours
* Labor productivity may decrease, perhaps because employers retain more workers than they need

b. Mathematical statements:

The gap version of Okun's law may be written as:

, where:

* Y is actual output
* Y̅ is potential GDP
* u is actual unemployment rate
* ū is the natural rate of unemployment
* c is the factor relating changes in unemployment to changes in output

The gap version of Okun's law, as shown above, is difficult to use in practice because Y̅, ū can only be estimated, not measured. A more commonly used form of Okun's law, known as the difference or growth rate form of Okun's law, relates changes in output to changes in unemployment

A picture containing object

Description automatically generated, where:

* Y and c are as defined above
* ∆Y is the change in actual output from one year to the next
* ∆u is the change in actual unemployment from one year to the next
* k is the average annual growth rate of full-employment output

II.ii. Experiment:

1. Data Imputation: We tried several method for dealing data missing:  
a. Single imputation methods:

để có thể áp dụng thành công mô hình Okun, ta cần tính các số sau:

Có tham số k, c.

Tính GDP\_diff.

Tính UPR\_diff.

Tính GDP\_annual\_diff\_rate = delta GDP / GDP cũ.

Tính GDP\_period\_diff\_rate

- Okun.

2. Data Exploration:

Some statistical information of given dataset:

GDP:

Full value from 1980 to 2018 (39 values)

Mean of the GDP values: 67.7161 (Billions of U.S. dollars)

Standard deviation of the GDP values: 68.1957 (Billions of U.S. dollars)

Minimum of the GDP values: 6.293 (Billions of U.S. dollars)

Maximum of the GDP values: 241.272 (Billions of U.S. dollars)

Median of GDP values: 33.873 (Billions of U.S. dollars)

Unemployment rate (UPR):

10 missing values from 1980 to 1989

Have 29 values from 1990 to 2018

Mean of the UPR values: 5.7034 (%)

Standard deviation of the UPR values: 2.8550 (%)

Minimum of the UPR values: 2.1 (%)

Maximum of the UPR values: 12.3 (%)

Median of UPR values: 5.6 (%) in 2004

Correlation of GDP and UPR: -0.8315774406

Distribution of GDP: see Figure.

Distribution of UPR: see Figure.

2.Some analysis information of given dataset:

GDP: Ranges from 6.293 (1989) to 241.272 (2018), with most values at the low end. The median is 33.8730.

UPR: Ranges from 2.1 (2014) to 12.3 (1990).

For every year, GDP decreases by about 5.6164 (Billions of U.S. dollars).

For every year, Unemployment rate (UPR) decreases by about 0.3607.

Từ năm 1980 đến năm 1989, GDP biến động thất thường, không ổn định, tăng/giảm đột ngột.

Từ năm 1990 đến năm 2018, GDP tăng khá ổn định, có xu hướng tăng theo hàm tăng trưởng.

Từ năm 1990 đến năm 2018, nhìn chung, UPR có xu hướng giảm theo hàm tuyến tính (chỉ tăng lên đột ngột vào các năm 1992, 1998 rồi lại giảm).

GDP tăng, UPR giảm cho thấy nền kinh tế Việt Nam đang trên đà phát triển.

There any relationship between GDP and unemployment rate:

For every increase of 100 in GDP, UPR decreases by about 3.23.

3. Features Extraction

4. Discussion

1. Evalution
2. Discussion

3. Evaluation

4. Discussion

References

1. https://vi.wikipedia.org/wiki/%C4%90%E1%BB%95i\_m%E1%BB%9Bi

2. <https://vi.wikipedia.org/wiki/Kinh_t%E1%BA%BF_Vi%E1%BB%87t_Nam>

<https://vi.wikipedia.org/wiki/%C4%90%E1%BB%8Bnh_lu%E1%BA%ADt_Okun>

<https://vi.wikipedia.org/wiki/Kinh_t%E1%BA%BF_Vi%E1%BB%87t_Nam>

<https://www.sbv.gov.vn/webcenter/portal/vi/menu/rm/apph/tcnh/tcnh_chitiet;jsessionid=h94gc53QDYGywsgVv4NZZW4ndHfNc7dfnj1m1MQ534pQbRwMhWNJ!-1383955375!-802188061?centerWidth=80%25&dDocName=SBV312772&leftWidth=20%25&rightWidth=0%25&showFooter=false&showHeader=false&_adf.ctrl-state=4634ta7s2_4&_afrLoop=14487903093432095#%40%3F_afrLoop%3D14487903093432095%26centerWidth%3D80%2525%26dDocName%3DSBV312772%26leftWidth%3D20%2525%26rightWidth%3D0%2525%26showFooter%3Dfalse%26showHeader%3Dfalse%26_adf.ctrl-state%3D18da7qpo3b_4>