Difference In Labor Force Participation Rate - A Linear Regression Analysis

Yusuf Elbana
yae13@rutgers.edu

Paul Samoylov
ps1172@rutgers.edu

Juan Ocampo
jao181@rutgers.edu

Rutgers University

Abstract

We analyzed the differences in labor force participation rates between men and women using factors like year, Gross Domestic Product (GDP), literacy rates, life expectancy, and the total female workforce. The data was gathered from the World Bank API, which also includes many other factors we didn't explore. Using linear regression, we examined how these variables relate to the gender disparities. This helped us identify key drivers of the participation gap, with some factors showing stronger connections than others.

We will use the terms 'parameters' and 'factors' interchangeably when discussing the regression model(s).

1 Level One

We collected data for over 250 countries, focusing on the time, GDP, and the gender gap in labor force participation rates.

Our first parameter was GDP. Logically as a country's GDP rises, it suggests that more job opportunities will become available over time, potentially reducing the gender gap as both genders benefit from increased employment.

We calculated the slope, intercept, and coefficient of determination(R2) by first minimizing the sum of squares regression (SSR = predicted - mean), then dividing it by total sum of squares (SST = observed - mean).

The slope and intercept came out to be ˜-0.00017 and 24.88 respectively. Our R2 came out to be ˜0.03. See Figure 1:

$$\text{slope} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\text{intercept} = \bar{y} - \text{slope} \times \bar{x}$$

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

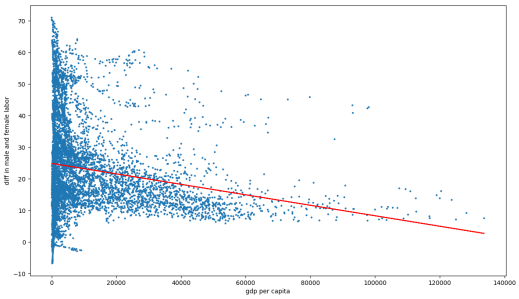

Figure 1: GDP vs Diff in Labor Participation Rates

Our findings show that our model only has a 3% predictability rate, signifying a weak relationship. We assume this is because of the concentration GDP at (0-5000), as the gender gap is scattered across all possible values showing no correlation between the two factors.

2 Level Two

Our findings regarding the relationship between a country's overall GDP and the gender gap were disappointing, as our model's predictive power was only marginally better than simply taking the average. However, this raises the question: are there additional factors within the data that might be more strongly correlated with the gender gap?

Could time(year) be a factor influencing the gender gap in male and female labor force participation? Over time, more women have been entering the workforce, breaking away from traditional roles and contributing to this (possible) shift. So we hypothesize that time could be a relevant factor regarding the gender gap in labor force participation rates.

In the next regression model, we will combine our two parameters (GDP and time).

In order to find our intercept and slopes, we opted to use matrix multiplication since we are are dealing

with multiple parameters.

The slope for GDP, Year, and the intercept came out to be ˜-1.42, ˜-1.8, and ˜3.85 respectively. Our R2 came out to be ˜0.05. See Figure 2:

$$\begin{bmatrix} \text{intercept} \\ \text{slope}_1 \\ \text{slope}_2 \end{bmatrix} = (X^T X)^{-1} X^T y$$

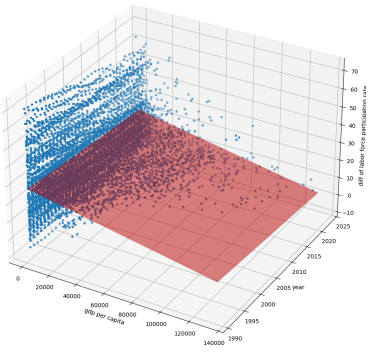

Figure 2: GDP and Year vs Diff in Labor Participation Rates

Disappointingly, our R2 indicates that our model's accuracy is only about 4% better than taking the average. What other parameters could we explore to better understand the relationship between the difference in labor force participation rates between genders?

3 Level Three

In this section we will find additional factors (other than GDP and date) that may influence the difference in labor force participation rates between men and women.

We've already created a model based off GDP and time, which had disappointing results. Via the World Bank API, we will collect more data in order to create a more intricate model. The following parameters will be added in our next model:

- Total female percentage of labor force
- Literacy rate (male)
- Literacy rate (female)
- Life expectancy (male)
- Life expectancy (female)

Let's first gauge the relationship of each independent factor against the gender gap regarding labor force participation.

Note, due to missing parameter values (null via api), the data has ***significantly*** *decreased in size.* See Figure 4.

Immediately, we can see an inverse relationship between the total percentage of women in the workforce and difference in labor force participation rate. As more women join the workforce, the gender gap in labor force participation rate decreases. With this insight in mind, let's proceed to calculate the model using seven parameters.

The calculations remain fundamentally the same, with the only difference being the size of the resulting matrix.

$$\begin{bmatrix} \text{intercept} \\ \text{slope}_1 \\ \text{slope}_2 \\ \vdots \\ \text{slope}_n \end{bmatrix} = (X^T X)^{-1} X^T y$$

After several calculations, we came to find that our seven parameter model had an R2 of ˜0.95! However, this intrigued us since two of those parameters had a low R2 (GDP:˜3% Year:˜4%). Could it be possible that the relationship between the gender gap and certain parameters are *exceedingly* stronger?

While our model shows a high R2 , the 2D plots of each parameter suggest that certain parameters may have a significantly stronger correlation to the gender gap in labor force participation rates. By taking the R2 of the full model and iteratively subtracting the R2 of each individual parameter model, we can determine the importance of each factor in the overall model. The smallest result will indicate the most significant factors, since that singular parameter made up a larger amount of the full model.

$$\Delta R^2 = R^2_{\text{full model}} - R^2_{\text{single parameter model}}$$

The two most important parameters were the **total percentage of women in the workforce** and the **life expectancy of males**, with R2 values of ˜0.92 and ˜0.08 respectively. See Figure 3 for the model:

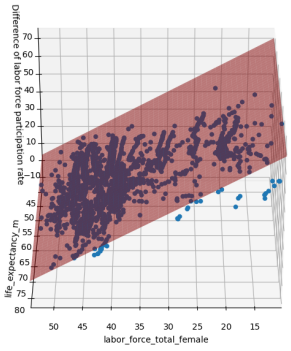

Figure 3: Important Factors vs Diff in Labor Participation Rates

4 Conclusion

Even though our data size significantly decreased because of null values, we found that the total percentage of women in the workforce and male life expectancy are the most important factors affecting the gender gap in labor force participation rate. However, the data varies *drastically* between countries. Typically, countries with a higher GDP usually have smaller gender gaps (Figure 1 GDP:60.000-140,000). Despite this, majority of countries we collected data from have a low GDP, which heavily effected our regression model. Adding on, we hypothesize that in developing countries with a lower GDP(Figure 1 GDP:0-5,000), the gender gap in labor force participation seems more sporadic, which most likely effected the GDP model's predictability accuracy.

5 Technology Used

5.1 Programming Language

- python

5.2 Libraries Used

- requests
- json
- pandas
- time
- mysql.connector
- matplotlib
- numpy
- tkinter

References

[1] Rencher, Alvin C.; Christensen, William F. (2012), "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, vol. 709 (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679

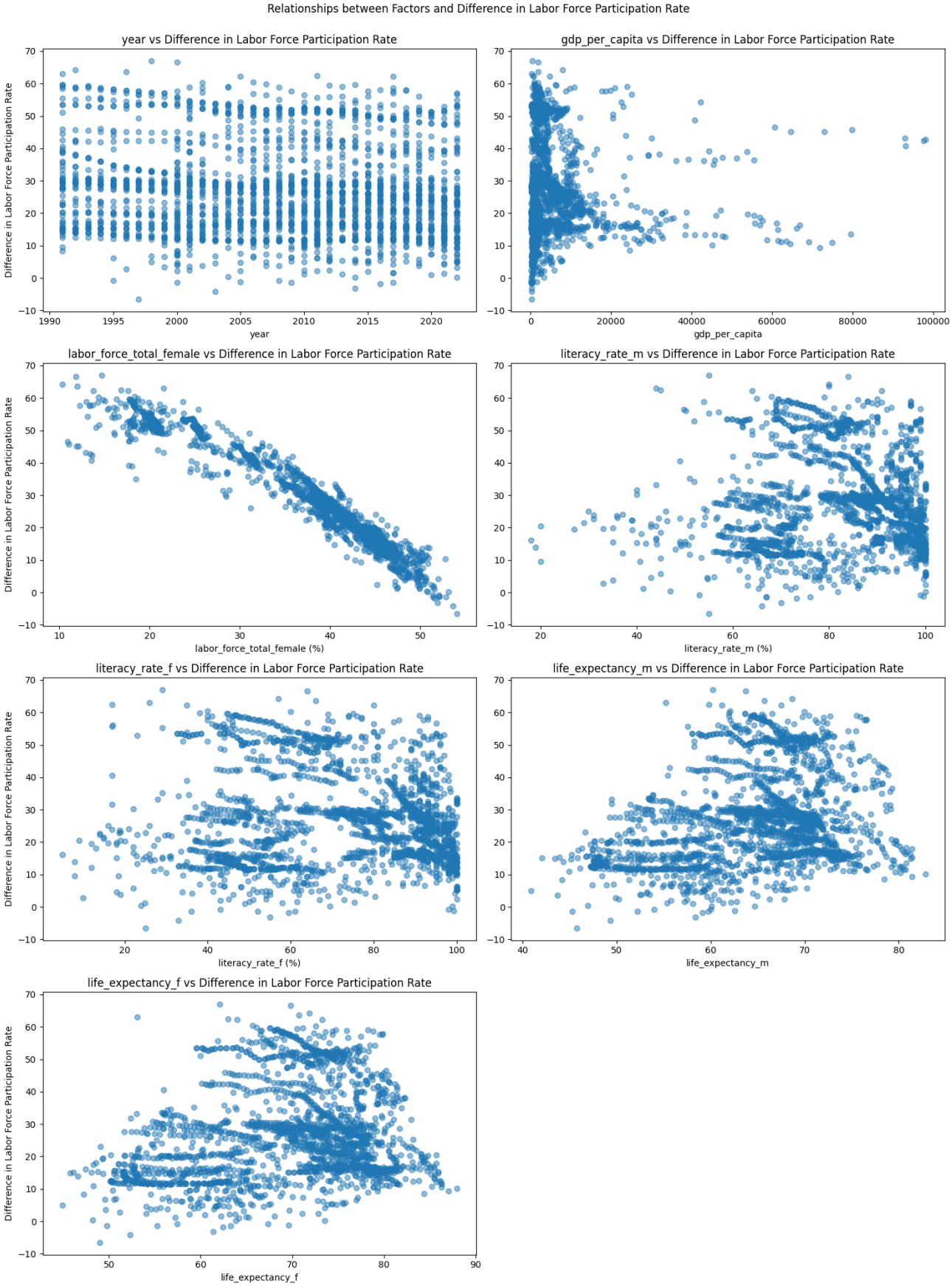

Figure 4: Independent Parameters vs Diff in Labor Participation Rates