Zsófia Rebeka Katona
https://github.com/zsofiarebeka/DA2_Final-term

Factors Affecting Women's Parliamentary Seats: Education, Gender Inequality, Labor, and Development

1. Introduction

The underrepresentation of women is not a new phenomenon under the sun, in particular when it comes to roles with higher authorities. One example are roles in the institutions of representative democracies, parliaments. Male majority in parliaments has been considered normal in the past decades, however female representation is gaining more and more attention and importance. There are several socio-economic factors that can be associated with this change, however female secondary education participation, a country's Gender Inequality Index (GII) and the female labour force participation are crucial benchmarks to assess female representation in parliaments across different countries. My aim with this paper is to study the correlation between:

- female secondary education participation rate and % of parliament seats held by women
- female secondary education participation rate, GII and and % of parliament seats held by women
- female secondary education participation rate, female labour force participation rate and % of parliament seats held by women
- all of the abovementioned rates and indexes and % of parliament seats held by women
- To broaden the scope of understanding gender disparities: Study the frequency distribution of the % of seats sorted by Human Development Index

2. Data

The data used in this report was found on [Kaggle](), but the original source is [Human Development Reports](), namely the *Table 5: Gender Inequality Index*. The dataset contains 195 rows and 13 columns, where data is categorized by countries and is ranked by the Human Development Index (HDI). Other variables include ranks, index values and % rates and the original data type of the columns are *objects*. The dataset contains several missing values denoted as '..' or '...'. The dataset is representative of 2021, except for the maternal mortality ratio, which reflects 2017 - and is not needed for our analysis. For this study, data cleaning and wrangling was necessary before performing regression models. Besides renaming and reformatting and dropping unnecessary columns, I replaced the missing values with the correct *NaN* format and transformed the data types of the columns from object to float to be able to perform regression models. The data cleaning resulted in the following conclusions:

- There are several missing values as the 'count' values are different
- The dependent variable, "parliamentseatsheldbywomenrate" ranges from 0% to 55.7% and has 2 missing values
- The explanatory variable, femalesecondaryeducation column has missing values (18, namely)
- The first conditioning variable, female labour force ranges from 6 to 83.1 and has 15 missing values
- The second conditioning variable, GII has 25 missing values
- In conclusion, we need to drop the rows where we have missing values in the variables that we carry out this analysis with

Variables with explanation used for the regression analysis:

- **Dependent variable (y)**: Share of seats in the parliament: Proportion of seats held by women in the national parliament expressed as a percentage of total seats.
- **Explanatory variable (x)**: Female secondary education: Percentage of the female population ages 25 and older that has reached (but not necessarily completed) a secondary level of education.

- **Conditioning variable (z1)**: Gender Inequality Index: A composite measure reflecting inequality in achievement between women and men in three dimensions: reproductive health, empowerment and the labour market. The exact calculation of the index can be found here.
- **Conditioning variable (z2)**: Female labour force: Percentage of the female population ages 15 and older that participate in the labour force.
- **Dummy variable**: To be able to perform a logistic regression, I needed to create a binary variable of the dependent variable, "seats held in the parliament by women". As it is a percentage rate, I created a binary outcome variable indicating whether the percentage of seats held by women is above or below a specific threshold:
- 1: Indicates that the percentage of seats held by women is above a certain threshold (25%)
- 0: Indicates that the percentage of seats held by women is below or equal to the threshold

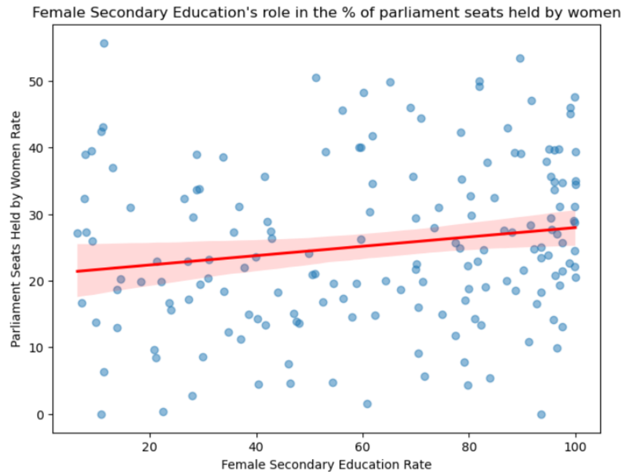
The other variables in the dataset include:
- Countries
- Human Development Index indicators: HDI rank and HDI level
- Gender Inequality Index indicators: GII (z1) and GII rank
- Reproductive Health indicators: Maternal mortality ratio: Number of deaths due to pregnancy-related causes per 100,000 live births and Adolescent birth rate: Number of births to women ages 15–19 per 1,000 women ages 15–19.
- % of seats in the parliament held by women (y)
- Educational indicators: Female secondary education rate (x) and male secondary education rate
- Workforce indicator: Female labour force participation (z2) and Male labour force participation

3. Models

$$parliamentseats = \beta_0 + \beta_1 x_1 x femalesecondaryeducation$$

Model 1 is a simple linear regression of % of seats held by women in the parliament on female secondary education rate. The regression table can be found at Appendix A1. According to the β1 of the model, females with secondary education have 7.01% more probability, on average, to earn seats in the parliament. Approximately 2.9% of the variation in the percentage of seats held by women in parliament can be explained by the variation in female secondary education rate. The CI of the slope is [0.008;0.132] which does not include zero, meaning that we can be 95% confident that the true effect of the female secondary education rate on the percentage of seats held by women in parliament is between 0.8% and 13.2%. The p-value < 0.05 also proves the above statement. To identify a potential nonlinear relationship, I created a regression graph with the lowess method. (Appendix A2) According to the chart, there is a slight decrease in the seats between 20% and 40% of female secondary education rate, which is not visible in the simple linear regression.

Zsófia Rebeka Katona
https://github.com/zsofiarebeka/DA2_Final-term



Scatter plot of female secondary education's role in the % of parliament seats held by women

$$parliamentseats = \beta_0 + \beta_1 femalesecondaryeducation + \beta_2 GII$$

The second model studies how the female secondary education and the GII combined have an overall impact on the seats held in the parliament. (Appendix B1) All covariates are significant at 99%. β1 of this model indicates that for a one-unit increase in female secondary education rates, and while holding the GII constant, the model predicts a decrease in the percentage of seats held by women in parliament by 0.208 units. The CI of the slope is [-0.295;-0.121] not including zero. Conditioning on z1, β2 shows that for a one-unit increase in the GII, and while holding female secondary education constant, the model predicts a decrease in the percentage of seats held by women in parliament by 52.114 units. The two variables' CIs don't overlap, suggesting a statistically significant difference. The robust SE of this coefficient is significantly higher (6.409) than the SE of β1, meaning that it might be less reliable or more sensitive to variations in the data compared to β1. R-squared is 0.267, meaning that by taking GII into consideration, we increase the model's goodness of fit. To identify further potential nonlinear relationships, I analyzed the second model with a lowess method and found a similar pattern to Model 1 (decrease between 20% and 40%). (Appendix B2)

$$parliamentseats = \beta_0 + \beta_1 femalesecondaryeducation + \beta_2 femalelabourforce$$

The third model examines how the female secondary education and the female labour force participation rate combined have an overall impact on the seats held in the parliament. (Appendix B3) All covariates are significant at 99% except for β0 which is significant at 95%. β1 shows a one-unit increase in female secondary education rates, and while holding the female labour force rate constant, the model predicts a 8.2% increase in the percentage of seats held by women in parliament. The CI of the β1 is [0.023;0.142]. β2 suggests that for a one-unit increase in the female labour force, and while holding female secondary education constant, the model predicts a 24.1% increase in the percentage of seats held by women in parliament. The CI of the slope is [0.130;0.352], meaning that we can be 95% confident that the true effect of the of the female labour force rate, while holding the female secondary education rate constant, on the percentage of seats held by women in parliament is between 13% and 35.2%. R-squared is higher than the R-squared only considering the female secondary education (0.117), but not higher when we only take the GII into consideration (0.259).

$$parliamentseats = \beta_0 + \beta_1 femalesecondaryeducation + \beta_2 GII + \beta_3 femalelabourforce$$

The fourth model challenges the assumptions about the direct relationships between all the variables and their correlation between parliament seats. (Appendix C1) All covariates are significant at 99%. β1 indicates that the model predicts a 1.4% increase in the percentage of seats held by women in parliament considering female secondary education. The confidence interval of the slope is [-0.259;-0.088] which does not include zero. None of the confidence intervals overlap, indicating a statistically significant difference between the variables. (Appendix C2) This model exhibits the highest R-squared value among all the examined models, indicating the strongest explanatory power in explaining the variation (0.299). To broaden the scope of understanding gender disparities, Appendix D shows a distribution of parliament seats by the Human Development Index (HDI). My expectations were skewed distributions with right long tails, but surprisingly, distributions for all HDI categories are normal, only slightly skewed to the right as the HDI level increases.

4. Probability Calculation

After evaluating the linear probability model (LPM) with predictions exceeding 1 (39.3141), I opted for logit and probit models for more accurate predicted probabilities (Appendix E1). In the logit model, a one-unit increase in female secondary education correlates with a 0.0063 decrease in log odds above the threshold, a statistically significant effect ($p < 0.05$). Similarly, a one-unit increase in GII associates with a substantial 1.5492 decrease in log odds, highly significant ($p < 0.001$). Additionally, a rise in female labor force participation indicates a 0.0051 increase in log odds, also significant ($p < 0.05$). Probit results align closely with logit expectations. Regarding goodness of fit, Brier scores indicate similar model performance, with logit (0.2011) and probit (0.2012) as the best fits. LPM, with a higher Brier score of 659.618, displays a notably worse fit. Log-loss values follow a similar trend, where LPM performs lower than logit and probit models. Prediction values in logit and probit models appropriately fall between 0 and 1. (Appendix E2)

5. Conclusion, Summary & Limitations

To enhance female representation, and considering the strong correlation between secondary education and % of seats, nations should invest and ensure comprehensive and accessible educational opportunities for women across all demographics. To tackle the matter of the significant impact of GII on parliamentary seats, I would suggest targeted policies promoting gender equity in healthcare, workforce participation which could contribute to a more equitable political landscape. In the case of the labour force, - as it has also a significant impact (24.1% increase) - I would recommend providing parental leave, affordable childcare, promoting flexible work arrangements, and providing equal career development opportunities for all genders.

The analysis revealed a strong correlation between female secondary education rates and parliamentary seats held by women, with a 1% increase in education linked to a 7.01% rise in representation. Notably, this relationship diminished between 20% and 40% education rates. Conversely, higher GII values corresponded to a significant decrease (52.114 units) in women's parliamentary representation per GII unit increase. While logit and probit models align closely with actual outcomes, the LPM model underperforms. The study's limitations encompass missing data resulting in incomplete country representation. Additionally, the GII values' interpretation was challenging due to high standard errors (6.409).

Zsófia Rebeka Katona
https://github.com/zsofiarebeka/DA2_Final-term

Appendices

Appendix A1

OLS Regression Results

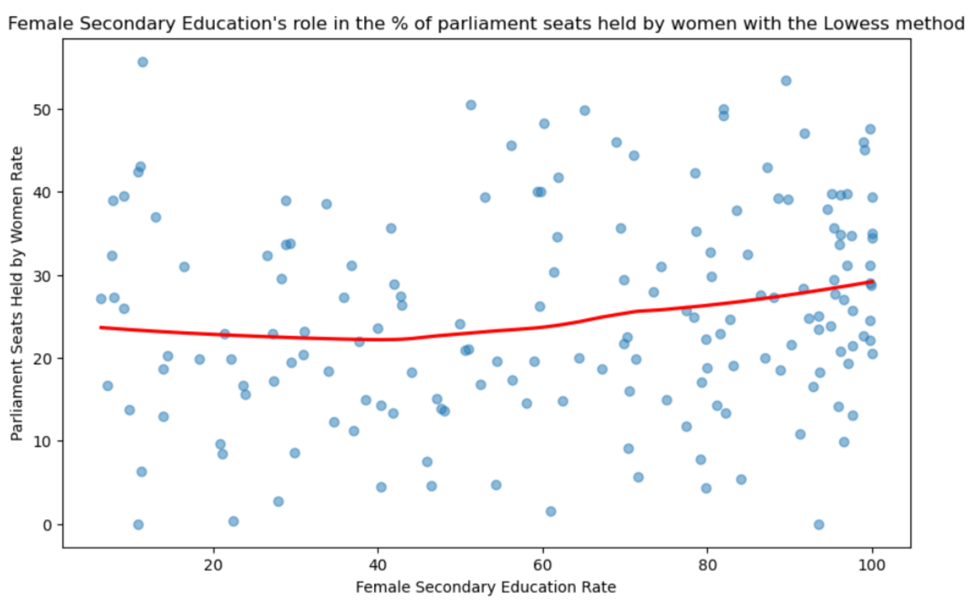
Dep. Variable:	parliamentseatsheldbywomenrate	**R-squared:**	0.029
Model:	OLS	**Adj. R-squared:**	0.023
Method:	Least Squares	**F-statistic:**	4.850
Date:	Fri, 22 Dec 2023	**Prob (F-statistic):**	0.0290
Time:	12:21:14	**Log-Likelihood:**	-665.59
No. Observations:	170	**AIC:**	1335.
Df Residuals:	168	**BIC:**	1341.
Df Model:	1		
Covariance Type:	HC1		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	20.9519	2.270	9.231	0.000	16.503	25.401
femalesecondaryeducation	0.0701	0.032	2.202	0.028	0.008	0.132

Omnibus:	3.489	**Durbin-Watson:**	1.608
Prob(Omnibus):	0.175	**Jarque-Bera (JB):**	3.034
Skew:	0.235	**Prob(JB):**	0.219
Kurtosis:	2.544	**Cond. No.**	159.

Regression table of Model 1

Appendix A2



Scatter Plot of Female secondary education’s role in the % of the parliament seats held by women with the lowess method

Zsófia Rebeka Katona
https://github.com/zsofiarebeka/DA2_Final-term

Appendix B1

OLS Regression Results

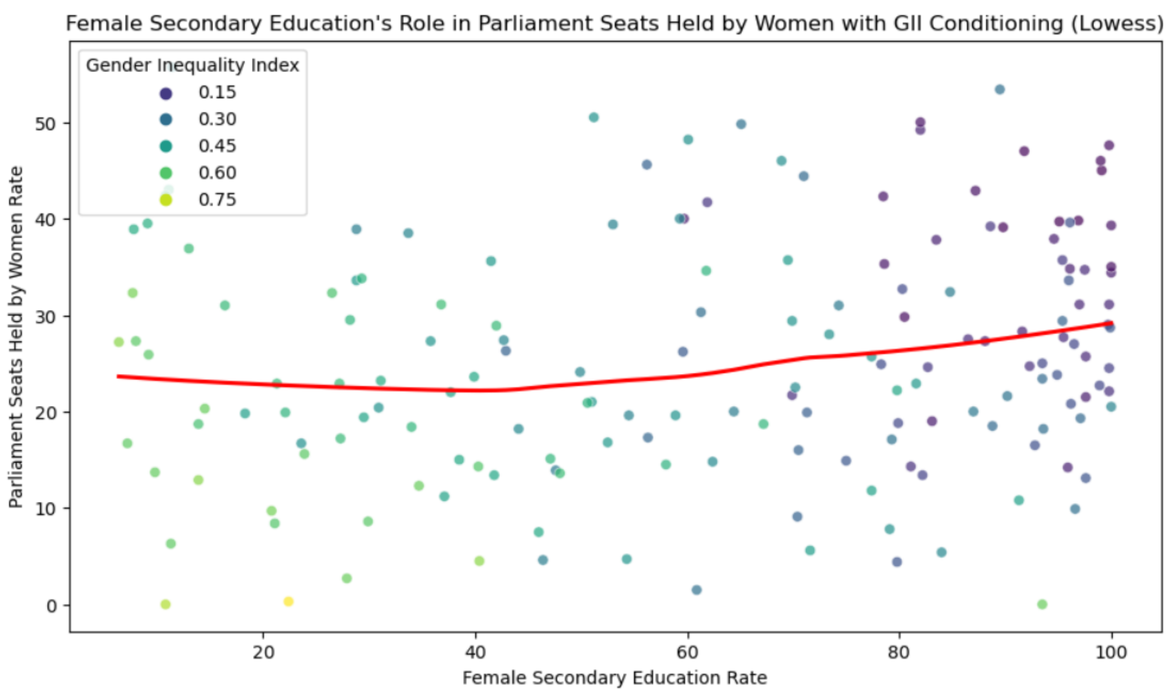
Dep. Variable:	parliamentseatsheldbywomenrate	**R-squared:**	0.267
Model:	OLS	**Adj. R-squared:**	0.259
Method:	Least Squares	**F-statistic:**	40.07
Date:	Fri, 22 Dec 2023	**Prob (F-statistic):**	6.11e-15
Time:	12:21:14	**Log-Likelihood:**	-641.62
No. Observations:	170	**AIC:**	1289.
Df Residuals:	167	**BIC:**	1299.
Df Model:	2		
Covariance Type:	HC1		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	56.1784	4.894	11.479	0.000	46.586	65.770
femalesecondaryeducation	-0.2083	0.044	-4.686	0.000	-0.295	-0.121
gii	-52.1136	6.409	-8.132	0.000	-64.674	-39.553

Omnibus:	2.486	**Durbin-Watson:**	1.635
Prob(Omnibus):	0.289	**Jarque-Bera (JB):**	2.240
Skew:	0.280	**Prob(JB):**	0.326
Kurtosis:	3.053	**Cond. No.**	726.

Regression table of Model 2

Appendix B2



Scatter Plot of Female secondary education’s role in the % of the parliament seats held by women, conditioning on GII with the lowess method

Appendix B3

OLS Regression Results

Dep. Variable:	parliamentseatsheldbywomenrate	**R-squared:**	0.117
Model:	OLS	**Adj. R-squared:**	0.106
Method:	Least Squares	**F-statistic:**	12.37
Date:	Fri, 22 Dec 2023	**Prob (F-statistic):**	9.76e-06
Time:	12:21:14	**Log-Likelihood:**	-657.49
No. Observations:	170	**AIC:**	1321.
Df Residuals:	167	**BIC:**	1330.
Df Model:	2		
Covariance Type:	HC1		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	8.0559	3.617	2.227	0.026	0.966	15.146
femalesecondaryeducation	0.0823	0.030	2.716	0.007	0.023	0.142
femalelabourforce	0.2411	0.057	4.249	0.000	0.130	0.352

Omnibus:	3.931	**Durbin-Watson:**	1.684
Prob(Omnibus):	0.140	**Jarque-Bera (JB):**	3.405
Skew:	0.257	**Prob(JB):**	0.182
Kurtosis:	2.534	**Cond. No.**	352.

Regression table of Model 3

Zsófia Rebeka Katona
https://github.com/zsofiarebeka/DA2_Final-term

Appendix C1

	Dependent variable: parliamentseatsheldbywomenrate			
	femalesecondaryeducation	gii	femalelabourforce	Intercept
	(1)	(2)	(3)	(4)
femalesecondaryeducation	0.070**	-0.208***	0.082***	-0.174***
	(0.032)	(0.044)	(0.030)	(0.044)
gii		-52.114***		-47.066***
		(6.409)		(6.330)
femalelabourforce			0.241***	0.149***
			(0.057)	(0.051)
Constant	20.952***	56.178***	8.056**	44.811***
	(2.270)	(4.894)	(3.617)	(5.936)
Observations	170	170	170	170
R^2	0.029	0.267	0.117	0.299
Adjusted R^2	0.023	0.259	0.106	0.286
Residual Std. Error	12.210 (df=168)	10.636 (df=167)	11.676 (df=167)	10.437 (df=166)
F Statistic	4.850** (df=1; 168)	40.071*** (df=2; 167)	12.372*** (df=2; 167)	30.393*** (df=3; 166)
Note:				*p<0.1; **p<0.05; ***p<0.01

Regression tables of all the models

Appendix C2

OLS Regression Results

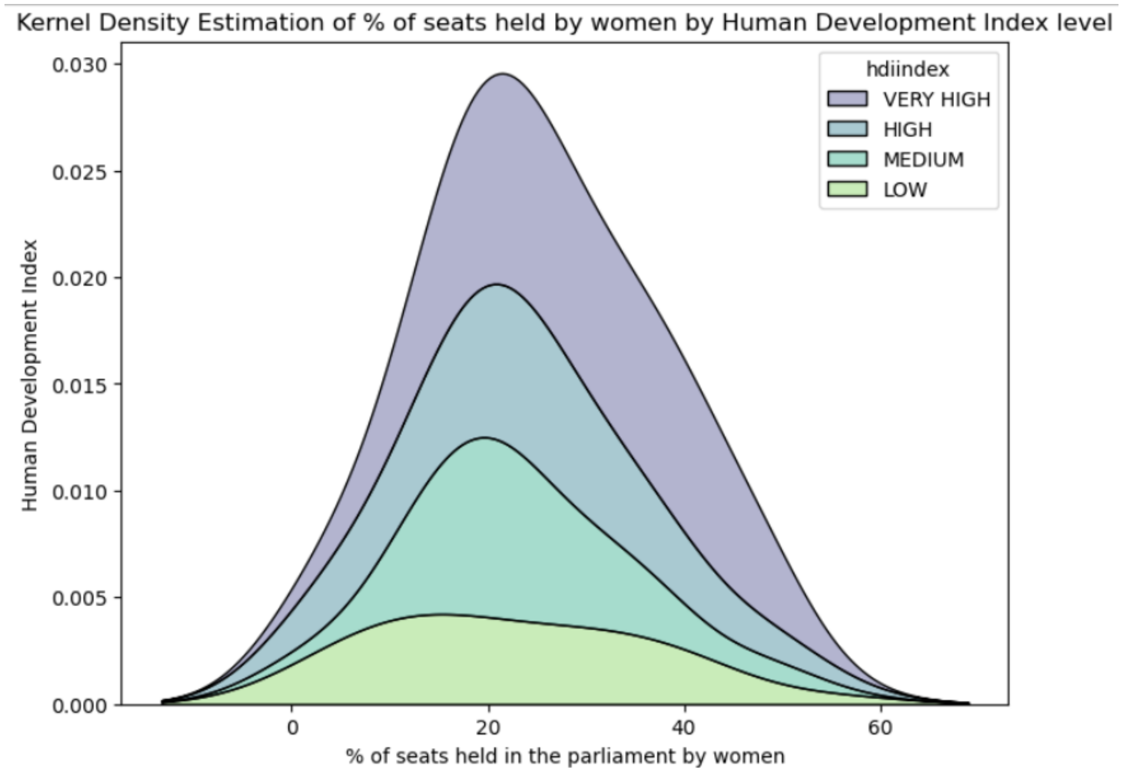
Dep. Variable:	parliamentseatsheldbywomenrate	**R-squared:**	0.299
Model:	OLS	**Adj. R-squared:**	0.286
Method:	Least Squares	**F-statistic:**	30.39
Date:	Fri, 22 Dec 2023	**Prob (F-statistic):**	1.03e-15
Time:	12:33:35	**Log-Likelihood:**	-637.90
No. Observations:	170	**AIC:**	1284.
Df Residuals:	166	**BIC:**	1296.
Df Model:	3		
Covariance Type:	HC1		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	44.8109	5.936	7.548	0.000	33.176	56.446
femalesecondaryeducation	-0.1738	0.044	-3.977	0.000	-0.259	-0.088
gii	-47.0661	6.330	-7.436	0.000	-59.472	-34.660
femalelabourforce	0.1487	0.051	2.924	0.003	0.049	0.248

Omnibus:	3.397	**Durbin-Watson:**	1.668
Prob(Omnibus):	0.183	**Jarque-Bera (JB):**	3.014
Skew:	0.316	**Prob(JB):**	0.222
Kurtosis:	3.162	**Cond. No.**	978.

Regression table of Model 4

Appendix D



Kernel Density Estimation of % of seats held by women by Human Development Index level

Appendix E1

```
count    170.0000
mean      25.3035
std        6.7512
min        3.2165
25%       20.7910
50%       25.4160
75%       30.2688
max       39.3141
Name: pred_lpm4, dtype: float64
```

Model 4's LPM prediction values (the max value goes above 1)

Appendix E2

	LPM	Logit	Probit
Brier-score	659.6177	0.2011	0.2012
Log-loss	-19.2940	-0.5855	-0.5852

Goodness of fit scores