ATLANTIS PRESS

Advances in Computer Science Research, volume 92

International Conference on Mathematics, Big Data Analysis and Simulation and Modeling (MBDASM 2019)

Factors That Influence Revenue Non-equalizing

Shanjun Liu and Juncheng Li

College of Mathematics and Fincance, Hunan University of Humanities, Science and Technology, Loudi, China

***Abstract*—Income inequality is an inevitable phenomenon in each country. National government can use the macroeconomic policy. This paper is try to find a relationship between the factors of index systems of macroeconomics and income inequality. Based on the factors of macroeconomics in world bank data, we acquire cross-sectional data of all over the world countries of 75 percentage as far as 2015. By literature review we found out that speed of global economic increase, population of work force, education, growth rate of population and unemployment will influence revenue non-equalizing. Therefore, we are choosing above factors as independent variables establish a stepwise multi-linear regression model. In the economics Gini coefficient measured by people's income differences, thus we are choosing the Gini index as the dependent variable. The important findings of this research is that with population of work force and jobless rate increase revenue non-equalizing growth and growth. In addition, some research also discovered economy situation of these countries was main influence and revenue non-equalizing of developed economics experience are less.**

Keywords—revenue non-equalizing; GiNi coefficient; global economic growth; population of labor; education; unemployment; stepwise multi-liner regression

I. INTRODUCTION

Revenue inequality means to irregular income among various stakeholders. Some research was found that with the development of globalization, the income level of per capita income has increased in various countries and regions. Moreover, some extent of income inequality will motivate people to competition and investment to go ahead in life. Such as, Edward P. Lazear and Sherwin Rosen (1981)[1] found that "the growth of inequality can through incentives by the Reform and Opening-up". This shows some degree of revenue non-equalizing is essential to insensitive economics growth. However, other studies has indicated that with the non-equalizing expansion and spread, it can influence operational of country and financial condition. For example, Stiglitz (2012)[2] found that non-equalizing will affects and is affected by every aspect of national policy, and causes unreasonable of resource allocation.

A social of country should try to keep a good symmetric of revenue non-equalizing and possible impacts. Thus will make the economic system of country even as a whole global economic system be better. Therefore, this study is aim to find out the factors that influence revenue non-equalizing and use econometric analysis to achieve this. Specifically, in this paper set up a stepwise multi-linear regression model, this way is seek to explore a connection between revenue non-equalizing and potential repercussions. People usually think that the speed of global economic growth, the development level of country, population of work force, educational attainments, growth rate of population and unemployment will influence a country's revenue non-equalizing. That is why we will set up a multi-linear regression model, these factors will act as dependent variables used to test and verify to whether a connection subsistent. Ultimately, this paper will reveal income inequality the impact on economics, readers will understand these factors and what it means in this relationship.

II. ECONOMIC DATA AND ECONOMETRIC MODEL

A. Choosing Dependent and Independent Variables

The combination of potential repercussions and explanation variables allows us to build an econometric model. It can be explained that Gini coefficient (revenue non-equalizing) is a function of population of work force, jobless rate, growth rate of population, education and development.

$$GiNi = f(PL,\ U,\ GRP,\ Edu,\ D) \tag{1}$$

The Table I provided an explanation of the dependent variables.

TABLE I. DEPENDENT VARIABLE DESCRIPTIONS

Dependent Variable	Description of Source Variable
Gini	Gini coefficient – Measure of income inequality. Perfect equality is equal to 0 and perfect inequality is equal.
PL	Population of Labor – refers to Labor participation. Count by the population of work force divided by total population.(elder than 15, include 15)
U	Unemployment rate –represents to the number of laborers who have not yet worked in the employed population who meet all the conditions of employment in a certain period of time.
GRP	Growth rate of population – Measuring Fertility rate refers to the ratio of the number of live births to the average number of women of childbearing age in a year.
Edu	Educational attainments–People above 15 years old and time of get the education completed.
D	Indicator variable – when D=1: developed country; when D=0: developing country;

B. Forming Econometric Model

The following procedure is the development of the multi-linear regression model that test and verify the foundation of empirical economic analysis.

To begin with we should obey the assumptions of a stepwise regression model, so the best linear unbiased estimators (BLUE) can be apply. We can make the best explanations of the data we are using, through carry out these supposes is necessary.

In addition, the Principles of Econometrics Textbook states[3], "we shall as much as possible satisfied the supposes of the multi-linear regression model, so we should try to found out a form can correspond the economic theory". Therefore, it is important that choosing the right functional form to achieve best linear unbiased estimators (BLUE). We are tested different models and different functional forms models include: Linear, Log-Linear, Linear-Log, and Log-Log, for find the best explanation model. At the same time, we compared all of models the sum of squared residuals (SSE) with each other. Finally, we found out can provide good outcomes, and best fit to the data is Log-Log formal. Therefore, we put forward the multi-linear regression model is as follows:

$$ln\, GiNi = \beta_1 + \beta_2\, ln\, PL + \beta_3\, ln\, U + \beta_4\, ln\, GRP + \beta_5\, ln\, Edu + \beta_6 D + e \quad (2)$$

C. Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
lnGiNi	145	3.648357	.2185532	3.218876	4.163559
lnPL	145	4.135889	.1672864	3.688129	4.459462
lnU	145	1.89002	.853022	-1.807889	3.386862
lnEdu	145	1.998762	.4949635	.3364722	2.595255
lnGRP	145	.9302897	.468447	.2070142	1.986504

FIGURE I. SUMMARY STATISTICS

Through data acquire and filter, in this model we select all over the world countries of 75 percentage as far as 2015. So far, we are explained the dependent and independent variables, and our model has been preliminary formation. Now that we can move onto the methodology part of this paper a stepwise multi-linear regression model.

III. METHODOLOGY

A. Econometrics Approach

1) Single Regression Model

Through start a single regression, it is necessary to identify how the importance of adding other explanatory variables. The first regression we choose checkout the connection of the GiNi index and the population of work force. Therefore, we can verify our hypothesis that the more population of labor in the labor market the more income inequality. The outcomes are as below:

$$\boldsymbol{Model\ 1}: ln\, GiNi = \beta_1 + \beta_2\, ln\, PL + e \quad (3)$$

The illustration confirms our suppose because the more people participating in the worker market the more gap of individual's revenue, as shown in Figure II. This suggests that the independent variable a one percent change will result in a 0.34% change in the dependent variable. When the independent variable is zero, the dependent variable is equal to 2.218. In this model, the dependent variable and the independent variable, were the revenue non-equalizing and the population of work force. The explanation for this is if there is no population of labor (complete equality equals zero), the Gini coefficient will be approaching entirely equally. People are not go to work, thus they haven't revenue, which means equality for everyone.

```
. reg lnGiNi lnPL

      Source |       SS       df       MS              Number of obs =     145
-------------+------------------------------           F(  1,   143) =   10.78
       Model |  .482075422     1  .482075422           Prob > F      =  0.0013
    Residual |  6.3961595    143  .044728388           R-squared     =  0.0701
-------------+------------------------------            Adj R-squared =  0.0636
       Total |  6.87823492    144   .04776552           Root MSE      =  .21149

------------------------------------------------------------------------------
      lnGiNi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnPL |   .3458723   .1053537     3.28   0.001     .1376204    .5541241
       _cons |   2.217868   .4360851     5.09   0.000     1.355862    3.079874
------------------------------------------------------------------------------
```

FIGURE II. Regression Results for Single Regression model

The p-value of the population of labor is 0.001, indicating that it is highly important.

The Figure II shows that the R-squared is 0.07, which means that the population of labor only explains 7% of the Gini index. In a sensible world, revenue non-equalizing of each country have different situation, it is extremely complex and unique, thus it is cannot depend on just one explanatory factor. Therefore, we need to add a more factors as independent variables to explore a perfect model.

2) Multi-linear Regression Model

We will regress the multiple factors and analysis discuss its feasibility the purpose is to improve the model and more comprehensive consider each country's complex and unique situation. The model and outcomes are follows:

$$\boldsymbol{Model\ 2}: ln\, GiNi = \beta_1 + \beta_2\, ln\, PL + \beta_3\, ln\, U + \beta_4 D + e \quad (4)$$

The Figure III shows the p-value of all variables are less than 0.05 which indicates that it is highly significant. It tells the model that the explanation variables have a statistical relationship with the Gini index. In addition, The R-squared to be 0.3430, which means all the independent variables can explains 34.3% of the Gini coefficient. It is further shown that the inclusion of education and growth rate of population variables has impose an evident no statistical influence on revenue non-equalizing, as changes in the dependent variables can interpreted equally well by these three variables.

```
. reg lnGiNi lnPL lnU D

      Source |       SS       df       MS              Number of obs =     145
-------------+------------------------------           F(  3,   141) =   24.53
       Model |  2.35899084     3  .78633028           Prob > F      =  0.0000
    Residual |  4.51924408   141  .032051376           R-squared     =  0.3430
-------------+------------------------------           Adj R-squared =  0.3290
       Total |  6.87823492   144  .04776552           Root MSE      =  .17903

------------------------------------------------------------------------------
      lnGiNi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnPL |   .4662981   .1043954     4.47   0.000     .2599155    .6726808
        lnU  |   .0886299   .0203364     4.36   0.000     .0484263    .1288336
           D |  -.2388854   .0365341    -6.54   0.000    -.3111109     -.16666
       _cons |   1.605007   .4532929     3.54   0.001     .7088779    2.501136
------------------------------------------------------------------------------
```

FIGURE III. REGRESSION RESULTS FOR MULTIPLE REGRESSION MODEL

Furthermore, the interpretation of the R-squared of the model is that changes in population of work force, changes in jobless rate, as well as changes in level of development can explain 0.343 of the changes in the income inequality. Compared with Model 1, a more interesting analysis of *R*-squared of the model is made. We can see that when we introduce more independent variables, the goodness of fit increases from 0.07 to 0.343. Therefore, approaching our goals step by step, that is found out factors that affect income inequality. We can get the economic significance, but we unable determine the quality of the model 2 based on the degree of prediction of the income inequality.

B. Model Checking and Innovation

To make sure all of the model follows the multiple-liner regression supposes. The multiple regression model should necessary to perform the necessary tests and evaluations to determine whether the model can be applied. Some tests have been performed, including:

- Determination of the degree of fit of the model;
- Significance test of regression equation and coefficient;
- Multiple collinear analysis;
- Test if add the jobless rate, will result other independent variables is significant;
- Test the heteroscedasticity;

After stepwise multi-linear regression model, the heteroscedastic is the only question of these tests. That means at all of independent factors the variance of the error term is not stable. That disobey any one of the supposes is mean the variance of the error term is stable. Accordingly, the previous model not follow the Gauss Markov theorem, so there is no best linear unbiased estimate available.

However, there has many ways to alleviate this problem like: White's heteroscedasticity- consistent standard errors, Estimated Generalized Least Squares, as well as method Generating Least Squares (GLS). Which is can lightens the variance suppose, and when heteroscedasticity exists, it can help to prevent error in calculating interval estimates and test statistics, hence we like to use the White's method.

The adjusted model outcomes are as follows:

$$\boldsymbol{Model\ 3}\text{: } ln\,GiNi = \beta_1 + \beta_2\, ln\,PL + \beta_3\, ln\,U + \beta_4\, ln\,D + e,\ vce(Robust) \quad (5)$$

The first thing people will notice is that the p-value of all variables are equal to 0 which indicates that it is still very important; Therefore, the dependent variables and independent variables has still strong in statistical sense. In addition, the R-squared is equal to 0.343. It compares with the multiple regression model is exact the same. Therefore, the explanation is the same.

```
Linear regression                                      Number of obs =     145
                                                       F(  3,   141) =   32.79
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3430
                                                       Root MSE      =  .17903

------------------------------------------------------------------------------
             |               Robust
      lnGiNi |      Coef.   Std. Err.      t    P>|t|     [95% Conf.  Interval]
-------------+----------------------------------------------------------------
        lnPL |   .4662981    .092246     5.05   0.000      .283934     .6486622
        lnU  |   .0886299   .0197577     4.49   0.000     .0495702     .1276896
           D |  -.2388854   .0313987    -7.61   0.000    -.3009586    -.1768123
       _cons |   1.605007   .3999979     4.01   0.000     .8142385     2.395776
------------------------------------------------------------------------------
```

FIGURE IV. REGRESSION RESULTS FOR MODEL 3

C. Boundedness of the Method

If we want to get more details from the model, we should make it more comprehensive, include:

- Use time series analysis can more explains the changes in revenue non-equalizing and have more data can be collected.
- Considering other independent factors to cause higher fitting and to better explain income inequality.
- Divide the data into different regions like Afican countries. And also divide the data into according to different degree of development. this way can better to explain factors of income inequality in countries around the world.

Therefore, we also should make many modify to innovation this study.

IV. Explanation and Economics Values

Despite the previous restricts, the article finalized define model is as follows:

$$\widehat{ln\,GiNi} = 1.605 + 0.4663\,\widehat{ln\,PL} + 0.0886\,\widehat{ln\,U} - 0.239\hat{D} \quad (6)$$

A. Explanation of Model

1) Population of Labor

The final model regression indicates that the population of work force increase 1% will result in an increase in the Gini index of 0.4663%. This relationship between income inequality and population of labor is positive, which confirms the assumptions we have described in the data segment. According to the confidence interval indicates that we have 95%

confident that the revenue non-equalizing will vary reasonable ranges from 0.2839% up to 0.6487%.

2) Jobless Rate

For the jobless rate, according to regression model indicates that add one percentage of jobless rate will result in a 0.0886 percentage add in the income inequality. We can obvious get positive connection between Gini index and unemployment rate. It also obey suppose at the information section. We have 95% confident that the income inequality will vary reasonable ranges from 0.0496% up to 0.1277%.

3) Development

Multiple regression analysis indicates that when countries have higher the degree of development, it will have a lower Gini index on average. This makes sense from an economic point of view, but it is important to have statistical evidence that revenue non-equalizing is a big problem if country at very lower level of development.

B. Economic Values

It has been suggested that explore the explanation of revenue non-equalizing is too complex and difficult, because there are exist a lot of explanation variables that can influence its outcome. In addition, different country has different regulations, lifestyles, political systems, cultural norms, etc., describing the income gap within its particular country. Therefore, it will become difficult to find a global explanation about which factors will affect income inequality.

However, this article shows some explanations. The results show that the population of labor and unemployment rate are very high, and have a positive correlation with the Gini index. Therefore, economies can be aware that connection of these independent variables and dependent variables, exists negative or positive are statistically proven to change the state's revenue non-equalizing. In addition, the paper also shows that the income inequality in developed countries will be lower than that in non-developed countries.

V. SUMMARY

The principal object of this paper is to study from macro-level to analysis influence factors of revenue non-equalizing. We choose the cross-sectional data of all over the world countries of 75 percentage as far as 2015. Using the stepwise multi-liner regression, we can get the national level of development, jobless rates, and population of work force were deemed essential and have an appreciable impact with the revenue non-equalizing.

In addition, in this article use different tests to evaluates the model whether consistent with the multi-linear regression hypothesis. After these tests, we get the model follow all previous hypothesis except MR3. It is indicated that the heteroscedastic is exist in the model. Therefore, the model violate the Gauss Markov theorem, so there is no best linear unbiased estimate available. We apply the White's heteroscedasticity to ameliorate the problem. Accordingly, we chose set up the model 3 to prevent incorrect range estimation or incorrect test statistics when the model have heteroscedasticity.

Finally, the paper explains the aspects of econometrics and economics. This thesis draws one conclusion that population of work force and jobless rates are positively correlated with revenue non-equalizing, but the gap of income inequality of developed countries lower than non-developed countries. We have 95% confidence to consider that with the small vary in population of work force, the income inequality will vary from 0.2839% to 0.6487%. If the jobless rate changes small, the income inequality will vary from 0.0496% to 0.1277%.

Ultimately, if more in-depth exploration research, we should use dynamic data. Therefore, we have more raw data can be use, and can comprehensive illustrate changes in revenue non-equalizing. In addition, more in-depth exploration research should form different aspect to choose influence factors as independent factors to better explain income inequality. Finally, because of the complexity of trying to explain global income inequality, it is recommended to divide the data according to different ways.

ACKNOWLEDGMENT

This work was supported by the Teaching Reform Research Project of Hunan University of Humanities, Science and Technology (RKJGZ1417).

REFERENCES

[1] Edward P. Lazear and Sherwin Rosen, “Rank-order tournaments as optimum labor contracts,” Journal of Political Economy 89(5):841-64.

[2] J. E. Stiglitz, “The price of inequality: How today’s divided society endangers our future,” W. W. Norton & Company, 2012.

[3] Hill, R. Carter, W. E. Griffiths, and G. C. Lim., “Principles of econometrics,” Econometrica 39.6(2011):1061.

[4] Deyshappriya, N. P. R. (2017), “Impact of macroeconomic factors on income inequality and income distribution in Asian countries,” Tokyo: Asian Development Bank Institute.

[5] Becker, Gary S., “Part 2: Investment in human beings || investment in human capital: a theoretical analysis,” Journal of Political Economy 70.5(1962):9-49.

[6] Kuznets, Simon, “Economic growth and income inequality,” American Economic Review 45.1(1955).

[7] Nolan, Brian, “Unemployment and the size distribution of income,” Economica (New Series) 53.212(1986):421-445.

[8] Surbhi S. (2015), “Difference between developed countries and developing countries,” Financial Times, June 18, 2015.

[9] “World development indicators,” https://datacatalog.worldbank.org

[10] Austin Hoggatt , Irma Adelman , Cynthia Taft Morris, “Economic growth and social equity in developing countries,” J. Technology and Culture, 1975, Vol.16 (4), pp.678