

The Impact of Machine Learning Techniques on Instrumental Variable Analysis

David Pruden, Anvith Andhavarapu, Ritvik Vasikarla, Reagan Mathai, Cooper Muery, Mohand Hassan
Texas A&M University: Alam M, Econ 470: Data Science for Economic and Social Issues, Spring 2024
May 4, 2024

Abstract

This paper aims to replicate the findings of Nathan Nunn's study over the long-term effects of Africa's slave trade. By using the historical data provided by Nunn in his replication package as well as economic health indicators from the world bank.

Introduction

- Our work builds on Nunn's paper by using a double lasso regression prior to performing a two-stage OLS regression using instrumental variables to help find which dimensions of the data have the highest impact on the causal relationship between our independent and dependent variables.

Literature Review



Figure: Literature Review

Methodology

- Double lasso performed in order to remove unnecessary variables from the OLS and 2SLS regression. Comparison of causal inference between double lasso regression and traditional regressions. Double lasso regression using several years' GDP per capita. Usage of double lasso to identify whether or not there is a significant difference between double lasso results coefficients and traditional regression coefficients.

Findings

- The double Lasso regression did not significantly impact the coefficient for the instrumental variable. The difference in standard error is not significant between the IV and LASSO regression for data for all years in the dataset. Ln export area consistently had the highest coefficient across the double lasso analysis demonstrating that it was the correct variable to choose. There was no significant difference between the results from the double lasso and from the Instrumental Variable regression.

Second Stage. Dependent variable is log income in 2000, ln y				
ln(exports/area)	-0.208*** (0.053)	-0.201*** (0.047)	-0.286* (0.153)	-0.248*** (0.071)
	[-0.51, -0.14]	[-0.42, -0.13]	[-∞, +∞]	[-0.62, -0.12]
Colonizer fixed effects	No	Yes	Yes	Yes
Geography controls	No	No	Yes	Yes
Restricted sample	No	No	No	Yes
F-stat	15.4	4.32	1.73	2.17
Number of obs.	52	52	52	42

Table: Predicted Export Area and PCGDP 2000 IV

	coef	std err	t	P> t	[0.025	0.975]
const	7.8114	0.187	41.879	0.000	7.437	8.186
predicted_ln_export_area	-0.2079	0.049	-4.285	0.000	-0.305	-0.110

Figure: Replication Scatterplot

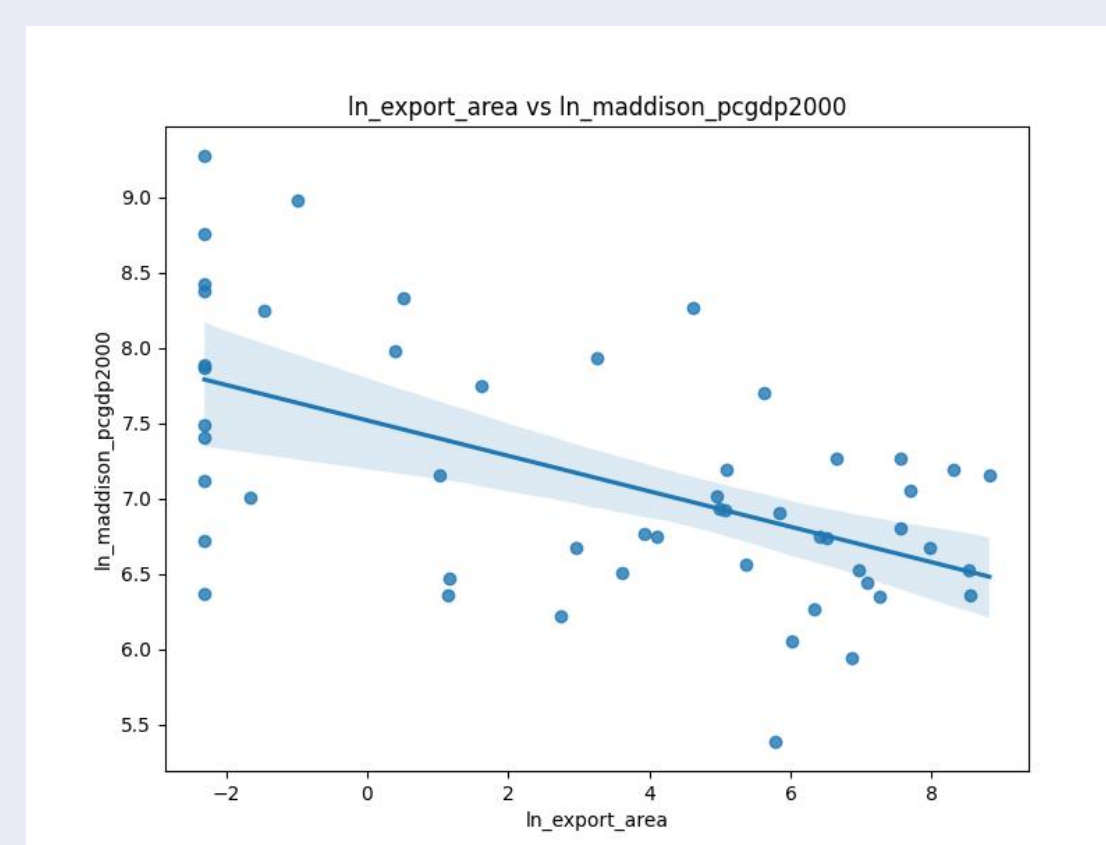


Figure: Nathan Nunn GDP vs Exports Scatter Plot

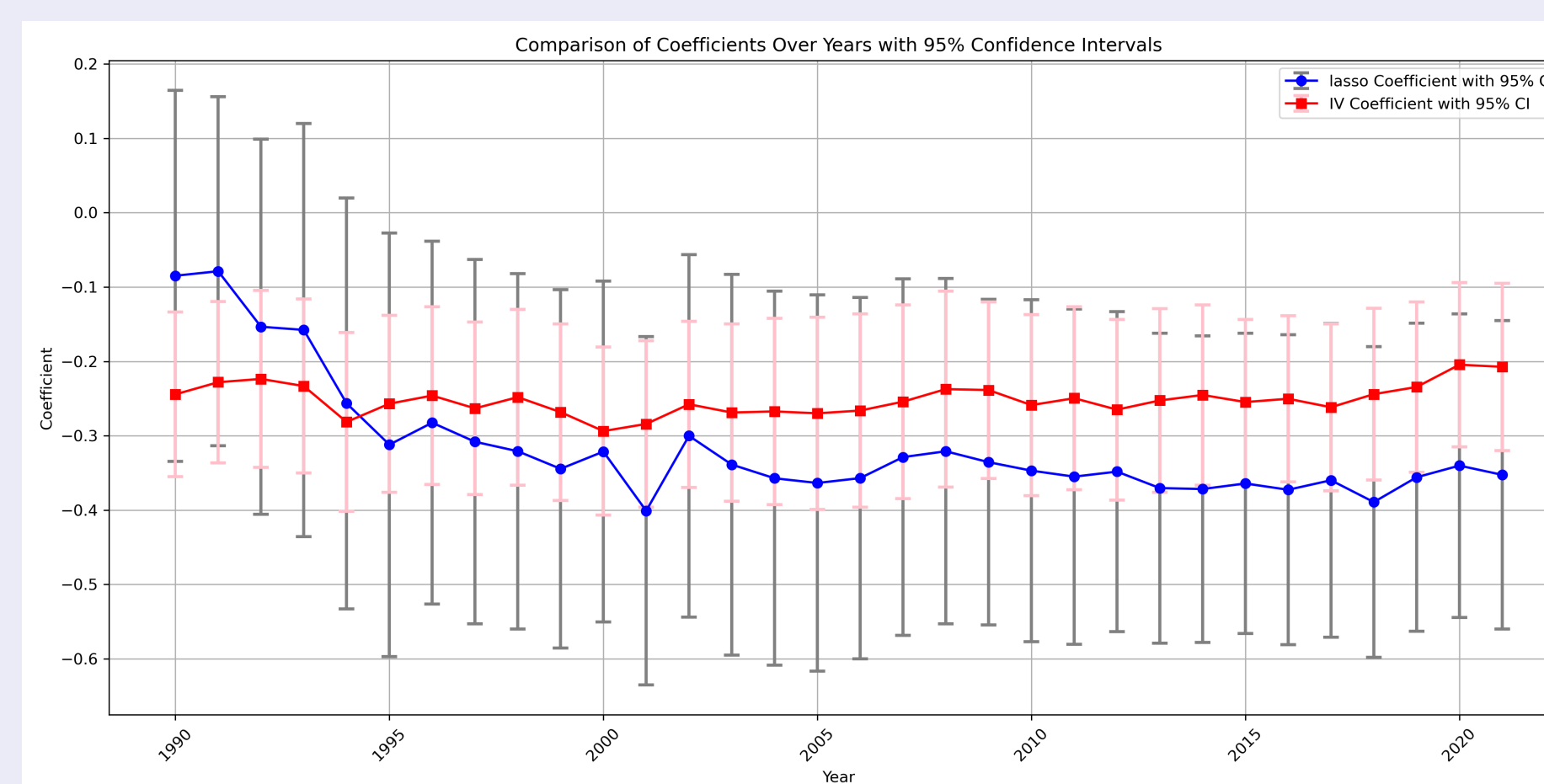
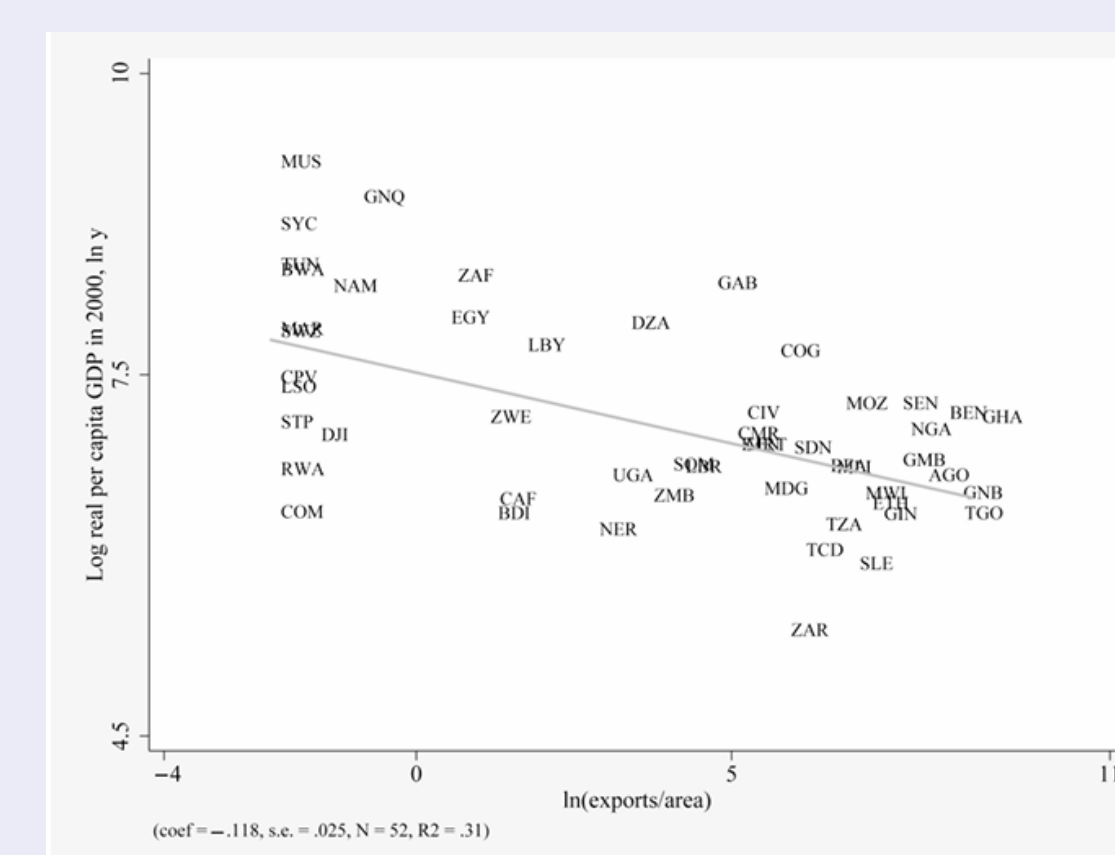


Figure: Time Series Graph

Discussion/Contribution

- Double Lasso method reveals no significant impact, with a slightly larger negative coefficient and increased standard error than the IV method. This suggests that simpler models may underestimate certain economic effects. Further longitudinal research is essential to verify these trends. This research advances knowledge in the field by demonstrating how different econometric models can significantly impact the interpretation of economic variables, using Double Lasso to uncover deeper insights into the economic effects of export areas. Shows the importance of model selection in accurately assessing economic impacts, ensuring that policy recommendations are based on the most comprehensive and reliable analyses.
- Increased model complexity, variable selection process, and high-dimensional nature of the data could be contributing factors to the discrepancy in the confidence intervals.

Conclusions

- Our research demonstrates that the African slave trade has had a long-lasting and detrimental impact on the continent's economic development.
- Areas with higher numbers of slaves exported show significantly lower per capita incomes today, reflecting the trade's destructive effects on institutions, social structures, and human capital.
- These findings underscore the importance of understanding the deep historical roots of Africa's development challenges. Our data is very complex therefore this demonstrates that when using medium-dimensional data there is no significant difference between an instrumental variable analysis and a double-lasso analysis.

References

- Alex Belloni, Victor Chernozhukov and Christian Hansen (2014), "High-dimensional methods and inference on structural and treatment effects," Journal of Economic Perspectives, Spring, 29-50.
- Urmitsky, O., Hansen, C., & Chernozhukov, V. (2019). The Double-Lasso Method for Principled Variable Selection. Retrieved from <https://osf.io/preprints/psyarxiv/2pema/>