Exploring the Impact of Public Data Transparency on Economic Complexity Index Growth Among Countries: A Comparative Analysis from 1995 to 2022

Mihir Kulshreshtha
Thomas Jefferson High School for Science and Technology
Research & Statistics 2
Dr. Catherine Scott
May 29, 2024

In a world driven by data, it is vital to understand the importance of decision making behind data governance. Data is powerful, and impactful decisions must be made about data transparency, such as who has access to what kinds of data. This study focuses on exploring the role of governmental data transparency in influencing economic stability and growth, a topic that is increasingly critical in the digital age. The Economic Complexity Index (ECI) offers a valuable tool to assess how different nations harness and utilize data, potentially driving diverse economic outcomes such as growth, income levels, and inequality. This study draws upon seminal research in the fields of economic complexity and data science to evaluate the potential correlation between public data transparency and economic complexity growth from 1995 to 2021.

In his work, "Economic complexity theory and applications," César A. Hidalgo (2021) describes the foundational methodologies for analyzing economic performance across nations. Hidalgo explains how economic complexity can be used to predict various economic outcomes, like income, growth, and inequality. Using dimensionality reduction techniques, Hidalgo shows how vast economic indicators and data can be reduced into interpretable forms, like ECI. His research lays the foundation for my study, as he explains the mathematical approach behind ECI, which I used as the response variable to measure countries economic prosperity. Complementing Hidalgo's theoretical framework, Liran Einav and Jonathan Levin in "Economics in the age of big data" (2014) discuss the transformative impact of big data on economic research. Their insights highlight the evolving nature of economic analysis, which is pertinent to this study's focus on the implications of transparent data handling for economic complexity (Einav & Levin, 2014).

Continuing the exploration of data transparency's impact on economic outcomes, the work of George Shambaugh and Elaine Shen offers pivotal insights into the practical effects of governmental openness during economic crises. In their study, they investigate how increased transparency of national governments—not just central banks—can significantly alter the duration and severity of economic downturns. According to Shambaugh and Shen, enhanced government transparency provides critical macroeconomic information that reduces information asymmetries, enhances the predictability of policy commitments, and boosts public confidence in policy effectiveness (Shambaugh & Shen, 2021).

Their findings suggest that transparency not only aids in shortening the duration of inflation and currency crises but also mitigates the severity of these downturns. This perspective is particularly relevant to this study, as it underscores the broader implications of open data policies beyond central banking practices and extends into the overall governmental strategy.

By integrating the theoretical frameworks of Hidalgo and Einav & Levin with the empirical findings of Shambaugh and Shen, this study aims to provide a comprehensive analysis of how transparency in data handling can influence economic conditions. This research is driven by the hypothesis that countries with higher levels of transparency in their data handling are likely to exhibit more substantial economic complexity growth, indicative of a dynamic and improving economy. More than just an academic pursuit, this study has practical interpretations, as understanding the impact of data transparency policies on economic growth can be extremely useful for policymakers and the financial markets. By exploring this relationship, this study aims to contribute to broader discourse on how data governance can shape economic conditions.

Methods

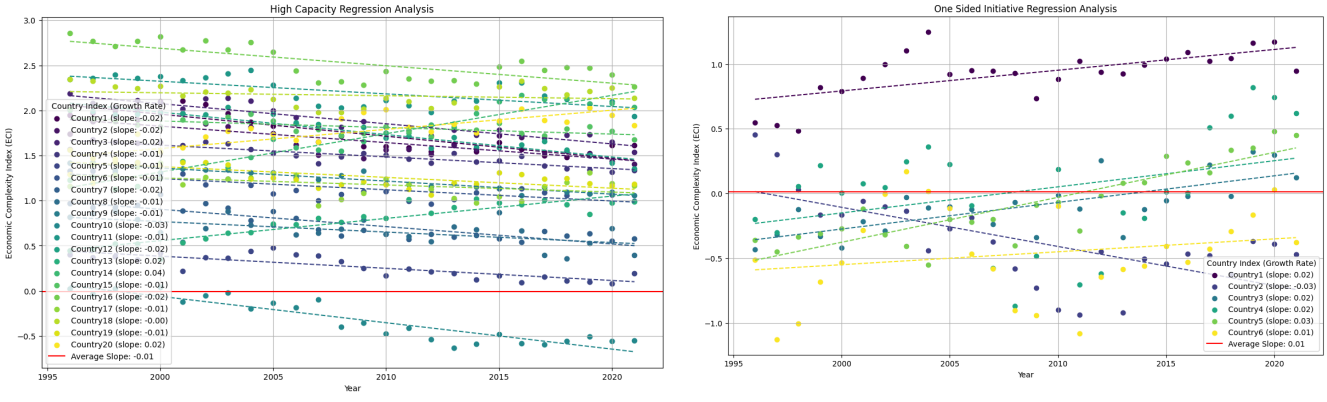
The study utilized data from two principal sources: the Harvard Atlas of Economic Complexity and the Open Data Barometer. These databases provided comprehensive metrics on economic complexity and public data transparency respectively, which are pivotal to understanding the relationship between governmental transparency and economic growth.

The countries analyzed in this study were segmented into four groups based on their data transparency ratings according to the Open Data Barometer: High Capacity, Emerging & Advancing, Capacity Constrained, and One-Sided Initiatives. High capacity countries all were determined to have established open data policies, generally with a strong political backing. Emerging & Advancing countries have emerging or established open data programs, but still face challenges before open data is mainstream and institutionalized as a sustainable practice. Capacity Constrained countries face challenges in establishing sustainable open data initiatives, often due to limited governments or private sectors. Finally, One-Sided Initiatives are countries that have some form of open data initiative, but lack government action and political freedom to allow large-scale, impactful data transparency. This classification facilitated a comparative analysis across varying levels of government data openness.

The Economic Complexity Index (ECI) data for each country from 1995 to 2022 was sourced from the Harvard Atlas of Economic Complexity. ECI is a measure that reflects how diverse a country’s export structure is, providing insights into its productive capabilities and economic complexity. The index is calculated based on the diversity and amount of a country's export products, highlighting the potential of a country's economy to adapt and innovate. More technically, ECI is derived from a mathematical equation for diversity and ubiquity, seen below (Hidalgo & Hausmann, 2009).

$$
\begin{aligned}
k_{c,n} &= \frac{1}{k_{c,0}} \sum_{p} M_{cp} \frac{1}{k_{p,0}} \sum_{c'} M_{c'p} k_{c',n-2} \\
&= \sum_{c'} k_{c',n-2} \sum_{p} \frac{M_{c'p} M_{cp}}{k_{c,0} k_{p,0}} \\
&= \sum_{c'} k_{c',n-2} \tilde{M}^{C}_{c,c'}
\end{aligned}
$$

The analytical procedure began with manually clustering and processing the raw data into four different spreadsheets, one for each data transparency group. Then, using Jupyter Notebook and Python's Sci-Kit Learning package, regression lines were produced for each country's ECI growth over the observed period. Regression analyses were conducted separately for each cluster. Specifically, the study focused on comparing the regression slopes of the two most contrasting clusters: 'High Capacity' (countries with open data policies) and 'One Sided Initiative' (countries with restrictive data practices), as seen in Figures 1.1 and 1.2 below. With the slopes in the two clusters, a two-sample t-test was used to compare the slopes and to investigate if the differences were statistically significant. Prior to conducting the t-test, a check for normality was performed to ensure the validity of the test results.



Figures 1.1 and 1.2

Since the analysis involved not only regressions but also t-tests, conditions must be checked. While it is not feasible to assume all the countries ECI growth are independent, as geopolitical trade amongst countries influences economic growth, when observing the residuals of each country's residuals, most seem to be evenly distributed without any patterns. However, the normality condition, crucial for the validity of t-tests, was found to be partially unmet. This suggests a deviation from the ideal assumptions of the test. The study design can be seen in the diagram in Figure 1.3, below.

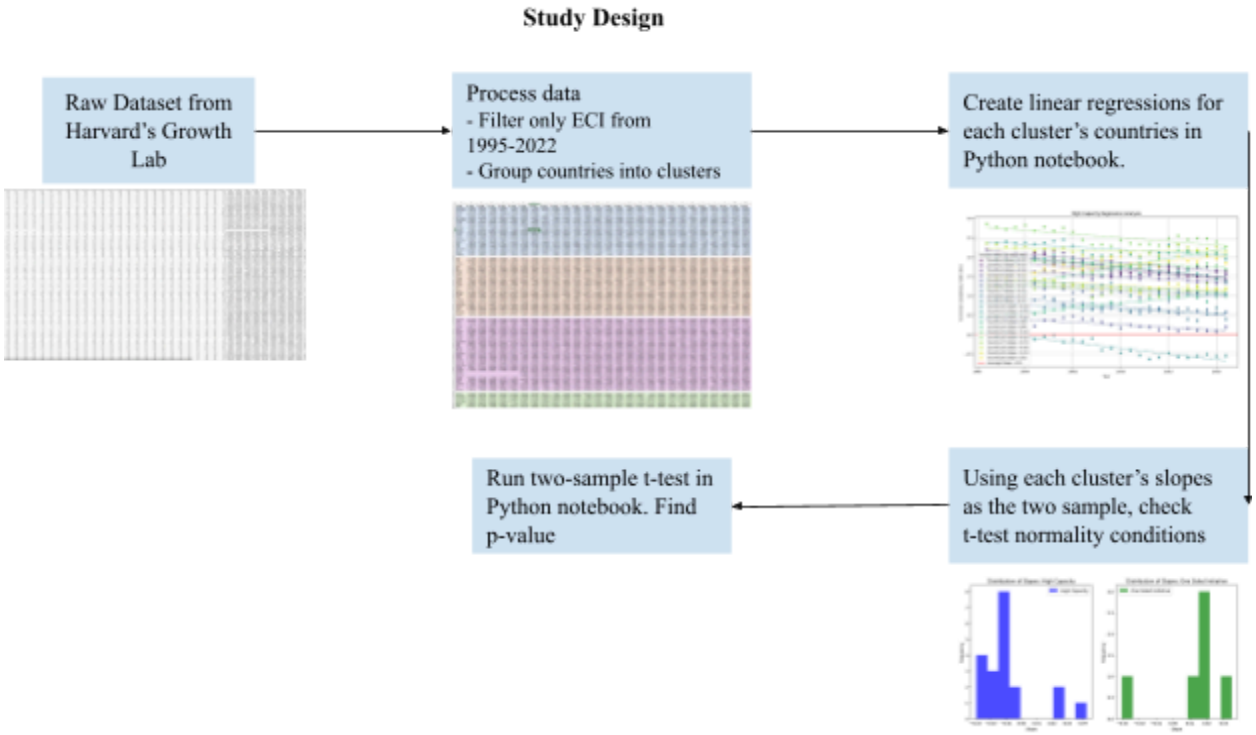


Figure 1.3

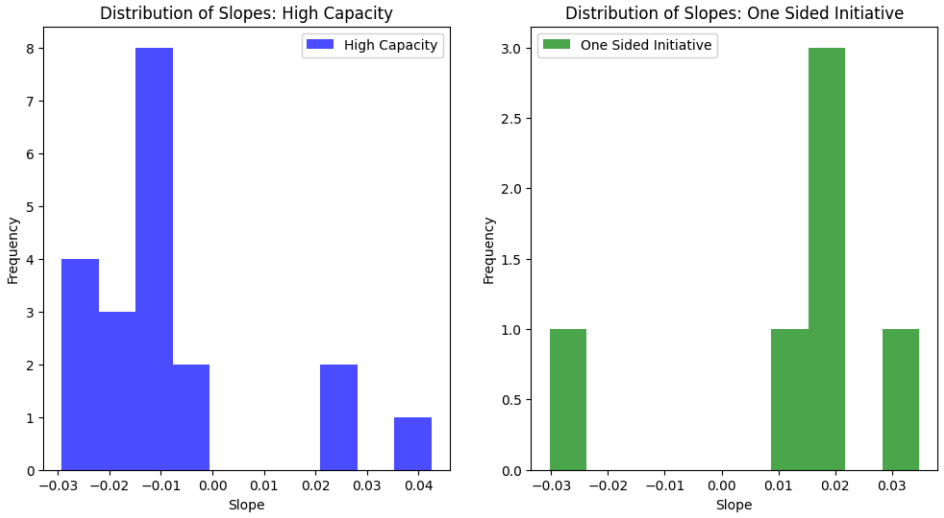
Results

The results of the study provide a compelling insight into the relationship between public data transparency and economic complexity growth from 1995 to 2022.

The regression analysis for the "High Capacity" cluster, as shown in Figure 1.1, displays a general trend where most countries have a slight negative slope, indicating a minor decline in

economic complexity over time. However, the average slope for this cluster is close to zero, suggesting stability rather than a marked decrease. The "One-Sided Initiative" cluster, depicted in Figure 1.2, generally shows slightly positive slopes, with an average slope indicating minimal growth in economic complexity. This is initially counterintuitive to my initial prediction that countries with more open data policies would have a higher growth over time, and vice versa for countries with more restrictive policies. However, One-Sided Initiative countries tend to be developing third world countries that have been experiencing a faster rate of growth recently than well established countries in the High Capacity category, which could explain this trend.

The histograms to the right (Figure 2.1) detail the distribution of slopes for the two clusters, revealing distinct patterns. The "High Capacity" cluster shows a broader range of slope values, whereas the "One Sided Initiative" has a concentration around a smaller range of values, particularly noting a peak in positive slopes.



The table below shows the results of running statistical tests on the two clusters, with p-values under the given test.

Normality Test (High Capacity Cluster)	Normality Test (One-Sided Initiative Cluster)	Levene's Test for Equality of Variances	Two-Sample T-test

0.00056	0.101	0.758	0.0275

Table 2.1

The normality test results indicate that the "High Capacity" cluster did not pass the normality test (P-value = 0.00056), suggesting that the data distribution deviates significantly from a normal distribution. On the other hand, the "One Sided Initiative" cluster showed a P-value of 0.101, which does not reject the hypothesis of normality.

The Levene's Test for Equality of Variances yielded a P-value of 0.758, suggesting that there is no evidence of unequal variances between the two clusters, thus fulfilling one of the prerequisites for running a two-sample t-test.

Two-Sample T-Test: The t-test between the two clusters resulted in a P-value of 0.0275, indicating a statistically significant difference between the slopes of the "High Capacity" and "One Sided Initiative" clusters. This leads us to reject the null hypothesis, which posited no difference between the groups. The alternative hypothesis, suggesting that there is a significant difference in economic complexity growth between countries with different levels of public data transparency, is thus supported. However, the test's result must be realized cautiously, as the normality condition was not met for the High Capacity cluster.

The results suggest a nuanced relationship between public data transparency and economic complexity growth. Despite the slight deviation from normality in the "High Capacity" cluster, the significant result from the t-test suggests that differences in data transparency levels are indeed associated with variations in economic complexity growth among countries.
These findings are crucial for understanding how transparent governance practices can influence economic outcomes and support the need for further research into specific mechanisms through which transparency can affect economic structures.

Discussion

The findings of this study offer valuable insights into the influence of public data transparency on economic complexity. By examining the growth of the Economic Complexity Index (ECI) from 1995 to 2022 across countries with varying degrees of transparency, the research highlights a discernible pattern: there is a significant distinction in economic complexity growth between countries with different transparency levels

The results corroborate the assertions made in the literature reviewed in the Rationale section. Studies by Shambaugh & Shen (2018) suggest that transparency fosters economic innovation and complexity. Specifically, the context of the "One Sided Initiative" cluster, where slightly positive growth in ECI suggests that even limited transparency can have a positive impact on economic complexity. This supports the theory that transparency can attract foreign investment and improve domestic production, hinting that openness might indeed enhance economic complexity.

The significant difference in ECI growth trends between the "High Capacity" and "One Sided Initiative" clusters suggests that transparency in data handling and governance might play a crucial role in economic development. However, these findings should be interpreted cautiously due to some limitations in the data's statistical properties. The failure to fully meet the normality condition, especially in the "High Capacity" cluster, raises concerns about the robustness of the t-test results. This deviation suggests that the data might not perfectly represent the underlying population, which could affect the reliability of inferential statistics used in this study. In other words, the findings from this t-test might not extend to any country with a given data policy.

Given that the findings seem to reject the normality assumption, my future research could explore alternative statistical methods that are less sensitive to such deviations, such as

non-parametric tests. Additionally, further studies could focus on a more granular analysis of the types of data transparency and their direct effects on specific sectors of economic complexity, potentially offering a clearer picture of these dynamics. For example, future research could involve studies that track changes in transparency and economic complexity over time within the same countries. Such studies could better isolate the relationship between data transparency and economic growth, eliminating the effects of other concurrent economic variables in my current study.

As improved economic indicators are being developed, new alternatives for future research are presenting themselves. Saleh Albeaik et al. (2017) introduce a new indicator for a country's economic productivity, the Economic Complexity Index Plus (ECI+). This provides a nuanced measure of economic complexity by considering exported products alongside their economic value. This revised metric, which they argue offers a more precise prediction of economic growth than the traditional ECI, creates new avenues for future research and improvement.

Overall, my study underscores the potential of public data transparency to influence economic outcomes significantly. While the results are promising and support the hypothesized positive relationship between transparency and economic complexity, they must be interpreted with caution due to the statistical limitations encountered. These findings contribute to a growing body of evidence suggesting that transparent governance can be a key driver of economic sophistication and growth, advocating for policies that enhance transparency in government data practices.

Reflection

Performing this statistical analysis has given me new experience with statistics and programming, allowing me to grow more confident as I aspire to perform more research as an undergraduate in college. Delving into the topics of economic complexity and data transparency has not only broadened my understanding but has also equipped me with a diverse set of skills applicable to both academic and professional settings.

One of the foremost advancements in my research capabilities involved enhancing my skills in analyzing literature. This process was crucial for positioning my research within the existing body of knowledge, identifying gaps, and formulating relevant research questions. These skills ensure that my future projects are grounded in a thorough understanding of the field and are contributing meaningful insights.

Another critical area of development was in sourcing and managing data. With the help of our librarians' compiled data libraries, I conducted extensive searches in various datasets and now have a valuable list of resources that I can use for future research. Additionally, mastering data management techniques in Excel, such as filtering, copying, pasting, and transposing, has prepared me to efficiently prepare datasets for detailed analysis.

Furthermore, utilizing Python for regression analyses and statistical testing significantly enhanced my technical skills. I gained familiarity with Jupyter Notebook, which I found to be a very structured approach to scripting, data analysis, and sharing code. I published my code on Github. These tools have not only improved my ability to conduct research but have also fostered a systematic approach to documenting and sharing my work.

Learning to perform and interpret complex statistical tests, including t-tests and checks for normality and variance, was instrumental in critically assessing the validity of my research

findings. These skills are crucial in understanding the limitations and robustness of statistical inferences, especially under conditions where certain statistical assumptions are not fully met.

The practical skills and theoretical knowledge acquired from this project are instrumental in enhancing my effectiveness and confidence as a researcher. I am better equipped to design studies, analyze complex data sets, and present findings in a coherent and scientifically rigorous manner. These competencies not only bolster my academic career but are also highly transferable to various professional research settings.

Overall, this project has been a great journey that significantly shaped my approach to research. The skills developed are not merely academic but are vital for my future dreams of working in a professional environment that values data analysis and critical thinking. I am eager to apply these new skills to my upcoming projects and continue my development as a knowledgeable and skilled researcher.

References

Albeaik, S., Kaltenberg, M., Alsaleh, M., & Hidalgo, C. A. (2017). Improving the Economic Complexity Index. *arXiv*. https://doi.org/10.48550/arXiv.1707.05826

Center for International Development at Harvard University. (n.d.). The Atlas of Economic Complexity. Retrieved from https://atlas.cid.harvard.edu/rankings

Einav, L. & Levin, J. (2014). Economics in the age of big data. *Science, 346*(6210). https://doi.org/10.1126/science.1243089

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics, 3*(1), 92-113. https://doi.org/10.1038/s42254-020-00275-1

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of Economic Complexity. *Proceedings of the National Academy of Sciences*, *106*(26), 10570–10575. https://doi.org/10.1073/pnas.0900943106

Shambaugh, G.. & Shen, E. (2018). A clear advantage: The benefits of transparency to crisis recovery. *European Journal of Political Economics 55*, 391-416. https://doi.org/10.1016/j.ejpoleco.2018.03.002

World Wide Web Foundation. (n.d.). Open Data Barometer (2nd Edition). Retrieved from https://opendatabarometer.org/2ndEdition/analysis/index.html