International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Coffee Cups, Brain Drain to Nobel Medals: Analyzing Factors Influencing Nobel Laureates per Million Using OLS Multiple Regression

Atharva Joshi[1], Ayan Mandal[2], Debanjali Dutta[3], Rohit Shelar[4], Subha Suresh Kumar[5]

Indian Institute of Technology, Kanpur, India
Email: *rohitsh24[at]iitk.ac.in*

Abstract: ***This study investigates the factors influencing the number of Nobel Laureates per million population across various countries. Using data from 62 countries, we examine the impact of key variables including R&D expenditure as a percentage of GDP, educational outcomes (PISA scores), coffee consumption per capita, the percentage of women in parliament, and brain drain. Employing Ordinary Least Squares (OLS) regression, our analysis reveals that both R&D expenditure and coffee consumption are statistically significant predictors of Nobel laureates per million. Notably, higher R&D spending correlates with more Nobel laureates, though this effect is attenuated by high levels of brain drain, which reduces the effectiveness of R&D investments. Coffee consumption, while also significant, may reflect broader socio - cultural factors related to intellectual engagement and productivity. The study highlights the need for comprehensive policies that not only promote research and development but also address talent retention to maximize scientific achievements. Future research should explore these dynamics further, considering more complex interactions and broader datasets to enhance our understanding of the factors contributing to Nobel laureateship.***

Keywords: Nobel laureates/prize, coffee consumption per capita, brain drain, regression

1. Introduction

The Nobel Prizes are the most prestigious global awards, recognizing outstanding contributions to humanity in the fields of Physics, Chemistry, Medicine, Literature, Peace, and Economic sciences. (The Swedish Nobel Prize, Sweden. se)

Nobel Prize awards represent the highest standard of excellence in our time. Each year, the Nobel Foundation in Sweden diligently seeks out the most exceptional individuals who have significantly benefited humanity. The Foundation strives to maintain impartiality in its selection process, often overcoming political pressures and influences to uphold its rigorous standards (Modis, 1988).

Receiving a Nobel Prize is considered a hallmark of scientific success" (Gallotti & De, 2019), and winning a Nobel Prize is not only a mark of individual brilliance but also a reflection of the environment that nurtured such achievement. This study seeks to evaluate how well various socio - economic, educational, cultural, and political factors explain the number of Nobel Laureates per million population across different countries.

Messerli (2012) highlights the cognitive benefits of dietary flavonoids, particularly flavanols in cocoa, green tea, red wine, and certain fruits, which include reduced dementia risk and improved cognitive function. This raises the question of whether a population's cognitive capacity, as reflected by Nobel laureates per million, could correlate with coffee consumption. Prinz (2020) explores this further, testing the correlation between chocolate consumption and Nobel laureates per capita using a two - stage Heckman model that controls for variables like scientific paper output, R&D expenditures, GDP per capita, and coffee and tobacco consumption. Despite adding these factors, the positive correlation between chocolate consumption and Nobel laureates persists. Additionally, Prinz finds a negative correlation between coffee consumption and Nobel laureates, though no clear causal relationship is established, suggesting possible hidden variables or chance factors. Dutton, te Nijenhuis, and Roivainen (2014) suggest that the traits associated with high PISA scores in Finland—lower psychoticism and average intelligence—might also explain the country's relatively few Nobel laureates in science. They propose that while these traits contribute to strong educational performance, they may not support the personality traits or intellectual curiosity needed for major scientific breakthroughs.

Our study focuses on key independent (explanatory) variables, including R&D expenditure as a percentage of GDP, Learning Outcome of Students in Science, Maths and Reading (PISA scores), Coffee consumption per capita, the percentage of Women in Parliament, and the Human Immigration (Flight and Brain Drain). Each of these variables represents a different dimension of a country's environment—whether it be investment in innovation, the quality of education, cultural habits, political inclusivity, or talent retention.

Through statistical analysis, we aim to understand the relationship and interplay between these variables in producing Nobel Laureates for the countries. The study will assess the strength and significance of these relationships that contribute to a nation's intellectual output.

The objective of this study is to explore and analyze the factors that influence the number of Nobel Laureates produced per million population across different countries. By examining the relationships between the dependent variable—Number of Nobel Laureates per million population—and various independent variables, we aim to uncover insights into how different socio - economic,

educational, cultural, and political factors contribute to a nation's success in producing Nobel Laureates.

2. Methodology

This study employs a cross - sectional quantitative approach using data from multiple countries. The dependent variable is the number of Nobel Laureates per million population for these 62 countries. Independent variables include R&D expenditure as a percentage of GDP, Learning Outcomes, Coffee consumption per capita, percentage of Women in Parliament, and the Human Flight and Brain Drain Index. Data is sourced from the latest available records, with minor variations in collection periods.

The analysis utilizes Ordinary Least Squares (OLS) multiple regression to examine the relationship between the dependent variable and the independent variables. Diagnostic tests, including multicollinearity and heteroscedasticity checks, are conducted to ensure model validity and result robustness.

3. Data Description

Dependent Variable: ***Number of Nobel Laureates per Million Population***

The dependent variable in this study is the number of Nobel Laureates per Million Population. By normalizing the number of laureates according to population size, this variable allows for a more equitable comparison across countries with differing population and size scales.

The Nobel Prizes are globally recognized symbols of exceptional intellectual and humanitarian contributions, often reflecting both individual excellence and the broader socio - cultural and economic contexts that foster such achievement. Therefore, this variable serves as a proxy for a country's overall intellectual and creative output, considering both the environment that supports high - level innovation and the talent cultivated within the nation. This study seeks to determine how well various independent variables—spanning socio - economic, educational, cultural, and political dimensions—explain the variance in this dependent variable across different countries.

The independent variables selected for this study represent different aspects of a country's resources (economic and intellectual), socio - economic conditions, and cultural dynamics. The objective was to select variables that are not only independently and identically distributed but also collectively cover various realms of influence to address the omitted variable bias in estimating a country's potential to produce Nobel laureates over the years.

Independent Variable (Regressor)	Realm of Coverage/ Influence	Description/Assumptions
R&D Expenditure (% of GDP)	Economic Resources and Innovation	Research and Development (R&D) expenditure as a percentage of GDP reflects a country's investment in scientific research and infrastructure, technological innovation, and intellectual development. Higher R&D spending typically indicates a greater emphasis on generating new knowledge and technologies, which are crucial for groundbreaking discoveries. This covariate seems plausible since research in fields like physiology/medicine, physics, and chemistry is costly and heavily reliant on state funding. (Prinz, 2020)
Learning Outcomes of 15 - year - old Students in Maths, Science, and Reading	Education and Human Capital Development	Learning outcomes of 15 year old students as measured by the The Programme for International Student Assessment (PISA) scores serve as a proxy for the quality of a country's education, robust educational foundations, and intellectual capital. High scores indicate a strong foundation in critical thinking, problem - solving, and scientific literacy, which are deemed essential for scientific discoveries.
Yearly Coffee Consumption per Capita (kg)	Socio - Cultural and Lifestyle	Coffee consumption is an intriguing variable that can act as a proxy for various socio - cultural factors, work habits, and intellectual engagement. This variable offers a unique perspective on lifestyle factors related to socializing, work, and productivity, which may influence creativity, innovation, and idea exchange. Also, caffeine, the primary active compound in coffee, is a psychomotor stimulant that notably impacts the brain's dopaminergic system (Fisone et al., 2004; Fredholm et al., 1999; Cauli & Morelli, 2005; Lee & Kim, 2019).
Women in Parliament (% of Women in the Lower House)	Socio - Political and Gender Parity	The percentage of women in parliament reflects the level of gender equality and inclusiveness in a country's politico - economic system. Gender inclusiveness in governance can reflect broader societal attitudes towards equality, more progressive policies, and support of diverse viewpoints, which can lead to more holistic and innovative problem - solving approaches and general welfare.
Human Flight and Brain Drain	Migration, Talent Retention, and Globalization	The Human Flight and Brain Drain Index measures the extent to which skilled professionals, academics, and intellectuals leave their home country in search of better opportunities abroad, potentially due to pursuit of an environment that is conducive to their intellectual aspirations. A high brain drain index suggests that a country might be losing its top intellectual talent, highlighting the importance of retaining intellectual capital within a country. It also reflects socio - politico - economic stability and the ability to retain talent, which are crucial for sustained intellectual contributions in an economy.

Note: Data for each of these variables has been collected from the most recent available sources. (2022 - 2023) However, it is important to note that due to varying periods of data collection, there may be some time lag in a few of the

attributes. This variation is recognized, but it is less concerning in the context of the study, as the dependent variable—the number of Nobel laureates per million people—is cumulative. This variable reflects achievements accumulated over many years per million of the population, and its relative nature helps mitigate the impact of these time lags.

Correlation Analysis

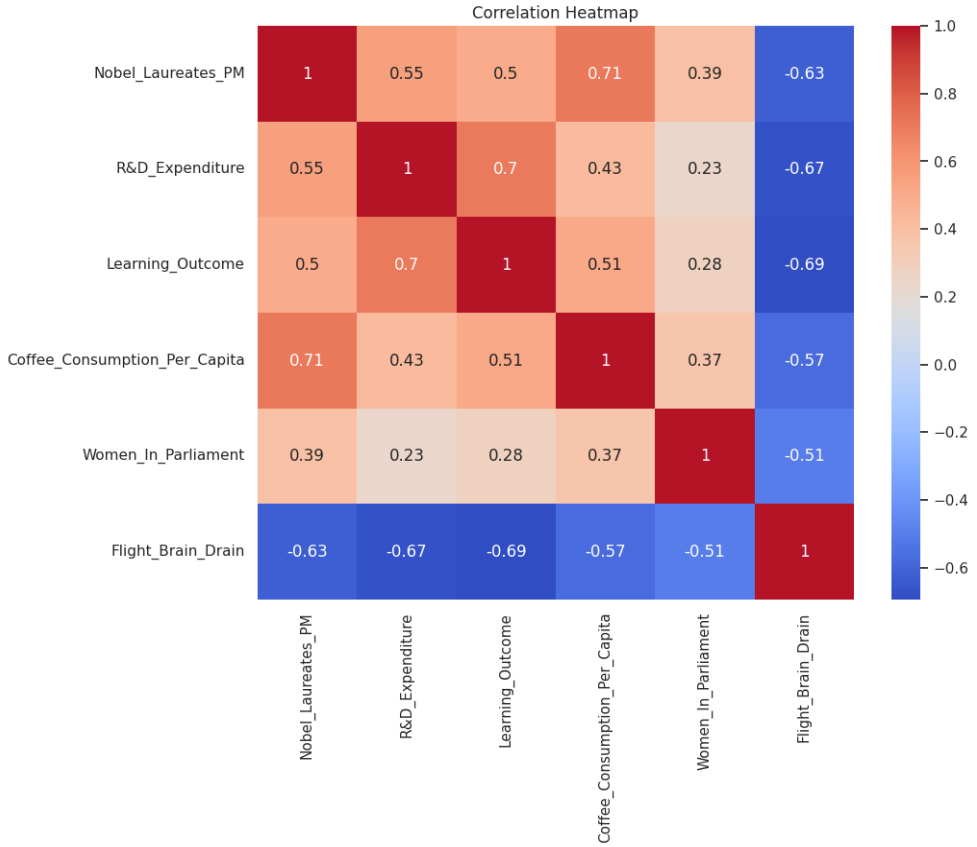

Figure 1: Correlation heatmap between the model variables

A persistent moderate correlation could be observed between the Nobel Laureates per million and its regressors (0.4 - 0.7). Also, between the regressors/predictors, low to moderate correlation is observed. (0.2 - 0.7). This warrants us to check for multicollinearity before moving ahead with model training.

For OLS regression to provide valid causal inferences, specific assumptions must be met, especially when analyzing multiple regressors (Stock & Watson, 2020)

4. Results and Interpretation

Model Summary

Note: Robust standard Errors (HC3) are used in all models after detection of heteroscedasticity. We also observe that the models have a very moderate multicollinearity.

Statistic	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
No. Observations	62	62	62	62	62	62
R - squared	0.306	0.426	0.58	0.594	0.599	0.608
Adj. R - squared	0.295	0.406	0.566	0.573	0.578	0.573
F - statistic	11.02	16.19	20.36	13.3	16.86	10.6
Prob (F - statistic)	0.00153	2.48e - 06	1.89E - 07	1.02E - 06	5.41E - 08	3.41E - 07
const	0.1182	1.2441	-0.213	-0.4482	0.3273	0.8023
R&D Expenditure (coef)	0.3904***	0.168	0.2144**	0.2062**	0.1374	0.1791
Flight Brain Drain (coef)	-	- 0.2085 **	-	-	-0.0925	-0.089
Coffee Consumption (coef)	-	-	0.1390***	0.1288***	0.1218***	0.1231***
Women in Parliament (coef)	-	-	-	0.0097	-	0.0056
Learning Outcome (coef)	-	-	-	-	-	-0.0016

Significance Levels

*Coefficients with *** (three asterisks) are highly significant ($p < 0.01$).*
*Coefficients with ** (two asterisks) are significant at the 5% level ($p < 0.05$).*
*Coefficients with * (one asterisk) are significant at the 10% level ($p < 0.1$).*
No asterisk: The result is not statistically significant (p - value ≥ 0.1).

Key observations–

a) Explanatory power with three or more different combinations of the chosen regressors stagnates to around 60%

b) As observed, the effect size (magnitude) and the significance level of the coefficients for coffee consumption per capita remain consistent even after accounting for other independent variables. This consistency suggests that the relationship is unlikely to be spurious. Across different models, coffee consumption alone could explain the same portion of the variance in the number of Nobel laureates produced per million, even when other predictors are held constant.

Hence, either this observed relationship could be coincidental or driven by omitted variables that are not included in the model but are correlated with both coffee consumption and Nobel laureates. Therefore, further investigation is needed to determine whether the relationship is causal or if it is due to some other underlying factors not accounted for in the current models.

Coffee consumption, as hypothesized during the selection of independent variables, may act as a proxy for various socio - cultural factors, work habits, and intellectual engagement, providing insight into lifestyle aspects related to social interactions, work routines, and productivity that could impact creativity, innovation, and the exchange of ideas. However, we caution about direct causation and warrant further study here.

c) **Interesting Case of Brain Drain**
The inclusion of Flight Brain Drain in Model 2 introduces a significant negative coefficient (- 0.2085), which is statistically significant at the 5% level & also improves the model's explanatory power by about 12 percentage points. While R&D Expenditure no longer has a strong, independent relationship with the number of Nobel Laureates per Million.

When we introduce *Flight Brain Drain* into the model (as in Model 2), it competes for the variance that was previously attributed to *R&D Expenditure*. If *Flight Brain Drain* is a confounder, it better captures the true relationship with the dependent variable. As a result, *R&D Expenditure* becomes less significant because the model now recognizes that *Flight Brain Drain* is a more relevant predictor.

The reduction in the R&D Expenditure coefficient from 0.3904 to 0.1680 when Flight Brain Drain is included suggests that the positive effect of R&D spending is partly mitigated by the loss of talent. (Omitted variable Bias produced a large estimated coefficient in model 1)

R&D Expenditure remains positive but less influential when brain drain is considered, suggesting that even countries with comparable R&D investments might not be able to convert that spending into scientific excellence if they face high levels of emigration amongst their most talented researchers. (Omitted variable bias in model 1 is partly addressed by Brain drain in model 2)

This brings us closer to accounting for multiple dimensions (both investment and retention of top talent) when modeling complex outcomes like Nobel laureateship.

The comparison between the two models illustrates the importance of considering both investments in R&D and the retention of talent. While R&D expenditure is undoubtedly a driver of scientific output, its effectiveness can be significantly reduced if a country experiences high levels of brain drain. This insight emphasizes the need for balanced policies that not only foster innovation but also retain and attract top - tier talent.

Higher R&D + Lower Brain Drain (implies) Higher Attraction of Top Intellectual Talent

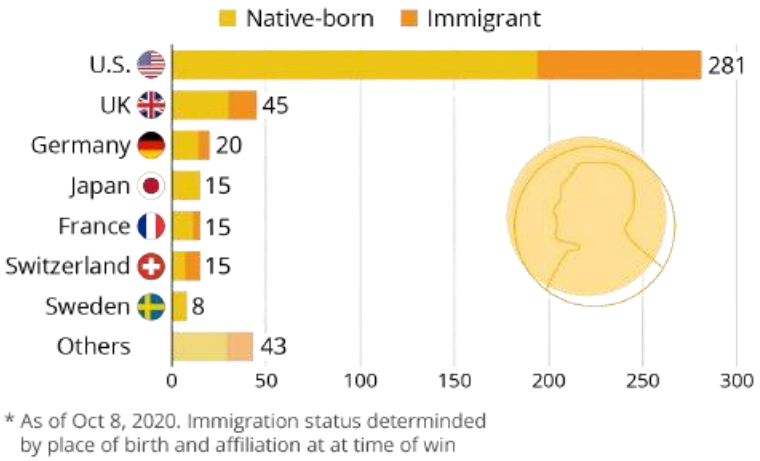

Figure 2: Number of Nobel laureates in physics, chemistry, medicine and economy per country (1969 - 2020) (Source: Nobel Prize Foundation)

The figure factually corroborates the above findings, as we can observe that the share of immigrants in the prize winners have been significant.

In summary, the lack of statistical significance for *R&D Expenditure* suggests that its influence on Nobel Laureates per Million is not strong enough, or is confounded by other

factors like *Flight Brain Drain*, reducing its relevance as an independent predictor in the current model.

Other independent variables, such as the learning outcomes of 15 - year - olds and the percentage of women in parliament, contributed minimally to explaining the variation in Nobel laureateship.

5. Limitations and Considerations

The model does not necessarily establish a causal relationship between the independent variables and the number of Nobel Laureates. (Correlation is not causation.)

For instance, while coffee consumption is statistically significant, it might be capturing other socio - economic characteristics rather than having a direct effect on Nobel Prizes.

Also, we must acknowledge that more complex interactions or non - linear relationships might exist between the variables. For example, the effect of R&D expenditure could vary depending on the country's level of economic development or cultural factors as well as its human capital.

Furthermore, with only 62 observations, the model might suffer from limited generalizability and limitation on the assumption 2 (**i. i. d.).** Small sample sizes can lead to high variance in estimates and less reliable inference.

Cook's Distance measures the influence of each data point on the overall regression model. A higher Cook's Distance indicates a data point that significantly affects the model's estimates. Top five countries [with threshold: 0.07] are>> Finland (0.4510), Sweden (0.1312), Switzerland (0.0944), United Kingdom (0.0903), and Israel (0.0757). These data points need further investigations to check for reliability.

6. Conclusion

This study explored the factors influencing the number of Nobel Laureates per million population across various countries, focusing on variables such as R&D expenditure, educational outcomes, coffee consumption, gender parity in politics, and brain drain. The analysis found that R&D expenditure and coffee consumption per capita are statistically significant predictors of Nobel laureates, highlighting their role in fostering environments conducive to scientific excellence. Specifically, while higher R&D spending is associated with a greater number of Nobel laureates, this relationship is mitigated by brain drain, which diminishes the potential impact of R&D investments.

Coffee consumption, though intriguing, seems to capture broader socio - cultural factors that may contribute to intellectual engagement and innovation. This suggests that while lifestyle factors like coffee consumption could offer insights into creative environments, the exact nature of this relationship requires further investigation.

Future research should continue to explore these relationships, incorporating more complex interactions and larger datasets to refine our understanding of the variables influencing Nobel laureateship. This ongoing investigation will contribute to a deeper comprehension of how various socio - economic, educational, cultural, and political factors collectively shape a nation's capacity to produce Nobel laureates.

References

[1] Aloys Leo Prinz, Chocolate consumption and Nobel laureates, Social Sciences & Humanities Open, Volume 2, Issue 1, 2020, 100082, ISSN 2590 - 2911, https: //doi. org/10.1016/j. ssaho.2020.100082
[2] B. B. Fredholm, K. Bättig, J. Holmén, A. Nehlig, E. E. Zvartau Actions of caffeine in the brain with special reference to factors that contribute to its widespread use; Pharmacological Reviews, 51 (1) (1999), pp.83 - 133
[3] Dutton, E., te Nijenhuis, J., & Roivainen, E. (2014). Solving the puzzle of why Finns have the highest IQ, but one of the lowest number of Nobel prizes in Europe. Intelligence (Norwood), 46, 192–202. https: //doi. org/10.1016/j. intell.2014.06.006
[4] Gallotti, R., De Domenico, M. Effects of homophily and academic reputation in the nomination and selection of Nobel laureates. Sci Rep 9, 17304 (2019). https: //doi. org/10.1038/s41598 - 019 - 53657 - 6
[5] G. Fisone, A. Borgkvist, A. Usiello, Caffeine as a psychomotor stimulant: Mechanism of action; CMLS Cellular and Molecular Life Sciences, 61 (2004), pp.857 - 872
[6] J. - B. Lee, T. - W. Kim; Ingestion of caffeine links dopamine and 5 - hydroxytryptamine release during half immersion in 42°C hot water in [a] humans; Journal of Exercise Rehabilitation, 15 (4) (2019), pp.571 - 575
[7] Messerli, F. H. (2012). Chocolate consumption, cognitive function, and Nobel laureates. N Engl J Med, 367 (16), 1562 - 1564.
[8] Modis, T. (1988). Competition and forecasts for Nobel prize awards. Technological Forecasting & Social Change, 34 (2), 95–102. https: //doi. org/10.1016/0040 - 1625 (88) 90059 - 5
[9] Stock, J. H., & Watson, M. W. (2020). Introduction to econometrics (4th ed.). Pearson.
[10] The Swedish Nobel Prize: Great minds think differently. https: //sweden. se/work - business/study - research/the - swedish - nobel - prize
[11] O. Cauli, M. Morelli; Caffeine and the dopaminergic system; Behavioural Pharmacology, 16 (2005), pp.63 - 77

Data set

[12] Link to the original compiled Dataset: https: //docs. google. com/spreadsheets/d/1UxyCa4XjYJK4_EgFrEznx5SHeuSadX29lPckEO - _Lmc/edit?gid=2020510912#gid=2020510912