



## Analysis of Labor Force Participation Rate Using Multiple Linear Regression

Muhammad Syazwan Romzi\*, Muhammad Fauzee Hamdan  
Department of Mathematics, Faculty of Science  
Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia  
\*Corresponding author: syazwanromzi@gmail.com

### Abstract

This study is carried out to investigate the relationship between independent variable inflation rate, urbanization rate, unemployment rate, and foreign direct investment toward dependent variable labor force participation rate. Data were obtained from the Department of Statistics Malaysia and World Bank Data website. Descriptive statistics and multiple linear regression were used in this study. The estimate regression function is labor force participation rate =  $69.938 + 0.003$  unemployment rate –  $0.021$  inflation rate –  $1.320$  Urbanization –  $0.357$  foreign direct investment. All the data used in this study are randomly distributed, so they cannot be transformed. The findings found that foreign direct investment and urbanization are important drivers to the model, whereas inflation and unemployment rate have a minimal effect on the labor force participation rate.

**Keywords:** Regression; Labor force participation rate; Malaysia.

### 1. Introduction

Labor force participation rate had been decreasing slightly in 2020 because of the pandemic covid-19. Now, labor force participation rate increasing back steadily from 68% to 70%. If Labor force drop it will effect country's growth because our country target to become developed country and be more progressive. For the thing to be done our country have to rise labor force participation above 70% to keep our country stable as a developing country.

Malaysia has a very big population and also many potential prospective labour force [7]. According to recent studies, employees with a higher education level are joining the labour force at longer ages in the United States, and comparable patterns are being noticed internationally. The labour force expanded rapidly in the 1970s and 1980s, as women's participation rates increased and the baby-boom generation entered the workforce (Toossi,2013).

This research focuses on two kind of method first descriptive analysis which is summary statistics. This study using summary statistics likely for us to observed the trend of the data throughout the year since 1990 to 2020. Total of the data that are used in this research was 155 including data regarding labor force participation rate, inflation, urbanization, unemployment and foreign direct investment. Second using multiple linear regression to construct a model that fit to the data. The scope of study will be mainly on labour force time series data in Malaysia.

This research aim to (1) To explore the combination of labour force data and describe the trend of data using plot method (2) To identify the significant factor which is independent variable to the dependant variable (3) To construct a multiple linear regression model that fit to the data labor force participation rate.

### 2. Literature Review

#### 2.1. Regression Analysis

##### 2.1.1. Regression analysis

Regression analysis is widely used for prediction. Regression analysis is a set of statistical processes used in statistical modelling to estimate the relationships between a dependent variable often known to

as the outcome or response variable and one or more independent variables often referred to as predictors, covariates, explanatory variables, or features [11].

### 2.1.2. *Linear regression*

Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables also known as dependent and independent variables. The case of one dependent variable is called simple linear regression.

$$Y = \alpha + \beta x + \varepsilon$$

## 2.2. Data

### 2.2.1. *Inflation*

Inflation is defined by economists as a rising price in the general price level of products and services in a sector over time. Inflation expectations have driven consumers to spend more aggressively, resulting in increased production activity and employment. Hence, high levels of employment can be achieved only at high levels of inflation. Despite inconsistencies about the severity of this relationship, recent literature is consistent about the overall nature of this relationship. For example, inflation has a negative impact on long-term growth (Fischer,1993).

### 2.2.2. *Urban Growth*

Urbanization is linked with an increase in population in cities or towns as a result of migration from agricultural regions, which is most prevalent as a result of changes in social and economic situations, resulting in a shift from rural to urban civilizations. Since the early twentieth century, most emerging countries have experienced rapid urbanization. When industrial growth emerges in urban regions, it causes a movement of job searchers from rural to urban areas, causing the urban employment market to become overloaded

### 2.2.3. *Unemployment rate*

Unemployment define as a state of not having a job or being unemployed or the fraction of persons who can work and actively seek jobs but are unable to do so. The unemployment rate is one of the most common economic indicators used to assess an economy's health. It fluctuates with the business cycle, rising during economic crises and dropping during upswing. It is one of the most closely monitored metrics by policy makers, investors, and the general public. Participation rates in the labour force have an impact on the amount of labour available therefore, as a consequence, on unemployment (Baharom,2013).

### 2.2.4 *Foreign direct investment*

A foreign direct investment (FDI) is a direct investment made by a company in one country to acquire a significant share in the company in another country (Ho et al.,2014). In general, the term refers to a business decision to buy a significant stake in or buy a foreign company completely in order to expand its activities to a new region. For a developing country like Malaysia, FDI has played a key role, particularly in the manufacturing sector, where it can create more jobs.

## 3. Methodology

### 3.1. *Research Data*

This research used secondary data from an open source website that is fully available online. The information was gathered from the websites of the Department of Statistics Malaysia (DoSM) and the World Bank site. Unemployment rate, Foreign direct investment, Urbanization, and Inflation rate are the independent factors, whereas the labor force participation rate is the dependent variable. The data gathering period spans 1990 to 2020, with a total sample size of 31. All of this information is collected on an annual basis.

### 3.2. Descriptive Analysis

Descriptive analysis is a type of data analysis that is used to organise and explain the structural elements of data collected by researchers. In this research, Labor force participation rate is calculated by the number of unemployed people, urbanization, Inflation Rate in percentage state, and Foreign Direct Investment. This included skewness and kurtosis of the data. Skewness is a measure of the probability distribution's "asymmetry," in which the curve appears distorted or skewed to the left or right. Kurtosis is a measure of the probability distribution's "tailedness," or whether or not the tails approach zero asymptotically.

### 3.3. Pearson Correlation Coefficient

The Pearson Correlation Coefficient test is a statistical test that's used to measure the relationship between independent variable and dependent variables. It's commonly used by researchers to analyze the relationship between the two. Pearson's correlation coefficient is the covariance of two variables divided by the product of their standard deviations. In Pearson correlation coefficient the correlation that allowed for the system is below 0.7. When the value of Pearson correlation coefficient is greater than 0.7 or less than -0.7 it will be considered as strong correlation (Vatcheva,2016).

### 3.4. Normality Test

The Kolmogorov-Smirnov and Shapiro-Wilk tests are commonly used to determine normality (Tsagris,2021). Null hypothesis is the data are normally distributed. The alternate hypothesis is the data are not normally distributed. If p-value, then the null hypothesis has to be rejected and if p-value, then the null hypothesis is not rejected (Kwak,2019). Since the sample size more than 30 therefore it appropriate to use both of the method to check normality. Although statistical tests can objectively assess normality, they occasionally lack sensitivity at small sample sizes or exhibit excessive sensitivity at high sample sizes.

### 3.5. Multicollinearity Test

In this research, VIF are used to diagnose multicollinearity between the independent variable, VIF start at 1 and has no upper limit which is it can be more higher. A value of 1 indicates that no connection exists between any of the model's predictor variables. A value of 1 to 5 indicates moderate correlation between a certain predictor variable and other predictor variables in the model, but it is not serious in nature in this stage. A correlation between a given predictor variable and the model's other predictor variables that is larger than 5 indicates a possible severe correlation. In this instance, the regression output's coefficient estimates and p-values are most likely unreliable.

### 3.6. ANOVA

ANOVA (Analysis of Variance) is the most commonly used statistical method for hypothesis testing nowadays. It covers a wide range of topics and can accommodate a higher number of experimental designs. ANOVA was performed for checking whether a significant relationship exist between dependent variable and independent variable. The ANOVA test is used as the first stage in examining the variables that influence a certain data set.

### 3.7. Multiple linear regression

Multiple linear regression attempts to model the linear relationship between independent and dependent variables. The formula of Multiple Linear Regression (MLR) is shown below:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \epsilon$$

- y represented labor force participation rate
- $x_1$  represented Unemployment rate (%)
- $x_2$  represented Foreign direct investment (Billion)
- $x_3$  represented Urban Growth (%)

- $x_4$  represented Inflation rate (%)
- $\beta_0$  is the y-intercept
- $\beta$  are slope coefficients for each independent variable
- $\varepsilon$  is the error or also known as residuals

In multiple linear regression one hypothesis had been test. The hypothesis are as below:

$$H_0 : \beta_1 = 0, \beta_2 = 0, \dots, \beta_n = 0$$

$$H_1 : \beta_1 \neq 0, \beta_2 \neq 0, \dots, \beta_n \neq 0$$

#### 4. Results and discussion

##### 4.1. Descriptive Analysis

Table 1 : Descriptive statistics

	N	Minimum	Maximum	Mean	Std. deviation	skewness		Kurtosis	
		statistic	statistic			statistic	Std. error	statistic	Std. error
LFPR	31	62.60	68.70	65.580	1.7896	0.137	0.421	-1.038	0.821
Unemployment	31	2.40	4.50	3.358	0.4515	0.663	0.421	1.795	0.821
FDI	31	-1.1	5.3	2.355	1.3902	-0.025	0.421	0.271	0.821
Urban Growth	31	2.00	5.00	3.4452	1.0576	0.132	0.421	-1.464	0.821
Inflation	31	-5.18	8.00	-0.6139	3.8720	0.920	0.421	-0.059	0.821
Valid N	31								

The lowest rate of labor force participation rate is at 62.60 while the highest is at 68.70. It has a skewness of 0.137 which means it is nearly symmetrical. The kurtosis of -1.038 indicates that it is a light-tailed distribution. Unemployment rate has the minimum level at 2.40 and the highest at 4.5. it shows a value of skewness 0.663 which mean it is positively skewed. The distribution has a light tail since the kurtosis is 1.795. The inflation rate has the lowest value at -1.1 and the highest value is 5.3. Because the skewness is -0.025, the inflation data is nearly symmetrical, with a light tailed distribution because the kurtosis is 0.217. Based on the table, the minimum value of urban growth is 2% and the maximum value is 5% annually. It shows a value of skewness 0.132 and this show that it is fairly symmetrical. The data has a light tail since the kurtosis value is -1.464. For the FDI, it has a lowest value at -5.19 billion and the highest 8 billion. it is positively skewed at 0.920. It is a light tail distribution since the kurtosis value is -0.059.

##### 4.2. Normality Test

	Kolmogorov-smirnov			Shapiro-wilk		
	statistic	df	Sig.	statistic	df	Sig.
LFPR	0.097	31	0.200	0.958	31	0.256
Unemployment	0.151	31	0.070	0.920	31	0.024
Inflation	0.120	31	0.200	0.980	31	0.721
FDI	0.152	31	0.150	0.897	31	0.007
Urban Growth	0.152	31	0.067	0.933	31	0.006

For normality test, we will look at kolmogorov-smirnov test. Since all the test statistics value are larger than 0.05, in show that the data is normally distributed.

**4.3. Multicollinearity**

Model	Collinearity statistics	
	Tolerance	VIF
Unemployment	0.838	1.135
FDI	0.672	1.773
Urbanization	0.495	1.826
Inflation	0.513	1.535

For each VIF of the variable is less than 5 and the tolerance is greater than 0.3. Therefore, the regression model does not have serious multicollinearity between each independent variable.

**4.4. Pearson correlation coefficient**

**Correlations**

		LFPR	INFLATION	UNEMPLOYMENT	FDI	UrbanGrowth
Pearson Correlation	LFPR	1.000	-.138	.078	-.337	-.360
	INFLATION	-.138	1.000	-.246	-.443	.586
	UNEMPLOYMENT	.078	-.246	1.000	-.085	-.015
	FDI	-.337	-.443	-.085	1.000	-.550
	UrbanGrowth	-.360	.586	-.015	-.550	1.000
Sig. (1-tailed)	LFPR	.	.229	.338	.032	.023
	INFLATION	.229	.	.091	.006	.000
	UNEMPLOYMENT	.338	.091	.	.324	.469
	FDI	.032	.006	.324	.	.001
	UrbanGrowth	.023	.000	.469	.001	.
N	LFPR	31	31	31	31	31
	INFLATION	31	31	31	31	31
	UNEMPLOYMENT	31	31	31	31	31
	FDI	31	31	31	31	31
	UrbanGrowth	31	31	31	31	31

**A. Unemployment and Labor force participation rate**

The correlation coefficient is 0.078 shows the relationship between the two variables is positive. Therefore, unemployment and labor force participation rate have very weak relationship between themselves.

**B. Foreign direct investment and Labor force participation rate**

The correlation coefficient value is -0.337. Hence, it is a negative relationship between two variable. Therefore, Foreign direct investment and labor force participation rate have a moderate relationship between themselves.

**C. Urban Growth and Labor force participation rate**

The value of correlation coefficient for urbanization and labor force participation rate is -0.36 which mean the relationship between the variables has negative and weak correlation.

**D. Inflation and Labor force participation rate**

The value of correlation coefficient for for inflation and labor force participation rate is -0.138. Therefore, there is an inverse correlation relation between two variable and shows that the two variable has negative and weak correlation

**4.5. Model summary**

Model	R	R square	Adjusted R square	Std. Error of the estimate
1	0.737	0.543	0.473	1.29958
a. Predictors: (constant), Unemployment, FDI, Urban Growth, Inflation				
b. Dependent variable: Labor force participation rate				

the summary shows the R-Square value is 0.543 implies that 54.3 percent of the regression model is significant for research results. In other words, the four independent variables account for 54.3 percent of the variability in the dependent variable. Other variables not included in the study are likely to account for the remaining percentages.

**4.6. Parameter estimate**

model	Unstandardized coefficient		Standardized coefficient	t	Sig.
	B	Std. Error	Beta		
(constant)	69.938	2.164		32.312	0.000
Unemployment	0.003	0.560	0.001	0.006	0.996
Inflation	-0.021	0.229	-0.016	-0.092	0.927
Urban Growth	-1.320	0.305	-0.777	-4.335	0.000
FDI	-0.357	0.076	-0.772	-4.700	0.000

From the table above unstandardized B value are use for the model to be build. The multiple regression model are as below

$$\text{Labor force participation rate} = 69.938 + 0.003 \text{ unemployment rate} - 0.357 \text{ FDI} - 1.320 \text{ Urbanization} - 0.021 \text{ inflation}$$

This model is multiple linear regression model with labor force participation rate as the dependent variable. The independent variable is Unemployment rate as  $x_1$ , Inflation rate as  $x_2$ , Urban Growth as  $x_3$  and Foreign direct investment as  $x_4$ . From the multiple linear regression obtained,  $\beta_0 = 69.938$  indicates that if there is no unemployment rate, no FDI, no urbanization, no inflation the estimated labor force participation rate increase by 69.938%. when urbanization is growth by 1% labor force participation rate will decrease by 1.320% holding all other variable constant.

**4.7. Hypothesis Testing**

Independent variables	Hypothesis	Conclusion
Unemployment	$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$	Do not reject $H_0$ . Unemployment has an insignificant relationship with labor force participation rate
Foreign direct investment	$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$	Reject $H_0$ . Foreign direct investment has a significant relationship with labor force participation rate

Urban Growth	$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$	Reject $H_0$ . Urban Growth has a significant relationship with labor force participation rate
Inflation	$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$	Do not reject $H_0$ . Inflation has an insignificant relationship with labor force participation rate

#### 4.8. ANOVA

Model	Sum of square	df	Mean Square	F	Sig.
Regression	52.163	4	13.041	7.721	0.000
Residual	43.911	26	1.689		
Total	96.074	30			
a. Dependent variable : Labor force participation rate					
b. Predictors : (constant), Unemployment, FDI, Urbanization, Inflation					

Value of statistic F in the model is 7.721 and this value is greater than the value in statistics table  $F_{0.05,4,26} = 2.76$ . This enough to show that  $H_0$  was reject at the level of significance 5%. Then, there exist linear relationship between dependent variable and at least one of the independent variables. From the table, p-value associated with the F-statistics is 0.000, so it concluded that the model is significant

#### 5. Conclusion

To summarize the conclusion, this research is to investigate the relation between unemployment, foreign direct investment, urban growth and inflation on labor force participation rate. Factor such as FDI and urbanization have significant relationship with labor force participation rate whereas Unemployment and Inflation do not significantly affect labor force participation rate in this research. Such outcomes could be the consequence of a regression model that uses a small sample size of data, making it difficult for the regression model to predict the exact relationship between the dependent and independent variables.

Through this study, the government and policymakers can make better decisions by understanding which factors have a substantial impact on Malaysia's labour force participation rate. It is found that urbanization can increasing labor force participation rate. For this part both government and private sector play a part.

#### Acknowledgement

The researcher would like to thank all people who have supported the research from start to finish. We would also like to thank the Department of Statistics Malaysia and World Bank website for the data that was used in this research.

#### References

- [1] Andrade, C. (2019). The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. *Indian journal of psychological medicine.* 41(3), 210–215.
- [2] Bakar, . A. N., A. (2007). Labor force participation of women in Malaysia. In Paper for International Economic Conference on Trade Industry.
- [3] Drezner, Z., Turel, O. and Zerom, D. (2010). A modified Kolmogorov– Smirnov test for normality. *Communications in Statistics—Simulation and Computation*®. 39(4), 693–704.

- Frees, E. W. (2006). Forecasting labor force participation rates. *Journal of Official Statistics*. 22(3), 453–485.
- [4] Groesbeck, . I. L. D., J. D. (1994). The relationship between household size, real wages, and labor force participation rates of men and women. *Economics Research Institute Study Paper*. 94(13).
- [5] Hayes, A. (2022). Multiple Linear Regression (MLR). Retrievable at <https://www.investopedia.com/terms/m/mlr.asp>.
- [6] Krueger, A. B. (2017). Where have all the workers gone? An inquiry into the decline of the US labor force participation rate. *Brookings papers on economic activity*. 2017(2), 1.
- [7] Kwak, S. G. and Park, S.-H. (2019). Normality test in clinical research. *Journal of Rheumatic Diseases*. 26(1), 5–11.
- [8] Leynes, A. G. (2021). Modeling the Labor Force Participation Rate of the Philippines through Multiple Linear Regression. *Turkish Journal of Computer and Mathematics Education*. 12(3), 4987–4998.
- [9] Mbah, A. K. and Paothong, A. (2015). Shapiro–Francia test compared to other normality test using expected p-value. *Journal of Statistical Computation and Simulation*. 85(15), 3002–3016.
- [10] Ni, Y. Z. M. M. M. . S. S. S., T. V. (2021). Assessing youth unemployment rate in Malaysia using multiple linear regression. *Journal of Mathematics Computing Science*. 7(1), 23–34.
- [11] Ozerkek, Y. (2013). Unemployment and labor force participation: A panel cointegration analysis for European countries. *Applied Econometrics and International Development*. 13(1), 67–76.
- [12] Paez, G. N. and Tin, M. S. (2021). Gender-based differences in Myanmar’s labor force. *World Development Perspectives*. 21, 100299.
- [13] Queiroz, . F. M. L. A., B. L. (2021). The evolution of labor force participation and the expected length of retirement in Brazil. *The Journal of the Economics of Ageing*. 18.
- [14] Ratner, B. (2009). The correlation coefficient: Its values range between+ 1/- 1, or do they? *Journal of targeting, measurement and analysis for marketing*. 17(2), 139–142.
- [15] Seshadri, A. (n.d.). A Meta-analysis of the decline in the labor force participation rate.
- [16] Shpresa, S. (2013). Application of Multiple Linear Regression Analysis of Employment through ALMP. *International Journal of Academic Research in Business and Social Sciences*. 3(12), 2222–6990.
- [17] Toossi, M. (2013). Labor force projections to 2022: The labor force participation rate continues to fall. *Monthly Lab. Rev.* 136, 1.
- [18] Tsagris, M. and Pandis, N. (2021). Normality test: Is it really necessary? *American journal of orthodontics and dentofacial orthopedics*. 159(4), 548–549.
- [19] Vatcheva, K. P., Lee, M., McCormick, J. B. and Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.)*. 6(2).