rics

A.Kiel

**An extension of debiased machine learning techniques towards traditional econometric approaches**

You have to learn the rules of the game. And then you have to play better than anyone else.
*Albert Einstein*

# Debiased Machine Learning Techniques

## An extension of debiased machine learning techniques towards traditional econometrics

by

A.Kiel

to obtain the Double degree BSc² in Econometrics and Economics
at the Erasmus Universiteit Rotterdam.

| | |
|---|---|
| Student number: | 386879 |
| Project duration: | May 7, 2012 – July 8, 2018 |
| Thesis supervisor: | Dr. A. Naghi, Erasmus Universiteit |

An electronic version of this thesis is available at
`https://github.com/C-o-r/Inference_DML`.

# Preface

We aim on illustrating a modern framework for the application of *nonparametric statistical learning techniques* in the presence of a high-dimensional nuisance parameter set. Traditional machine learning techniques cause distortions on the estimator of interest, for example on a *treatment effect* $\hat{\theta}_0$, targeted by a variety of modern economic research. We thoroughly introduce and examine the proposed techniques on their underlying structure regarding both their underlying model specifications, their aim to overcome the regularization bias and several algorithmic properties for the implemented techniques. For illustration purposes, those branded *"debiased machine learning techniques"* (DML) coefficients estimators are compared to traditional OLS and IV estimation regression results. Thence, we empirically evaluate this nonparametric framework by consulting traditional, social economic research results, which we also explain, originating at Nunn (2011) and Acemoglu (2005).

*A.Kiel*
*Rotterdam, July 1 2018*

# Contents

# 1 Introduction

## 1.1. What are we doing?

Our goal is to illustrate a general ”machine learning” framework on doing inference about a low-dimensional causal parameter. This parameter is framed in this context as a policy or a treatment variable in the presence of a high-dimensional nuisance parameter set, accounting for a traditional machine learning bias. Under this framework, we consult traditional OLS/IV estimations and compare them with the following nonparametric statistical methods, which shall be thoroughly analyzed in the *Methodology* of our work:

- Lasso
- Ridge
- Trees
- Forest
- Boosting
- Deep and Standard Neural Nets
- Aggregations and Cross-Hybrids of these methods.

We analyze the algorithms by assessing their performance and comparing the corresponding ”double biased machine learning” (DML) parameter estimators. We also compare DML to traditional OLS estimation results under the situation of a high amount of covariates.

In particular, for our extended DML framework, in order to construct high quality points that should theoretically concentrate in a $N^{-\frac{1}{2}}$ neighborhood of the true parameter value, we apply Neyman-orthogonal scores and sample splitting (cross-fitting). We construct valid confidence statements of the ”causal” parameters obtained from an IV-estimation with additional control variables. This method is outlined in the underlying research papers by Nunn (2011) and Acemoglu (2005).

Double biased techniques (Chernozhukov et al., 2018) are in this context compared with traditional approaches of OLS and IV-estimations by Nunn (2011) and Acemoglu (2005). We thoroughly discuss the constructed valid confidence statements of the ”causal” parameters obtained from both techniques and derive conclusions.

## 1.2. Why are we doing that?

In modern applications relatively familiar traditional machine learning techniques are not able to capture this high-dimensional nuisance parameter set up ($\eta_0$). These traditional techniques potentially cause a distortion in the treatment effect estimator $\hat{\theta}_0$. On the contrary, the discussed DML methods are presumably able to efficiently handle an enormous amount of covariates, which we are going to test.

## 1.3. Why is it important and how do we contribute to the existing literature?

Valid inferential statements about a low-dimensional parameter in the presence of a complex nuisance parameter set are of high importance. Using DML allows us to relax linear control assumptions for the covariates and replace them by weaker assumptions allowing non-linear relations.

With the empirical validation of these debiased machine learning techniques we contribute to the findings of ”traditional” nonparametric methods such as IV-estimation (Van Der Vaart, 1991). We are aiming for a complementarity of DML with traditional econometric techniques. The results of this paper have different applications based on the underlying research field. This goes beyond the level of purely social economic relevance and can be applied in, for example, observational studies in the presence of a large amount of covariates or controls.

Chernozhukov et al. (2018) sets the groundwork for this paper, by applying in their recent research a generation of nonparametric statistical methods to do estimation over this high-dimensional nuisance parameter set up, targeting $\theta_0$. Their aim to overcome a bias, produced by the off-the-shelf applications of standard machine learning techniques, is hereby their biggest contribution.

Chernozhukov et al. (2018) allows us go beyond prior work done on the field of nuisance parameter inference. They extend a framework, contemporary framed as a ”naive” or prediction-based machine learning approach where ML estimators of $\eta_0$ are naively plugged into the estimating equation of $\theta_0$. Chernozhukov et al. (2018) name their extension to this traditional machine learning scheme ”Debiased Machine Learning” (DML).

Under certain conditions, the ”traditional” methods require the estimators to take values in the Donsker set, a set which complexity is bounded such that the complexity of the functions does not increase with the sample size i.e. it limits the complexity of the function space. This vital assumption rules out most of the high-dimensional techniques. The nonparametric statistical learning methods of Chernozhukov et al. (2018) extends this framework by obtaining estimators in a high-dimensional parameter (set) $\eta_0$, which goes beyond the constraining Donsker set.

Despite the remarkably effectiveness of these methods in predicting, the estimation and inference about ”causal” parameters could be misleading. This regularized estimator has a non-trivial effect on the estimation of $\theta_0$ caused by a regularization bias which originates from keeping the variance in this highly complex setting reasonably small. The naïve estimator fails to be $N^{\frac{1}{2}}$ consistent. A canonical running example in Chernozhukov et al. (2018) is the partially linear model which shall be outlined in the $Methodology$.

The examined *regularization bias* of traditional machine learning techniques which Chernozhukov et al. (2018) aims to overcome via the contribution of Neyman-Orthogonality and sample splitting or ”k-cross validation” serves together with a short algorithm analysis of those implemented techniques as the theoretical framework of our paper.

## 1.4. Why did we choose the two applied papers for illustration?

We aim on comparing traditional estimator findings of $\theta_0$ to the techniques of Chernozhukov et al. (2018). Therefore, we compare coefficient estimators of traditional economic research with our modern DML techniques, aiming to estimate the equivalent underlying parameter of interest. Consequentially, We ”regress” and ”train” at the same time. As our running example, we account hereby for regularization bias in the work of of Nunn (2011) about long-term socialistic effects of slave trade on the current African population and Acemoglu (2005), strengthening the importance of Atlantic trade for the development of western European countries in the early modern world. While these papers add value to a new and growing literature in social economics that seeks to better understand the role that culture, norms, and beliefs play in individual decision making via econometric models, their results serve in our case as a convenient object of investigation. Specifically, due to the prosperous covariate structure of both papers.

## 1.5. What are our main results?

We stress the performance of the hybrid technique, emphasize the relative similarity of the techniques which respect to parameter estimates, however, not with respect to the duration of the implemented algorithm computations. Our results differ from the IV/OLS regression results in a spurious manner, which requires further research.

## 1.6. What is the structure of this paper?

Before comparing DML estimation results with traditional literature results we give a detailed insight into both the models obtained by the traditional econometric research techniques and the DML algorithms' structure. Firstly, the question of interest arises, whether we obtain broadly consistent results, regardless of which DML technique we employ. Based on our results, this question is denied. Our theoretical framework, introducing the (double biased) machine learning techniques to the unfamiliar reader in the *Methodology* is the key pillar of our work. In the Data section we introduce our datasets as well as the consulted 'traditional' models of both underlying research papers. In the *Methodology*, we additionally outline the contribution of Chernozhukov et al. (2018) to traditional machine learning techniques i.e. regularization bias and sample splitting. Finally, we present in the result section the comparison of parameter estimators, and give further comments. We conclude with potential extensions which involved from the limited amount of time for our research.

# 2
# Data

## 2.1. Traditional OLS estimation results

In order to empirically validate the quality of our double biased machine learning algorithms we compare them with standard OLS research results. Therefore, we consult prior research of Acemoglu (2005)[1] and Nunn (2011)[2]. The dataset, STATA replications files and, most importantly, the results which validate the hypotheses of the authors for both papers are available online via the American Economic Association (AEA). We replicate those via STATA. Both papers conduct various regression techniques in order to substantiate their research hypothesis, which can be summarized as follows:

a) Trust levels among the contemporary African population can be traced back to costal slave trade from Africa to the Western Hemisphere. In particular, the environment of ”insecurity caused by the slave trade” generated a culture of greater mistrust to develop which persists up to this day (Nunn, 2011).

b) Atlantic Trade induced institutional changes and economic growth, highly beneficial for Western Europe (Acemoglu, 2005).

The coefficient estimators of both papers serve in this context as the running example with which we aim to compare our DML techniques estimators, ”trained” by the same underlying data. We examine in the consecutive subsections the main results of both papers in order to both understand the underlying structure of the models with respect to the research hypothesis and to give valid comparisons with respect to our obtained DML estimators.

[1]available via https://www.aeaweb.org/articles?id=10.1257/0002828054201305
[2]available via https://www.aeaweb.org/articles?id=10.1257/aer.101.7.3221

### 2.1.1. "The Slave Trade and the Origins of Mistrust in Africa" (Nunn, 2011)

Several strategies in this paper aim to verify a causal correlation the authors uncover. The work mainly complements on Knack (1997) and Guiso (2008), documenting the importance of trust and Europe's colonial influence (Tabellini, 2010). The paper establishes a causal relationship of less trusting individuals nowadays whose ancestors were heavily raided during transatlantic slave trade. Their main results obtained via three different strategies are illustrated in Table **1**, **2**, **3**, **5** and **6**, involving both OLS and IV-estimation in order to additionally control for endogeneity. The results of these tables, we aim to analytically compare via our aforementioned DML techniques. Nunn finds in his research in 2011 that slavery had a "significant negative effect on long-term economic development". Slave trade caused a culture of mistrust to develop within Africa, based on heuristic decision- making strategies (Rindos, 1985) which describe the state of prior slavery as an environment of ubiquitous insecurity which caused individuals to both turn on others, including friends and family members and to kidnap or trick. Convincingly, Nunn (2011) establishes the hypothesis that in this environment, a culture of mistrust may have evolved, which may persist to this day. The applied models in the work of Nunn (2011) contain a large amount of covariates which could lead to an erroneous inference on the "causal" parameter estimator of interest.

Descriptive Statistics

Nunn combines individual-level survey data from seventeen sub-Saharan countries, concentrated in western and southern Africa, with historical data on slave shipments by ethnic group. In particular, the data from the Afrobarometer survey, contains 21,702 respondents who list a valid ethnicity. Moreover, country-level slave exports (1400-1900) from Nunn (2008) are used. For the research, Nunn defines in a first step a set of $baseline\,controls$, a set of individual covariates ($\mathbf{X'}_{i,e,d,c}$) in order to mainly control for the correlation of an individual's income with her trust within the model. This set consists of the following variables:

$age$, $age^2$, , $male\,dummy$, $urban\,dummy$, $education\,(Levels)$, $occupation\,(Levels)$, religion (Levels), $living\,conditions\,(Levels)$, $isocode(Levels)$.

Note hereby that we create level dummies for both isocode, education, occupation, religion and living conditions based on the categorical character of those variables.

Moreover, he defines additional district controls in order to both, again, control for income and to account for a respondent's district environment ($\mathbf{X'}_{d,c}$[3]):

$fractionalization\,of\,an\,ethnic\,per\,district$ , $proportion\,of\,ethnic\,group\,per\,district$.

Furthermore, Nunn categorizes the following local colonial controls ( $\mathbf{X'}_e$), appearing from Table 3 onwards, aiming to control for additional European colonial effects:

$malaria$, $ecology$, $total\,mission\,area$, $explorer\,contact$, $railway\,contact$, $cities$ 1400 $dummy$, $v30\,(Levels)$, $v33$.

Finally, Nunn's research applies 3 more variables for the IV estimations:

$log(\,inital\,population\,density)$, depicted as vector $v$,
$fishing$, $dist\,Saharan\,node$,$dist\,Saharan\,line$, whereby the variables are summarized in a matrix as $\mathbf{X'_{IV}}$.

[3]note the slide abuse of notation since the second variable in this vector varies over both district and ethnicity

Regression Models

We subsume the results of the paper by focusing on Table 1, 2, 3, 5, 6 for now. We ignore for our DML comparison table 9 and table 10 of Nunn's results, depicting the potential causality channels of the present-day mistrust. Nunn (2011) obtains different standard errors based on clustering with ethnicgroups or districts, accounting for potential within-group correlation of the residuals. We ignore in the follow up of the replication the construction of the type specific standard errors, since all three of them are essentially identical. In the general model structure we index for individuals ($i$), ethnic groups ($e$), districts ($d$), countries ($c$) and ergo apply $trust$ as the dependent variable. $\alpha_c$ denotes the country fixed effects, while $trust_{i,e,d,c}$ denotes one out of five measures capturing country-specific factors:

$$trust_{i,e,d,c} = \alpha_c + \beta\, slave\, exports_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \epsilon_{i,e,d,c}. \tag{2.1}$$

In Table **1** the estimator $\beta$ measures the effect of various transformations of the dependent variable slave exports (=exports) onto the neighbor trust of an individual. Aiming to reduce skewness, etc. in the explanatory variables, we apply several different modifications of the explanatory variable slave exports (column 1-6), such as the natural logarithm of one plus slave exports, normalized by land area:

$$neighbortrust_{i,e,d,c} = \alpha_c + \beta\, exports_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.2}$$

$$neighbortrust_{i,e,d,c} = \alpha_c + \beta\, \frac{exports}{area}_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.3}$$

$$neighbortrust_{i,e,d,c} = \alpha_c + \beta\, \frac{exports}{hist\, pop}_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.4}$$

$$neighbortrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + exports)_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.5}$$

$$neighbortrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + \frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.6}$$

$$neighbortrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + \frac{exports}{hist\, pop})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}. \tag{2.7}$$

In Table **2** the estimator $\beta$ measures the effect of $slave\, exports_e$, the slave export of an individuals' ethic group, on the same person's trust, whereby she is acting as a survey respondent who indicates her current level of e.g. neighbor trust:

$$relativestrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + \frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.8}$$

$$neighbortrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + \frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.9}$$

$$counciltrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + \frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.10}$$

$$intratrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + \frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}, \tag{2.11}$$

$$intertrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + \frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \epsilon_{i,e,d,c}. \tag{2.12}$$

A standard variance decomposition which we shall not discuss via the DML framework is carried out in the aftermath of Table **2**. This modern approach essentially determines the ratio of observable to unobservable factors in order to completely explain away the negatively established relation Nunn (2011) aims to verify in his hypotheses. Nunn concludes hereby that the estimator findings are unlikely caused by unobserved heterogeneity. Moreover, a number of robustness and sensitivity checks are carried out. Thereafter, OLS results are compared to a logit model. Nunn keeps the data, in accordance with our DML analysis, on individual level to account in the later process better for causal mechanisms. We skip these, in our opinion, less important steps for the DML analysis.

In Table **3** Nunn (2011) controls for an extensive set of additional covariates ($\mathbf{X'}_e\mathbf{\Phi}$) which capture the potential effects of non-slave trade European influence on long term-trust:

$$relativestrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \epsilon_{i,e,d,c}, \tag{2.13}$$

$$neighbortrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \epsilon_{i,e,d,c}, \tag{2.14}$$

$$counciltrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \epsilon_{i,e,d,c}, \tag{2.15}$$

$$intratrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \epsilon_{i,e,d,c}, \tag{2.16}$$

$$intertrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{exports}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \epsilon_{i,e,d,c}. \tag{2.17}$$

In Table **5** and **6** Nunn makes use of an instrumental variable. We use the distance of an individual's ethnic group to the coast as an instrument for the number of slaves taken, due to the unique history of the regions of interest (Table **5**). We use a two-stage lest square estimation (2SLS). In stage 1 we regress the endogenous variables on the exogenous variable, the instrument. In stage 2, we regress the resulting $ln(1+\widehat{\frac{exports}{area}})_e$ on our respective trust variable:

$$Stage1 : ln(1+\frac{exports}{area})_e = \alpha_c + \beta\, distance\, coasts_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + v\delta + \epsilon_{i,e,d,c}, \tag{2.18}$$

$$Stage2.1 : relativestrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + v\delta + \epsilon_{i,e,d,c}, \tag{2.19}$$

$$Stage2.2 : neighbortrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + v\delta + \epsilon_{i,e,d,c}, \tag{2.20}$$

$$Stage2.3 : counciltrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + v\delta + \epsilon_{i,e,d,c}, \tag{2.21}$$

$$Stage2.4 : intratrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + v\delta + \epsilon_{i,e,d,c}, \tag{2.22}$$

$$Stage2.5 : intertrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + v\delta + \epsilon_{i,e,d,c}. \tag{2.23}$$

However, distance might be correlated to early trade in Sahara desert. Therefore, Nunn (2011) additionally adds the aforementioned control variables distance to Sharan routes, fishing and other European instruments to the model depicted in Table **6**:

$$Stage1 : ln(1+\frac{exports}{area})_e = \alpha_c + \beta\, distance\, coasts_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \mathbf{X'_{IV}}\mathbf{\Delta} + \epsilon_{i,e,d,c}, \tag{2.24}$$

$$Stage2.1 : relativestrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \mathbf{X'_{IV}}\mathbf{\Delta} + \epsilon_{i,e,d,c}, \tag{2.25}$$

$$Stage2.2 : neighbortrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi}\mathbf{X'_{IV}}\mathbf{\Delta} + \epsilon_{i,e,d,c}, \tag{2.26}$$

$$Stage2.3 : counciltrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \mathbf{X'_{IV}}\mathbf{\Delta} + \epsilon_{i,e,d,c}, \tag{2.27}$$

$$Stage2.4 : intratrust_{i,e,d,c} = \alpha_c + \beta\, ln(1+\frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \mathbf{X'_{IV}}\mathbf{\Delta} + \epsilon_{i,e,d,c}, \tag{2.28}$$

$$Stage2.5: intertrust_{i,e,d,c} = \alpha_c + \beta\, ln(1 + \frac{\widehat{exports}}{area})_e + \mathbf{X'}_{i,e,d,c}\mathbf{\Gamma} + \mathbf{X'}_{d,c}\mathbf{\Omega} + \mathbf{X'}_e\mathbf{\Phi} + \mathbf{X'_{IV}}\mathbf{\Delta} + \epsilon_{i,e,d,c}. \quad (2.29)$$

Table 7 and 8 involve a second falsification test outside Europe in order to validate the power of this applied instrumental variable. The assumption of Nunn (2011) for this IV model is that distance had an impact on trust only though slave trade. We omit those falsification tests for now. Table 9 and 10 seek to extend the hypotheses of less trusting individuals today by trying to understand the importance of diverse channels of causality. Therefore, Nunn (2011) differentiates between an external and internal environment which influences the contemporary individual. Relocations plays hereby an essential role. The tests are stressing the internal norms of the individual, however, we omit these models and result tables with respect to our overall DML results testing framework, focusing specifically on the 5 aforementioned model results.

### 2.1.2. "The rise of Europe: Atlantic trade, institutional change, and economic growth" (Acemoglu, 2005)

The second paper presents itself as a "marriage between the Marxist thesis linking the rise of the bourgeoisie and development of the world economy" by emphasizing the beneficial influences of Atlantic trade on Western Europe, creating a higher economic prosperity compared to other regions, such as Asia or Eastern Europe. The neoclassical motion of developing political institutions and securing property rights is connected to the Atlantic sea trade between the 16ht and 19th century. The paper contributes the rise of Western Europe largely to the rise of the "Atlantic Europe" i.e. rise of Atlantic ports. Hereby, the author tries to weigh against a theory of purely internal dynamics causing the rise of European growth such as catholic/protestant religion or a roman heritage. Atlantic trade caused large profits, beneficial for property rights and initial non-absolutist institutions. In general, access to the Atlantic ocean seems to be the most important point in accounting for the growth in Western Europe during the 16th to 19th century, compared to other regions on the globe. The author emphasizes in his research hypothesis the importance of profits made in Atlantic trade, colonialism and slavery.

#### Descriptive Statistics

Acemoglu (2005) uses urban population numbers from Bairoch et al. (1988), which is a dataset about 2,200 European cities with more than 5000 inhabitants. Estimates of McEvendy to calculate urbanization. Barioch and Vries provide additional estimates for Asia. Urbanization serves in this instance also as a proxy for the GDP. GDP measures per capita are obtained from Maddison (2001). European city-level data is retrieved from Bairoch et al. (1988). He aims to control for endogenity via controls. Acemoglu's research involves variables such as:

$sjurbanization$, $westerneurope1600$, $westerneurope1700$, $westerneurope1750$, $westerneurope1800$, $westerneurope1850$,
$atlantictrader1500$, $atlantictrader1600$, $atlantictrader1700$, $atlantictrader1750$, $atlantictrader1800$, $atlantictrader1850$,
$country1, ..., country29$,
$year1000$, ..., $year1850$.

#### Regression Models

We emphasize for our DML analysis the results of Table **2**. In the general model structure of this paper, $u_{j,t}$ represents the amount of urbanization in country j a time t, $WE_j$ is a dummy for a Western Europe location, $d_t$ denotes year effects, $\delta_j$ denotes country effects, $X'_{jt}$ a vector of other covariates and $\epsilon_{jt}$ is the disturbance term. Moreover, we apply either $PAT_j$ which is a dummy for an Atlantic trader or the in another model, the Atlantic coastline-to-area ratio (time invariant of the country). We are mainly interested in the $\beta_t$ coefficients which we are also estimate via DML. The general model can be framed as:

$$u_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \sum_{t \geq 1600} \beta_t \, PAT_j \, d_t + X'_{jt} \gamma + \epsilon_{jt}. \tag{2.30}$$

In order to optimally replicate the regression results we aim for a replication of column 5 and 10 Table **2**, which represent unweighted regressions. Moreover, in contrast to Acemoglu (2005) we do not test for the joint significance of the interaction terms $\sum_{t \geq 1600} \alpha_t \, WE_j \, d_t$, but instead focus on the single OLS parameter estimators of column 5 and column 10 in Table **2**, applying a both structured and unstructured unweighted specification. The models of interest for Table **2** can be depicted as follows:

unstructured (col. 5)

$$u_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \sum_{t \geq 1600} \beta_t \, PAT_j \, d_t + \epsilon_{jt}, \tag{2.31}$$

structured (col. 5)

$$u_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \beta \, (PAT_j * log(AT_t)) + \epsilon_{jt}, \tag{2.32}$$

unstructured (col. 10)

$$u_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \sum_{t \geq 1600} \beta_t \, CTA_j \, d_t + \epsilon_{jt}, \tag{2.33}$$

structured (col. 10)

$$u_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \beta \, (CTA_j * log(AT_t)) + \epsilon_{jt}. \tag{2.34}$$

In Table **3** we replace the dependent variable urbanization with $log(GDP \, per \, capita)$.
unstructured (col. 5)

$$log(GDP \, per \, capita)_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \sum_{t \geq 1600} \beta_t \, PAT_j \, d_t + \epsilon_{jt}, \tag{2.35}$$

structured (col. 5)

$$log(GDP \, per \, capita)_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \beta \, (PAT_j * log(AT_t)) + \epsilon_{jt}, \tag{2.36}$$

unstructured (col. 10)

$$log(GDP \, per \, capita)_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \sum_{t \geq 1600} \beta_t \, CTA_j \, d_t + \epsilon_{jt}, \tag{2.37}$$

structured (col. 10)

$$log(GDP \, per \, capita)_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \beta \, (CTA_j * log(AT_t)) + \epsilon_{jt}. \tag{2.38}$$

In Table **4** we check the robustness of our results by adding a number of covariates. We aim to verify whether internal European dynamics effect our results. Hereby, religion is added as an important determinant of economic and social development. Moreover, we add a variable which indicates the average number of years at war, as well as a dummy indicating whether the country was part of the roman empire. Finally, we add more explanatory variables (distance from equator + dates from 1600):

col. 1.1: (via Atlantic trader dummy measure of potential for Atlantic trade)

$$u_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \beta \; (PAT_j * log(AT_t)) + X'_{jt}\gamma + \epsilon_{jt}, \tag{2.39}$$

col. 1.2: (via Atlantic coastline-to-area-measure of potential for Atlantic trade)

$$u_{jt} = d_t + \delta_j + \sum_{t \geq 1600} \alpha_t \, WE_j \, d_t + \beta \; (CTA_j * log(AT_t)) + X'_{jt}\gamma + \epsilon_{jt}. \tag{2.40}$$

For Table 4 we have in total eight different regressions for the analysis, where $X'_{jt}$ for the additional covariate term in the regression ($X'_{jt}\gamma$) depicts either:

- religion: $protestant1600$, ..., $protestant1850$
- war: $warsperyear$
- roman heritage : $romanempire1600$, ...,$romanempire1850$
- latitude: $latitude1600$, ...,$latitude1850$

We omit the rest of the results in Acemoglu's paper, deriving mainly positive effects of trade on institutions, property rights, etc., focusing solely on the first Table (table **2**) for the DML replication of this paper for now.

## 2.2. Further Comments

The finer transmission mechanisms of both papers causing the explored correlations/causations, potentially, arise from a high dimensional nuisance parameter set which cannot be sufficiently captured by the afore-described models. The paper of Nunn (20111) is based on the vital assumption of a dominant presumption that cultural change occurs slowly (Bisin, 2000). The author identifies a negative causal relationship between slave trade and income today. The author stresses in the research their main finding, mainly that the underlying heuristic decision- making strategy theory is confirmed. This reasoning seems fairly one-sided. Under this context, the paper does not discuss potentially other emotional feelings attached with the state of slavery such as potential protectiveness and care-taking among the slaves. Also, modern day trust can be indeed fairly difficult to measure. A survey might not be sufficient. Based and the results and conclusion, the question of concern arises whether slaves really become mistrusting solely due to the fact that they were betrayed and whether they pass this attribute among generations. The control variables have a vial function in this context which we aim to better account for via our new modern methods. In general, it seems rather doubtful that the emotional state of the survey correspondent is largely based on the ancestors betrayal, therefore DML driven estimator coefficients could be potentially much "smaller".
On the other hand, the results of Acemoglu (2005), intuitively, seem to be in inclination with a potential high dimensional nuisance parameter set up. Nonetheless, we expect the affect of potential inter dynamic European control variables for Table 4) i.e. religion, war, etc., to have a greater effect on the dependent variable in our high dimensional nuisance parameter set up via DML. Our first hypothesis shall be verified in the results section of this paper, along with various other findings. DML could be a better tool than relatively "simpler" linear regression. It accounts for nonlinearties and interactions in a much more flexible way which shall be outlined the the following section.

# 3

# Methodology

In this section we depict several regularization methods. DML can be applied to any of those. We frame these methods as " traditional machine learning techniques" and introduce them in the next subsection shortly. For a thorough understanding additional material from the R-libraries, etc., should be consulted. Afterwards, we explain the contribution of Chernozhukov et al. (2018): Neyman-Orthogonality and Sample Splitting. A short algorithm analysis, involving both computation times and a graphical depiction of the parameter estimator deviations with respect to changing the amount of covariates, based on the relatively small AJR dataset, gives the reader a first illustration of these regularization methods.

## 3.1. Machine learning techniques

This explanation builds upon the R-code provided by Chernozhukov et al. (2018). The following packages are of major interest: 'MASS', 'randomForest', 'neuralnet', 'gbm', 'sandwich', 'hdm', 'nnet', 'rpart', 'glmnet'.

### 3.1.1. Trees-library(rpart)

For out Trees technique we specify as an input parameter a regression method e.g. anova. Prediction or regression trees (in case of a continuous response variable) are used to predict a class $Y$ from input $X_1$,$X_2$,...,$X_n$. We subbranch by checking iteratively regressors at each node, up to a leave which results in the "optimal" prediction. Contrary to ordinary regression models where the OLS estimate formula holds for example for the entire data space, trees try to partition the data space into small enough parts, whereby a simple different model can be applied in each step towards an "ideal" leave node. Prediction trees are adaptive nearest-neighbor methods e.g. k-nearest-neighbors.

The main idea of this technique is to partition the covariance space and to do separate things. We use regression trees in order to estimate a regression function non-parametrically in a flexible way. In regression trees we partition covariate space constantly, where the regression function is estimated as the average outcome for units with covariate values in that subspace. We partition sequentially, one covariate at a time, in order to reduce the sum of squared deviation from the regression estimated as much as possible. Hereby, we estimate the regression function in each point (node) of the subspaces.

The question of interest arises on how to partition this covariance space, since we cannot analyze all subsets due to a combinatorial explosion. As a starting point we take the whole covariance space and estimate $g(x) = \hat{Y}$, whereby the sum of squared deviations is

$$Q(g) = \sum_{i=1}^{N}(Y_i - h(X_i))^2 = \sum_{i=1}^{N}(Y_i - \hat{Y})^2. \tag{3.1}$$

With respect to the covariate k and the treshold t, we split the data depending on whether $X_{i,k} \leq t$ versus $X_{i,k} \geq t$

We let the two averages be

$$\hat{Y}_{left} = \frac{\sum_{i:X_{i,k}\leq t} Y_i}{\sum_{i:X_{i,k}\leq t} 1}, \hat{Y}_{right} = \frac{\sum_{i:X_{i,k}\geq t} Y_i}{\sum_{i:X_{i,k}\geq t} 1}. \tag{3.2}$$

We look for the covariate or threshold that minimizes the sum of squares. Under that context we define the estimator for covariate k and threshold t as

$$g_{k,t}(x) = \begin{cases} Y_{l\hat{e}ft}\; if x_k \leq t \\ Y_{ri\hat{g}ht}\; if x_k \leq t. \end{cases} \tag{3.3}$$

We find the covariate k* and the threshold t* that solve

$$(k*, t*) = argmin_{k,t} Q(g_{k,t}(.)). \tag{3.4}$$

We partition the covariate space into two subspaces by deciding whether $X_{i,k*}$ is bigger than t* or not, choosing simultaneously between covariances, thresholds and subspaces.

We additionally penalize the amount of splits via $(Q(g) + \lambda * \#(leaves))$. For this flexible step function the establishment of formal asymptotic properties seems again difficult as there are, for example, no confidence intervals available.

In order to select an appropriate penalty term for this method, we have to account for the number of leaves. For each $\lambda$ value, we grow the tree by excluding our b-th cross validation $(g(b,\lambda))$. We sum up for each $\lambda$ the squared errors over the cross validation sample in order to get:

$Q(\lambda) = \sum_{b=1}^{B} \sum_{I_i=b} (Y_i - g(b,\lambda))^2.$

We choose the $\lambda$ that minimizes this criterion and estimate the tree, given this particular chosen penalty parameter. Note that we only look at a discrete set of lambdas. Prunning the tree, similar to other supervised learning techniques should lead to an improvement of the underlying problem. We can ”grow” much bigger trees with the hope that later splitting, which does not improve the subjective function initially, improves things in later stages i.e. go back and prune branches or leaves that do not collectively improve the objective function sufficiently. In this way we are able to find possible interactions among the covariates.

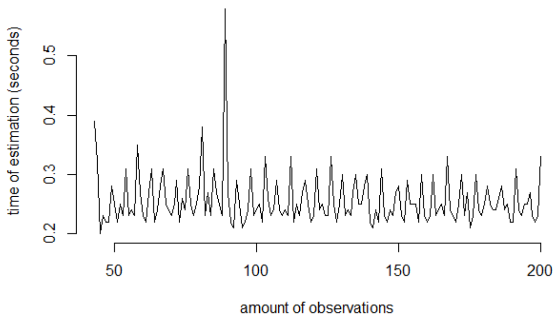


Figure 3.1: Duration of the DML Tree implementation.

In Figure 3.1 we analyze this algorithm on its properties with respect to handling a dataset. We use the AJR dataset, artificially increase its size from 43 to 200 observations and measure the amount of time the Tree DML algorithm takes to compute for every of the 167 subproblems its ”ideal” parameters. In Figure 3.2 we depict the changing parameter estimators, once we add covariates to the estimation procure. The idea of of Figure 3.1 and 3.2 shall be applied to the upcoming DML techniques. Finally, we compare those findings.

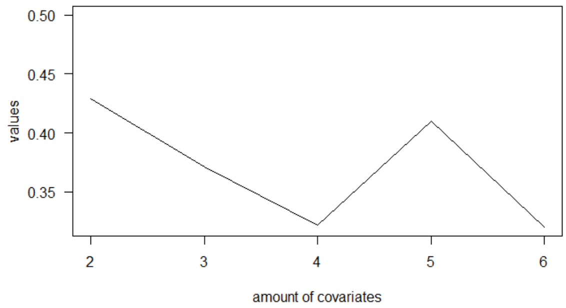


Figure 3.2: Parameter estimations of the DML Trees method by iteratively increasing the amount of covariates.

### 3.1.2. Random Forest-library(random Forest)

The Forest implementation has various inputs such as nodesize (class), an amount of trees, etc. First proposed by Tin Kam Ho, Random Forests came into the spotlight in 2001 after their description by Breiman and serve as an extension to the original decision trees. These supervised ensemble-learning begging models, based on low-bias decision trees, effectively work via bootstrap samples. By finding an optimal threshold and covariance for the $d$ arbitrarily selected regressors in each step, we construct trees or models that are not correlated with each other (bootstrapping). The node splitting criteria is again based on the a search for the best feature among a random subset of features. Majority voting beneficially adds to the idea of overfitting and reduces the effect of outliers compared to the trees. Averaging our results over decor related trees improves the predictive accuracy. Also used in the context of classification, we apply this procedure, instead, for our specific regression context. Pre-pruning options should be specified e.g. the depth (vital for the complexity) or the smallest subset that can be split. Recent research emphasizes the use of Random Forest and even suggests that asymptotic normality may hold Potential advantages of this (D)ML technique are beneficial effects of majority voting, however, potential disadvantages are clearly the not easily interpretable output of this technique.

**Algorithm 1** Random Forrest Pseudocode

```
procedure RANDOM FORREST(t, n...)                              ▷ Input variables
    i ← 0                                                      ▷ # trees
    j ← 0                                                      ▷ # nodes
    while i ≠ t do
        while j ≠ n do
            Randomly select k regressors from total m
            Among all k, find node d with best cov + treshold
            Split node using the best split
        end while
    end while
    Assess results by averaging trees over bootstrap samples
    return estimator
end procedure
```

Figure 3.3: Pseudo code for a Random Forest implementation.

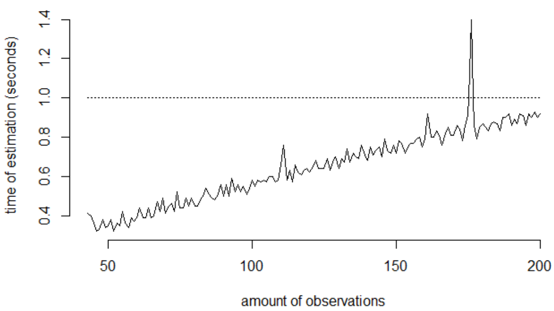


Figure 3.4: Duration of the DML Forest implementation.

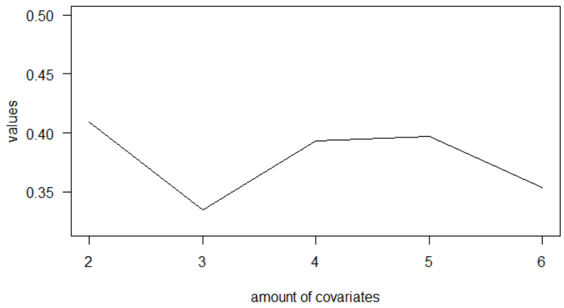


Figure 3.5: Parameter estimations of the DML Forest method by iteratively increasing the amount of covariates.

### 3.1.3. Neural Network-library(nnet)

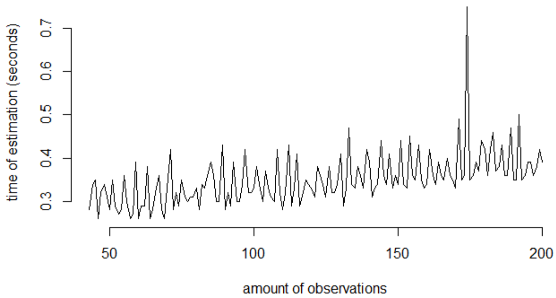


Figure 3.6: Duration of the DML Nnet implementation

For the provided Neural Net DML implementation, we specify of the Neural Net size (e.g. 8), a maximimum amount of iterations (e.g. 1000), a decay rate (e.g. 0.01) and the axNWts (e.g. 10000). In its basics, an elastic net, for example, combines ridge and lasso techniques, establishing a penalty term which is a combination of an $L_2$ and $L_1$ shrinkage norm (explained in 3.1.4), leading to the general formula for a simple elastic net:

$$min \sum_{i=1}^{N} (Y_i - (X_i\beta))^2 + \lambda(\alpha * ||\beta_1|| + (1-\alpha)||\beta_2||). \tag{3.5}$$

By finding $\lambda$ and $\alpha$ we find the required tuning parameters. Note that if $\alpha$ is 0 we obtain a Lasso model, if $\alpha$ is 1, we obtain a Ridge model. In this setting we combine positive properties of both models. More complicated regularization methods lead to more complex implementations. We choose various predetermined values for $\alpha$ and optimize over those and $\lambda$, in order to avoid a certain instability of the problem. The strength of this DML method is that we aim to find hidden layers within the explanatory variable $X_i$ via $Z_i$. Hereby, we aim on better capturing the relations between, for example, the underlying covariate structure:

$$Z_{i,m} = \sigma(\alpha_{0m} + \alpha'_{1m}X_i), \quad for\ m = 1, ..., M, \tag{3.6}$$

$$Y_i = \beta_0 + \beta'_1 Z_i + \varepsilon_i, \tag{3.7}$$

$$\sum_{i=1}^{N} (Y_i - g(X_i, \alpha, \beta))^2, \tag{3.8}$$

$$\lambda(\sum_{k,m} \alpha_{k,m}^2 + \sum_k \beta_k^2). \tag{3.9}$$

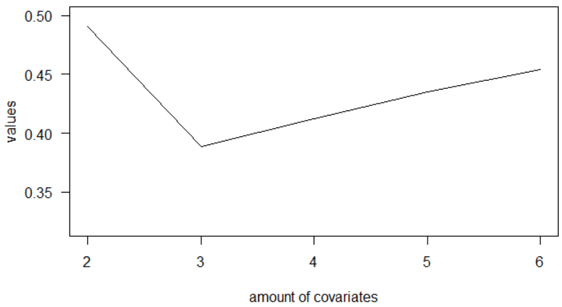


Figure 3.7: Parameter estimations of the DML Neural Network by iteratively increasing the amount of covariates.

### 3.1.4. Lasso, Ridge-library(glmnet,rlasso)

Lasso's (Leasts Absolute Selection and Shrinkage operator) approach to look at the estimates and to do regularization works incredibly well. It is the most popular traditional machine learning technique in econometrics. We define the aforementioned $L_p$ norm as:

$$||x||_p = (\sum_{k=1}^{K} |x_k|^p)^{\frac{1}{p}}. \tag{3.10}$$

In the Ridge framework, we generally aim to minimize the sum of squared deviation by using a penalty term. Preferably, we normalize the covariates. We make use of the $L_2$ norm in this case. We shrink estimates continuously towards a common value, typically 0, in order to achieve a sparse collection, choosing only particular covariates (sparsity principle). We select submodels to select even smaller models. For this subset selection we use the $L_0$ norm, limiting non zero coefficients by a penalty term. Computationally, this procedure can be again extremely hard, due to the fact that all combinations have to be verified which is not scalable). Overall, we differ from the standard OLS regression model by applying:

$$Y_i = \sum_{k=1}^{K} X_{ik} + \beta_k \epsilon_i = X_i + \epsilon_i. \tag{3.11}$$

We estimate $\beta$ as

$$\beta_{ridge} = (X'X + \lambda * I_K)^{-1}(X'Y), \tag{3.12}$$

or

$$\beta_{ridge} = \sum_{i=1}^{N} (Y_i - X_i\beta)^2 + \lambda||\beta_k||_2^2 X_i. \tag{3.13}$$

whereby we inflate $X'X$ by $\lambda * I_k$, so it becomes positive definite and independent of K, in cases where K > N. From a Bayesian perspective the $\beta_{ridge}$ estimator seems appealing. Overall, we shrink OLS coefficients to 0. If maximally correlated, we shrink them by $1/(1+\lambda)$. A main disadvantage of this method is that a general matrix inversion will get hard at some point. In case of a modest size effect Ridge seems to perform better than the Lasso (Tibshirani, 1996). Due to the $L_2$ norm, Ridge shrinks more than Lasso. However, shrinking the estimate a lot might not be effective in some set-ups. For example a thick-tailed distribution might be better handled by Lasso ($L_1$ norm) than Ridge.

For the DML RLasso technique we prespecify values for $\lambda$, an intercept etc. In this framework we make use, instead, of the vital $L_1$ norm. We sum in this context the absolute values of betas. We minimize the sum of squared but in contrast to a raw subset selection add we add, similar to Ridge, a penalty term.

For very large values of coefficients, we could obtain a difference in estimators between Ridge and Lasso. Ridge will shrink the large coefficients much more than lasso due to the different norm. Hence, Lasso might be doing, again, better in certain instances:

$$min(\sum_{i=1}^{N}(Y_i - X_i\beta)^2 + \lambda||\beta_k||. \tag{3.14}$$

Lasso uses hereby the "$L_1$" norm. In case we obtain big difference between Lasso and Ridge model and literature recommends that Lasso is likely to give better results. The Lasso technique has become extremely popular. For further insights we refer to Belloni et al. (2014).

We finally conclude for the comparison between Lasso and Ridge that similar sized coefficient are presumably better treated with Ridge while in a set up with a differing importance in covariates, Lasso seems to be preferred. Moreover, if the coefficients are close to null, Lasso does worse. Highly correlated regressors can lead to unstableness for the Lasso technique, whereby Ridge can handle those situation relatively better.

We choose a Lasso penalty parameter, by standardizing $X_i$ and $Y_i$ where no intercept is needed.

Hence we formulate,

$$min(\sum_{i=1}^{N}(Y_i - X_i\beta)^2 + \lambda||\beta_1||, \tag{3.15}$$

as

$$min(\sum_{i=1}^{N}(Y_i - X_i\beta)^2 s.t. \sum_{k=1}^{K}|\beta_k| \leq t\sum_{k=1}^{K}|\beta_k^{OLS}|, \tag{3.16}$$

where t is a scalar between 0 and 1, in particular 0 if we shrink all estimates and if 1 we apply no shrinking. We choose this penalty term via cross validation. Let $I_i \in 1, ...B$ indicate the $b$-th crossvalidation sample via an integer. If we select for example for $B = 10$ we obtain for this sample for b = 1, ...,B estimate $\beta_b(\lambda)$:

$$\beta_b(\lambda) = argmin(\sum_{i;I_i \neq b}(Y_i - X_i\beta)^2 + \lambda\sum_{k=1}^{K}|\beta_k|, \tag{3.17}$$

on all data with $B_i \neq b$. Finally we calculate the sum of squared errors via

$$Q(b,\lambda) = argmin(\sum_{i;I_i \neq b}((Y_i - X_i'\beta_b(\lambda))^2, \tag{3.18}$$

by taking the average of these cross validation samples and its standard error

$$\hat{Q}(\lambda) = \frac{1}{B}\sum_{b=1}^{B}Q(b,\lambda), se(\lambda) = (\frac{1}{B^2}\sum_{b=1}^{B}(Q(b,\lambda) - \hat{Q}(\lambda)^2))^{\frac{1}{2}}, \tag{3.19}$$

and we choose

$$\lambda_{min} = argmin\, Q(\lambda), \tag{3.20}$$

which tends to overfit a bit, hence we use the largest $\lambda$ such that

$$Q(\lambda) \leq Q(\lambda_{min}) + se(\lambda_{min}). \tag{3.21}$$

Hence, we choose the largest $\lambda$ that is reasonably close in minimizing the cross validation standard error. Various ad-hoc decisions are involved in this process.

Overall Lasso and Ridge scale up very well for bigger data and generally have high interpretive power e.g. about the amount of null coefficients, compared to other (D)ML techniques.

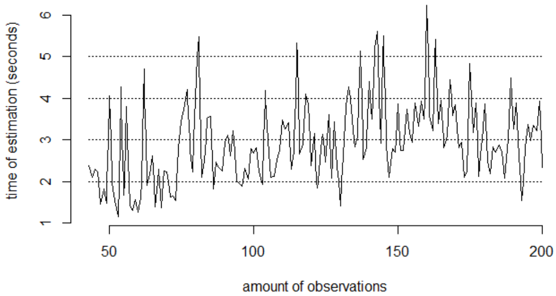


Figure 3.8: Duration of the DML (R)Lasso implementation

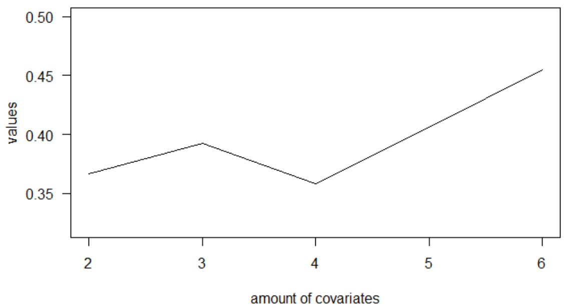


Figure 3.9: Parameter estimations of the DML Lasso method by iteratively increasing the amount of covariates

### 3.1.5. Boosting-library(gbm)

In our algorithm analysis Figure 3.9 graphically illustrates that this method is by far the slowest DML technique. For this technique we prespecify a bagging and training fraction, an amount of trees and an underlying distribution (e.g. Gaussian" (squared error)). The applied, underlying AdaBoost algorithm has a particular loss function and optimization algorithm associated by using Friedman's Gradient Descent (Boost) algorithm.

We refrain from a deeper insight and state at this point that in any function estimation we wish to to find a regression function $f(\hat{x})$, that minimizes the expectation of some loss function $\phi(y, f)$. For a thorough discussion we refer both to the respective *gbm* library and to Friedman (2001):

$$f(\hat{x}) = arg\,min\, E_{y,x}\phi(y, f(x)). \tag{3.22}$$

We iteratively compute the negative gradient , fit a regression model, predict and choose a gradient step size in order to update the estimates of $f(\hat{x})$.

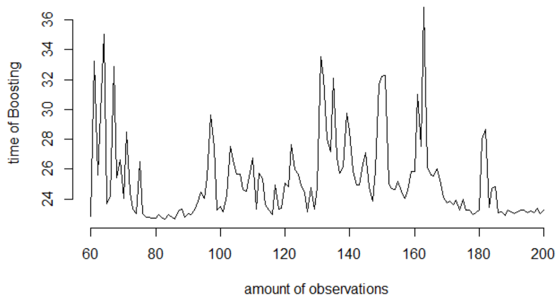


Figure 3.10: Duration of the DML Boosting implementation

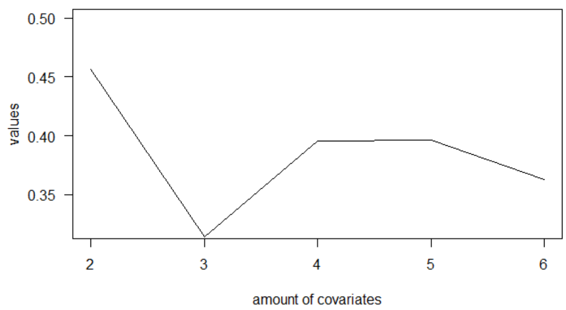


Figure 3.11: Parameter estimations of the DML Boosting method by iteratively increasing the amount of covariates

### 3.1.6. Hybrid/Ensemble techniques

A good Ensemble or Hybrid consists of strong independent predictors and an appropriate combiner. Our Ensemble constitutes of the DML methods ”Lasso”, ”Boosting”, ”Forest”, ”Nnet”. From an intuitive information theory standpoint the use of combinations of predictors in order to achieve better performances seems attractive. In 1785 Marquis de Condorcet established his famous jury theorem:

Votes:
$P_i \in [-1, +1]$,

Majority:
$S_n = \sum_{i=1}^{n} P_i$,

if each voter has a probability larger than $\frac{1}{2}$, being correct

$$Pr[P_i = correct] > 1/2 \tag{3.23}$$

then the probability of the majority vote being correct will go to 1 as the size of the jury increases:

$$Pr[S_n = correct] -> 1,\ as\ n -> \infty. \tag{3.24}$$

Favorable for this method is the fact that we can parallelize, hence, run each single DML predictor technique on a separate computing server, which potentially reduces runtime and memory storage. Similar to our jury theorem the Ensemble's methods performance increases if the underlying DML techniques remain independent. A disadvantage of that method is that we have to wait for all predictors to finish training, hence we are only as fast as the slowest single predictor method, in our case Boosting.

In general, various sorts of combiners in these Ensemble techniques are applied such as fixed combiners which use a majority/ average voting rule or a confidence weighting. Besides, trained combiners are used to level according to a linear weighting. For the specification of the Ensemble method in the code of (Chernozhukov et al., 2018), we refer to the Moment_Functions.R file in our online repository.

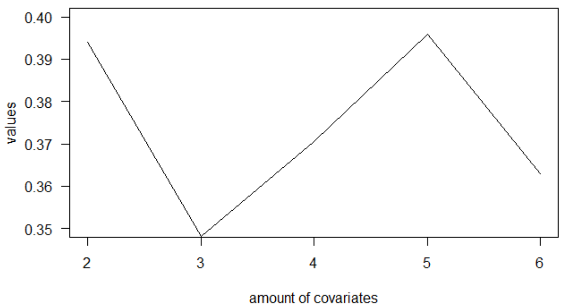


Figure 3.12: Parameter estimations of the DML Hybrid/ Ensemble method by iteratively increasing the amount of covariates. Notice the much smaller y-scale. The results of the Ensemble method seem to be the most stable ones.

### 3.1.7. Overall comparison

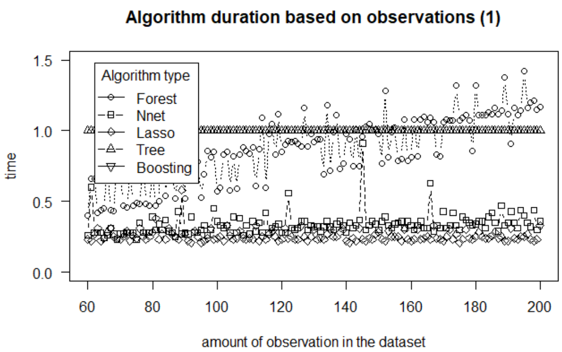


Figure 3.13: Duration of all DML implementations (Boosting not visible)

Overall, we illustrate that the (D)ML Boosting method is much slower than any of the other DML methods (not visible in Figure 3.12 but 3.13). While the Forest clearly unveils a positive trend in the computation time with respect to an increasing amount of observations, the computation for the other techniques seems to be only slightly effected by increasing the amount of observation from 43 to 200 in the dataset. Moreover, with respect the covariates, the estimators seems to equally differ in size for all DML methods, except for the Ensemble method for which the estimators seem to involve relatively stable (see —).

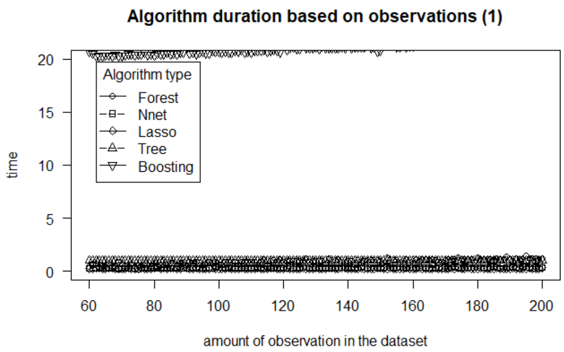


Figure 3.14: Duration of all DML implementations (Boosting visible)

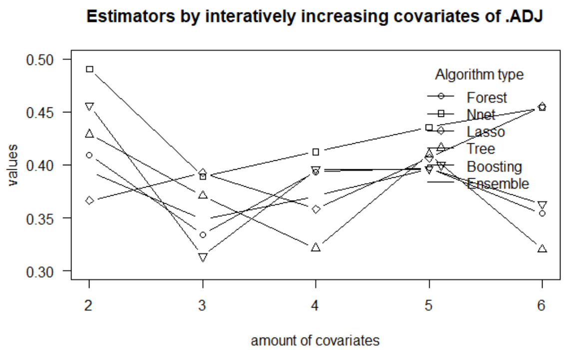


Figure 3.15: Parameter estimations of all DML implementations by iteratively increasing the amount of covariates

# 3.2. Debiased Machine learning

## 3.2.1. Overcoming Regularization Bias

Chernozhukov et al. (2018) depart from the classical machine learning setting by allowing $\eta_0$ to be so high-dimensional that the Donsker properties break down. ”Here, highly complex formally means that the entropy of the parameter space for the nuisance parameter is increasing with the sample size in a way that moves us outside of the traditional framework considered in the classical semi-parametric literature where the complexity of the nuisance parameter space is taken to be sufficiently small. Offering a general and simple procedure for estimating and doing inference on $\theta_0$ that is formally valid in these highly complex settings is the main contribution of [their] paper” (Chernozhukov et al., 2018). Modern analyses aims to model $p$, the amount of confounding factors as increasing with the sample size. This causes traditional assumptions, limiting the complexity of the parameter space for the nuisance parameters to fail. The bias causing naive estimators failing to be $N^{1/2}$ consistent is the result of both the regularization bias and overfitting for estimators $\eta_0$ into estimating equations for $\theta_0$, depicted below in the partially linear model.

The following two sub points outline the key ingredients for DML. Those rely on only weak theoretical requirements and can be applied to the list of the aforementioned ML methods.
By modelling in modern frameworks the size of the covariates $p$, as increasing with the sample size, we go beyond those traditional semi-parametric nuisance parameter assumption.
Chernozhukov et al. (2018) illustrates the problem, as well as the solution of Neyman-Orthogonality and Cross Splitting via the help of the partially linear model of Robinson:

$$Y = D\theta_0 + g_0(X) + U,\ \ E[U|X, D] = 0 \tag{3.25}$$

$$D = m_0(X) + V,\ \ E[V|X] = 0 \tag{3.26}$$

with

$$X = (X_1, ..., X_p) \tag{3.27}$$

where the naive ML estimator is defined as:

$$\widehat{\theta_0} = (\frac{1}{n}\sum_{i\in I} D_i^2)^{-1}\frac{1}{n}\sum_{i\in I} D_i(Y_i - \widehat{g_0}(X_i)), \tag{3.28}$$

however, the estimator for $\theta_0$ fails to have a slower convergence than $1/\sqrt{n}$,

$$|\sqrt{\widehat{\theta_0} - \theta_0}| \to p\,\infty. \tag{3.29}$$

This behavior can be analyzed by decomposing the estimator further into:

$$\sqrt{\widehat{\theta_0} - \theta_0} = (\frac{1}{n}\sum_{i\in I} D_i^2)^{-1}\frac{1}{n}\sum_{i\in I} D_i U_i + (\frac{1}{n}\sum_{i\in I} D_i^2)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I} D_i(g_0(X_i) - \widehat{g_0}(X_i)), \tag{3.30}$$

where b is the second summation of the aforementioned equation

$$b = (ED_i^2)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I} m_0(X_i)(g_0(X_i) - \widehat{g_0}(X_i)) + o_p(1), \tag{3.31}$$

by regularizing we keep the estimator from exploding, however, also introduce a substantive bias, affecting the rate of convergence for $\widehat{g_0}$. Accordingly, $\widehat{g_0}$ will typically be within $n^{-\phi_0}$ where $\phi_0 < \frac{1}{2}$ and b of order $\sqrt{n}n^{-\phi_0}$.

By constructing an "orthogonalized" formulation, we are able to partial out the effect of X from D:

$$V = D - m_0(X), \tag{3.32}$$

$$\widehat{V} = D - \widehat{m_0}(X). \tag{3.33}$$

Via the main sample we obtain:

$$\widehat{\theta_0} = (\frac{1}{n}\sum_{i\in I} \widehat{V_i}D_i)^{-1}\frac{1}{n}\sum_{i\in I} \widehat{V_i}(Y_i - \widehat{g_0}(X_i)). \tag{3.34}$$

Alternatively, we obtain also

$$\widehat{\theta_0} = (\frac{1}{n}\sum_{i\in I} \widehat{V_i}\widehat{V_i})^{-1}\frac{1}{n}\sum_{i\in I} \widehat{V_i}(Y_i - \widehat{l_0}(X_i)) \quad l_0(X) = E[Y|X]. \tag{3.35}$$

The benefits of this auxiliary prediction step , whereby now $\theta_0$ can be clearly interpreted as a linear instrumental variable estimator can be depicted from the aforementioned decomposition of the estimator:

$$\sqrt{n}(\widehat{\theta_0} - \theta_0) = a^* + b^* + c^*, \tag{3.36}$$

whereby $a^*$

$$a^* = (EV^2)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I} V_i U_i \to N(0, \Sigma), \tag{3.37}$$

$b^*$ is

$$b^* = (EV^2)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I}(\widehat{m_0}(X_i) - m_0(X_i))(\widehat{g_0}(X_i) - g_0(X_i)), \tag{3.38}$$

$b^*$ is upper bounded by $\sqrt{n}n^{-\phi_m+\phi_g}$.

Finally, for $c^*$ we want to guarantee via sample splitting that $c^* = o_p(1)$. $c^*$ vanishes indeed in probability by

$$\frac{1}{\sqrt{n}} \sum_{i \in I} V_i(\widehat{g_0}(X_i) - g_0(X_i)) \tag{3.39}$$

and trough Chebyshev's inequality it holds that

$$\frac{1}{\sqrt{n}} \sum_{i \in I} (\widehat{g_0}(X_i) - g_0(X_i))^2 \to p \to \quad 0. \tag{3.40}$$

These two additions to the "traditional" machine learning framework can be summarized as follows:

#### Neyman-orthogoanl moments/scores

In the partially linear model outlined by Chernozhukov et al. in 2018, the the treatment effect estimator's ($\hat{\theta}_0$) rate of convergence in a traditional machine learning set-up is smaller than $n^{-1/2}$. This behavior is caused by the bias of the parameter coefficients $g_0$. Accordingly, the estimator of $\theta_0$ does not converge

$$| \sqrt{n}(\hat{\theta}_0 - \theta_0) | \to \infty. \tag{3.41}$$

Under this context, $\hat{\theta}_0$ is estimated in the following manner

$$\hat{\theta}_0 = (\frac{1}{n} \sum_{i \in I} D_i^2)^{-1} \frac{1}{n} \sum_{i \in I} D_i(Y_i - \hat{g}_0(X_i)). \tag{3.42}$$

However, by applying double prediction i.e. partialling out the effect of X on D, where we make use of Neymann orthogonality we obtain the estimator

$$\check{\theta} = (\frac{1}{n} \sum_{i \in I} \hat{V}_i D_i)^{-1} \frac{1}{n} \sum_{i \in I} \hat{V}_i(Y_i - \hat{g}_0(X_i)), \tag{3.43}$$

which convergences.

#### Cross-fitting as an efficient form of data-splitting

Trough sample splitting our remainder term will contrary to the inital estimator approach 0 in probability

$$\frac{1}{\sqrt{n}} \sum_{i \in I} (\hat{g}_0(X_i) - g_0(X_i))^2 \to 0 \tag{3.44}$$

The applied algorithm goes beyond the partially linear model and has an implementation for IV-estimation or a LATE effect. We specify the number of splits in order to apply data-spiting in the algorithm. Chernozhukov et al. apply 100 splits in 2018. We, however, encounter for the first DML estimation for the parameter coefficient of Nunn (2011), involving 20000 observations and over 60 covariates (level-dummies), a computation duration of approximately 2 hours with a split size of 2. Therefore, we limit for the results of this paper our split size to 2 and regard an analysis of a further split size as a potential future extension.

# 4 Results

Finally, its time to compare DML with OLS/IV (2SLS).

In a first step, we replicate Nunn's and Acemoglu's results in STATA. Secondly, we replicate the same results in R via an assisting IV-package, in order to get familiar with the underlying datasets and in particular the many covariates (see online repository).

For the DML implementation we recommend the unfamiliar reader NOT to test DML code by purely both "plugging in" a dataset and various input parameters into the DoubleML-function but, instead, we recommend the reader to get exposed to the provided examples at first (AJR, Bonus, etc.). In that matter the interested reader gets familiar with the allowed/not-allowed data structure of the required input. According to the principle "making already working things work", we iteratively, fill the provided data of a sample (e.g. AJR) with our variables, whereby we add regressors step by step and hereby test in every step whether an additional covariate causes trouble in the execution of DML. This rather lengthy procedure (histograms of the covariates can give an indication for a failing Neural net covariate), can avoid run-time complications with respect to the rather difficult code. For a more through understanding for this issue, see the provided code in the online repository. Due to the limited debugging function in R and the not easy structure of the underlying functions ML_Functions.R and Moment_Functions.R, the implementation of Chernozhukov et al. can cause tedious troubles which we would like to warn the reader of. Specifying solely a dataset, the input parameters for the DoubleMl-function and pressing CTR+ALT+R will not work in most situation. For the AJR dataset DML begins to work "bug-free" from 43 observations onwards.

The rather extensive size of Nunn's dataset, observing approximately 21000 observations, leads to diverse problems regarding the DML implementation. Encountering around 90 covariates for various coefficients, the sparsity of various dummies leads to problems for the Neural Net implementation. We exclude therefore, around 10% of the data for various categorical control dummy variables in Nunn's work, including various level dummies of occupation and religion. Moreover, we cope with the same issue in the work of Acemoglu. A trade of between the loss of this data and the, consequentially, beneficial omission of the Neural Network DML method is considered in order to replicate the same exact data, used in the OLS/IV research. However, we accept the loss of quality in favor of implementing the Neural Network method in order to compare all provided techniques. This feature represents undoubtedly the biggest limitation of our paper.

While the algorithm analysis in the *Methodology* analyzed a relatively small dataset (AJR) with 64 up to 200 (artificially constructed) observations, Acemoglu's regressions involve around 190 observations if we restrict for dates (e.g. < 1850 and >1200) and exclude Asia, while Nunn's research counts for approximately 21000 observation. Consequently, our algorithm duration rapidly increases. As already stated in the *Methodology*, the replication of the five IV estimations for Table 5 in Nunn's work takes around 10 hours in total with a split size of 2. Consequently, we keep the split size low i.e constant at 2, enabling us potential future extensions with a higher split size in the future. Another limitation of our results.

We apply, based on the normality theory of the DML estimators (Chernozhukov et al., 2018) a t-test in order to check for the statistical significance of the DML coefficients estimators, which are normally distributed. We refer to the Gauss-Markov conditions and the Central Limit Theorem for the DML coefficient testing procedure under this context. We formulate, similar to a standard regression coefficient a null hypothesis, mainly that the DML coefficient is equal to zero. In this case, the t-test is simply the ratio of the sample ML coefficient to its provided standard error. The test statistic follows a t-distribution. In our case the sample DML coefficient is either much bigger than the magnitude of its standard error, hence, statistically significant at the conventional 0.05 critical level (1.96 as critical value) or the other way around, primarily in Acemoglu's DML replication much smaller (insignificant). We test the aforementioned socio-economic results for which several of the discussed coefficients of interest are significant. We divide the DML estimate by the provided standard error in order to obtain a "t-value". This estimator is t-distributed if the expected value for the estimate is zero, which is ergo our underlying null hypothesis ($H_0$). T can be used to get the p-values for a given t-value. That is the probability that a value at least as extreme as the test statistic would be observed under the null hypothesis. Given the null hypothesis that the underling DML parameter value is 0, we reject often for Nunn (2011), however, less often for Acemoglu (2005), concerning the constructed DML coefficient estimates.

In Appendix part 2 we report our identical results for the OLS/ IV replication of the work by Nunn (2011) and Acemoglu (2005).. These traditional coefficient results are for the first paper all significant and negative. In the work of Acemoglu (2005) the coefficient estimators of interest steadily become more significant dependent on the date. With respect to our DML estimates, we report for Acemoglu's and Nunn's work "Median" DML estimates for the coefficients of interest. We report tables for each coefficient depicting the output of each DML technique including an Ensemble of the five aforementioned technique and a final, "Best" technique which synthesize similar to the Ensemble technique, the five methods plus the Hybrid/ Ensemble technique. Note that the results of a table are obtained from calling the DoubleML function twice (s=2). We report the median standard error from across the 2 splits, as well as the report standard errors adjusted for variability across the sample splits using the median method which are, in general, relatively equal. The results imply 2-fold cross-fitting.

## 4.1. DML results of "The Slave Trade and the Origins of Mistrust in Africa" (Nunn, 2011)

### 4.1.1. Table 1

The first OLS coefficient of Nunn's work in Table 1, the exports of slaves in thousands estimator (-0.00068; se: 0.00014) is rather close the DML Lasso estimation (-0.00060). According to our expectation the "best" method, combining various DML techniques retrieves a slightly less negative effect on the trust effect, which is, however, still significant. All DML coefficients are significant and negative on a 0.05 level. The trend of a negative, significant effect of slave exports per ethnic on an individual's trust is confirmed.

OLS coefficient 2, with an estimate of -0.019 and a standard error of 0.005 is again closest captured by the DML method Lasso (-0.017). The other techniques differ by at most 6 hundredths. Overall, the DML estimates are, again, slightly smaller.

OLS coefficient 3, (-0.531; se: 0.147) is greater than the DML coefficients which are significantly negative but between -0.26 and -0.34, hence, again smaller.

OLS coefficient 4, -0.037 with a standard error of 0.014 is not captured by the machine learning techniques. We obtain partially insignificant coefficients (Lasso, Trees) while the significant others range from -0.07 to -0.014, which is much less than the OLS estimator.

OLS coefficient 5, the probably most interesting coefficient in this table, since this explanatory variable is used in the rest of the paper as the explanatory variable (-0.159; se: 0.034), is again not captured by DML. Instead, we obtain significant negative estimates almost half the size smaller.

Finally, OLS coefficient 6 (-0.743; se: 0.187) has a much more negative effect on trust than our DML coefficients, which are efficient and range between -0.33 and -0.44.

We summarize our main findings, mainly that our DML coefficient estimators support Nunn’s negative and significant effect, however, less pronounced as stated by Nunn (2011).

### 4.1.2. Table 2

For the first OLS coefficient we obtain a value of -0.133 and a standard error of 0.037. However, our DML estimators are again smaller and negative. Ensemble and Best, the hybrid techniques, obtain a value of around -0.06 instead.

The second OLS coefficient, estimating the effect on neighbors trust (-0.159; se: 0.034) is again much bigger than the DML estimates which are all significantly negative between -0.048 and -0.079.

The third OLS coefficient which shows the council trust effect (-0.111; se: 0.021) is bigger then DML estimators which are significant and lie around -0.05 to -0.085. Out of all DML coefficient estimators Lasso retrieves the largest negative value which is closest to the OLS estimate, indicating a certain similarity between OLS and the DML Lasso technique.

For the fourth OLS coefficient (-0.144; se 0.032) we obtain DML estimators which are negative and significant but around less than half of the size of OLS, specifically for the DML methods Boosting.

Finally, the last last coefficient of Table 2 (-0.097; se 0.028) retrieves DML estimators which are again negative and significant but around less than half of the size or smaller than the OLS estimator.

### 4.1.3. Table 3

Using DML allows us to relax the linear control assumption and replace it by a weaker assumption, mainly that explanatory variables such as exports can be sufficiently controlled by an unknown function of the distance to the sea, not necessarily linear, which can be learned by DML methods. In this context we are trying to interpret our DML coefficients:

For the first OLS coefficient estimator we obtain a value of -0.178 and a standard error of 0.032. However, our DML estimators are again smaller and find themselves in a neighborhood of -0.11.

The second OLS coefficient value, estimating the neighbors trust via European controls (-0.202; se: 0.031) is again much bigger than the DML estimates which are all significantly negative between -0.09 and -0.13.

The third OLS coefficient estimate which shows the effect on council trust in a controlled model (-0.129; se: 0.022) is bigger then the DML estimators and is for the DML technique RLasso insignificant. The DML methods differ hereby. The DML Neural Net estimator obtains a value of -0.12, se(median) 0.01, while the Ensemble obtains a significantly smaller estimate of 0.037 and se(median) of 0.08. In this instance varying DML techniques clearly differ. The significant DML estimators remain, however, all negative.

For the fourth OLS coefficient estimator (-0.188; se 0.032), measuring the effect on intra group trust, DML estimators are negative and significant but much less than half of the size for the DML methods Boosting for example (around -0.05).

Finally, for the last OLS coefficient estimator (-0.115; se 0.030), we obtain DML estimators which are negative and significant too, but, again, much smaller (around -0.05).

### 4.1.4. Table 5

Using DML allows us to relax linear control assumption not only for a partially linear model but also for a feasible IV-estimation, where we apply useful instruments and construct a 2SLS regressions. Again, the relationship between exogenous variables and the instrument could also be estimated by an underlying unknown function not captured via the linear regression models. The relations can, however, be ”learned” by debiased machine learning methods.

Under this context we interpret the coefficient of both 2SLS from the second stage and DML:

For the first 2SLS coefficient we obtain a value of -0.190 and a se of 0.067. However, our DML estimators are differing again and are for the first time positive. While estimates of DML Rlasso are positive but not significant (se (median has same size), only the negative Neural Net estimator seems to be highly significant (T>3), with a much smaller estimator though (-0.07).

The second 2SLS coefficient(-0.245; se: 0.07), estimating the effect on neighbors trust via European controls and the distance to sea instrument, is again much bigger than the DML estimates. For those, DML Rlasso is positive but insignificant. DML Forest is also insignificant. The DML estimators differ again. While DML Ensemble is negative and insignificant (-0.06; se: 0.081), the Neural Net is significant and twice the size (-0.14; se: 0.05). Only negative estimators are significant while RLasso is positive but insignificant.

The third 2SLS or IV coefficient which shows the council trust effect in a controlled model with an instrument (-0.221; se: 0.060) is captured surprisingly well by the DML estimators. we obtain significant negative DML coefficient estimators between -0.21 and -0.26 with a standard error (se (median)) of around 0.04 to 0.03. This close similarity of both techniques, whereby RLasso captures almost the exact value of the coefficient (-0.21), potentially indicates again the similarity of RLasso and the traditional IV-set up for this instance. DML Ensemble and DML Best show slightly smaller negative coefficients (-0.24). These "more negative" significant coefficient estimators confirm Nunn's hypothesis in a more drastic way and stand in contrast to our hypothesis of a much softer negative effect of exports on trust.

The fourth 2SLS coefficient (-0.251; se: 0.088), measuring the effect on intra group trust in the instrumental set up, outlined in the data section, is significantly negative, like all other coefficients of interest. DML estimators differ again significantly. While all estimators are negative, RLasso and Ensemble are insignificant. Trees, Boosting, Forest and Neural Net obtain coefficients between -0.13 - -0.17 instead. A softer effect of the modified export variable on intra-group trust in this instrumental setting, allowing to capture a more complex nuisance parameter set in our covariate relationship, illustrates for this coefficient the purpose of DML remarkably well.

Finally, for the last 2SLS coefficient (-0.174; se 0.08) DML estimators are mostly insignificant, except the Nnet (0.052; se: 0.03).

We conclude that in an IV-instrumental setting DML results for the research of Nunn (2011) are less pronounced negative and partially insignificant. These results disable us to further disapprove the hypothesis of Nunn (2011). In accordance with his results, our estimators remain mostly negative.

### 4.1.5. Table 6

A feasible IV-estimation with additional controls, extending the model of table 5, is applied. We focus hereby on the coefficient estimators of $ln(1+exports)$. Under this context we interpret again the difference between 2SLS coefficient estimators from the second stage and the DML parameter estimators:

For the first 2SLS coefficient we obtain a value of -0.190 and a se of 0.067. However, our DML estimators are differing again. While Boosting and Forest is insignificant, RLasso the first time positive and significant, Ensembles are insignificant but also positive. The Neural Net remains significant and negative (-0.07; se: 0.02).

The second 2SLS coefficient estimator (-0.245; se: 0.07), estimates the local council trust via European controls and additional instrument controls for the distance to sea instrument. We obtain highly insignificant estimators for all DML techniques. These debiased machine learning results seem erroneous.

The same holds for the third, fourth and fifth coefficient where we obtain positive, negative but, overall, for all DML techniques insignificant results, where the standard error is on average around four times bigger than the estimator.

We conclude the DML research on Nunn (2011) with those results. We encounter most likely erroneous results for most coefficients of table 6. We see that DML can possibly not capture the IV-set up. A potential solution could be the increase of the split size i.e. applying discussed DML theory, which remains for our results constant at 2. This could be a potential extension to this research. Unfortunately, we are, based on these results, not able to rebut

Nunn’s main research finding of an significant, negative effect of slave exports on present day trust[11]. We omit further DML results for this paper.

## 4.2. DML results of ”The rise of Europe: Atlantic trade, institutional change, and economic growth” (Acemoglu, 2005)

### 4.2.1. Table 2

For the comparison of results with our debiased machine learning techniques we focus on the unweighted regressions of Table 2. For completeness, we check for differential growth for countries engaged in Atlantic trade, by focusing on the coefficient estimators of PAT (dummy for Atlantic trader) regressors, linked to a date dummy. We exclude the DML replication for Acemoglu’s joint significance/Wald test for the European dummies. Significant positive PAT estimates imply hereby that Atlantic traders grew, starting in the period where estimators become significant for the first time (dependent on the year dummy).

Column 5

Column 5 describes the unweighed panel for 1300-1500, by using the Atlantic trader regressor (PAT). We obtain for the 1500 OLS estimator an insignificant value of 0.055 with a se of (0.026). The DML estimators are all insignificant for 1500. While the significance and strength of the PAT estimators increases with the date (significant from 1700 onwards in the traditional OLS estimation), the 1600 PAT coefficient remains for all DML techniques insignificant. This DML pattern proceeds for all the upcoming dates (1600,1700,1750,1800,1850). Specifically, for the 1850 coefficient standard errors seem to be at least twice the size of the estimates, while for the DML techniques Ensemble, Best and Boosting they are more than 5 times as high. These results are not in accordance with the traditional results from Acemoglu (2005). We justify this behavior with potential adjustments of the covariate structure (sparsity). Additionally, a positive, increasing trend of coefficient estimator significance, similar to Aecomglu’s coefficient estimators in column 5 table 2 is not detectable for the DML results, which are all insignificant.

For the final row of column 5 (Panel B) we account for potential Atlantic trade x volume of Atlantic trade (logsept2002atradexatlantictrader). The OLS coefficient estimator is 0.016 with a se of 0.0034. Our DML estimators are in this case in accordance with Acemoglu’s results, fluctuating from 0.019 to 0.021 with fairly small standard errors (around 0.003). The DML Ensemble and Best technique depict a ”stronger” effect than the OLS estimation, supporting the results of Acemogly (2005) in a favorable manner.

Column 10

Column 10 describes the unweighed panel for 1300-1500, by using an Atlantic coastline to area measure as an explanatory variable, including various controls (see models). The effects are, interestingly, statistically significant starting in 1700, indicating a developing prosperous trading strategy across the European Atlantic side. For 1500, our DML coefficients remain again highly insignificant, except for the Boosting method. Here, we are closely under the critical 1.96 value ($|1.88|$), and allow for a significance at a p-value of 0.01 instead. The coefficient is, however, negative. These insignificant results confirm repeatedly the findings of Acemoglu (2005).

For the 1600 OLS coefficient estimator, we obtain 0.94 and a se of 0.94 which is again insignificant. The same holds for the DML estimators which are close to null.

For 1750, our OLS coefficient value is 2.60 and has a se of 0.94, hence, significant. All our DML techniques remain, however, insignificant.

For 1850, the OLS coefficient estimator takes, understandably, the biggest value of 4.67 and has a se of 0.94, hence, highly significant and in support of Acemoglu’s research hypotheses of a developing trade. Our DML estimators remain insignificant, however, less insignificant than for the prior dates, indicating a correct direction of the trained DML estimators, which potentially requires further research.

For the final row of column 10 (Panel B), we account again for potential Atlantic trade x volume of Atlantic trade (logsept2002atradexatlantictrader), however, this time in the Atlantic coastline to area regressor set-up. The OLS coefficient estimator is significantly positive with a value of 0.62 and a se of 0.11. Our DML estimators are, again, in accordance with those results. We obtain, except for Trees for all DML methods significant positive estimators raging from 0.48 to 0.78 and a standard error between 0.11 and 0.32. We note, contrary to the findings of Chernozhukov et al., the difference of DML estimation results with respect to both the parameter estimators and the median standard error estimates.

Within the DML estimators, we are not able to observe an insignificant- significant transition similar to the OLS estimators. Panel structure B, involving the logsept2002atradexatlantictrader covariate, seems to be better captured by the DML estimators than the rest of the coefficients of interests, potentially caused by the higher amount of dummies for the Panel A (e.g. more date dummies).

# 5 Conclusion

The imposed structure in the field of traditional econometrics modeling might be unable to cope with a high-dimensional nuisance parameter set in the future. A lot of the discussed methods in this paper do not really fit into the traditional, formally defined, econometric set up with, ideally, ”BLUE” properties. We emphasize the vital use of the discussed, debiased techniques for future economic research. Hereby, we stress in the framework of Chernozhukov et al. in 2018 more the algorithms’ power to deal with relatively large datasets than the formally defined asymptotic statistical properties. Many formal properties remain unfortunately unspecified up to this points, requiring an extensive future research. Similar to the findings of Chernozhukov et al. small sample sizes cause stability problems in training the Neural Network. Another problem for this technique is the encountered sparsity of many dummies.

We conclude the positive scalability of DML with respect to big datasets, able to deal with many predictors, covariates or features. Another positive aspect is hereby that we are able to ”map-reduce” or parallelize problems. However, we also stress the problem of scarcity with regards to the covariates for DML. Hereby, especially the algorithmic Neural Network technique implementation encounters problems which require future research. We stress, based on our results, the importance of Hybrid or Ensemble methods, combining in a beneficial manner positive characteristic traits from various different DML techniques.

Contrary to Chernozhukov et al., the choice of the particular ML method used in estimating nuisance functions does substantively change the conclusion in various of our findings. We obtain different results dependent on which method we employ. Comparing the DML coefficients to traditional parameter estimations of OLS and IV, we obtain spurious results. While, we support Chernozhukov et al.’s conclusion of sufficiently high quality approximations to the underlying nuisance functions via DML, we tend to several times erroneously diverge from the results of traditional techniques, however, with no interpretable pattern. This requires a further investigation. Regarding the fact that we used a constant sample split of two, we cannot make any conclusions regarding the robustness of the the parameter estimates with respect to sample splitting. Finally, we strengthen in this context the complementarity of DML with traditional Econometric techniques. Overall, we conclude that a paradigm shift on the way we traditionally do econometrics might potentially be unavoidable.

# 6 Appendix

Acknowledgments. I would like to thank Mr. Brinkhuis and Mr. Heij.

### 6.0.1. Research Results

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.00060214 | -0.00047417 | -0.00056184 | -0.00046597 | -0.00054004 | -0.0004853 | -0.000482 |
| se(median) | (8.442e-05) | (6.21e-05) | (6.478e-05) | (9.745e-05) | (5.542e-05) | (5.998e-05) | (4.355e-0 |
| se | (3.574e-05) | (3.825e-05) | (3.773e-05) | (4.415e-05) | (3.733e-05) | (4.328e-05) | (4.334e-0 |

Table 6.1: Table 1, first coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.01666405 | -0.01534519 | -0.01548436 | -0.01236733 | -0.01490538 | -0.01307643 | -0.01 |
| se(median) | (0.00131733) | (0.00183794) | (0.00188939) | (0.00302942) | (0.00130426) | (0.00198307) | (0.00 |
| se | (0.0009767) | (0.00105903) | (0.0010579) | (0.00125395) | (0.00104482) | (0.00122418) | (0.00 |

Table 6.2: Table 1, second coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.34947219 | -0.2748388 | -0.33212388 | -0.24451371 | -0.30112995 | -0.26142146 | -0.26 |
| se(median) | (0.06142907) | (0.06845625) | (0.06613624) | (0.08052448) | (0.04685202) | (0.05470592) | (0.04 |
| se | (0.04005127) | (0.03925384) | (0.04109092) | (0.0445157) | (0.04024849) | (0.04462757) | (0.04 |

Table 6.3: Table 1, third coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.00720739 | -0.00072657 | -0.01113165 | -0.0144293 | -0.01278326 | -0.01412898 |
| se(median) | (0.00528172) | (0.01250772) | (0.0041817) | (0.00569127) | (0.00799942) | (0.00378452) |
| se | (0.00328304) | (0.00331427) | (0.00334433) | (0.00380794) | (0.00332237) | (0.00374652) |

Table 6.4: Table 1, fourth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.09892208 | -0.08530543 | -0.09333447 | -0.08953858 | -0.08794845 | -0.09074742 |
| se(median) | (0.01178898) | (0.01031396) | (0.00932804) | (0.01159752) | (0.00959981) | (0.00969391) |
| se | (0.00797732) | (0.00822755) | (0.00828625) | (0.00961935) | (0.00820974) | (0.0094637) |

Table 6.5: Table 1, fifth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.44744788 | -0.3417925 | -0.43717029 | -0.33729181 | -0.38658041 | -0.34967693 |
| se(median) | (0.08159923) | (0.08641725) | (0.08649801) | (0.08732787) | (0.06545477) | (0.06005528) |
| se | (0.04893368) | (0.04966838) | (0.05058242) | (0.05738749) | (0.05030979) | (0.05628493) |

Table 6.6: Table 1, sixth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.08656522 | -0.08080327 | -0.08208676 | -0.06273562 | -0.07519046 | -0.06546162 |
| se(median) | (0.00824706) | (0.01453857) | (0.01405287) | (0.01520651) | (0.00940411) | (0.00852635) |
| se | (0.00729876) | (0.00720743) | (0.00707177) | (0.00611972) | (0.00697657) | (0.00617606) |

Table 6.7: Table 2, first coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.07952039 | -0.0651328 | -0.07080106 | -0.04874489 | -0.06828319 | -0.05164454 |
| se(median) | (0.01106949) | (0.0109778) | (0.0133802) | (0.02018957) | (0.00782621) | (0.01380939) |
| se | (0.00665708) | (0.00648022) | (0.00650689) | (0.00570457) | (0.00644988) | (0.00574728) |

Table 6.8: Table 2, second coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.08573544 | -0.07509353 | -0.07407893 | -0.05371172 | -0.07664708 | -0.05644775 |
| se(median) | (0.00915245) | (0.01412522) | (0.0125844) | (0.0189703) | (0.00635769) | (0.01344631) |
| se | (0.00607114) | (0.00585286) | (0.005914) | (0.00520475) | (0.00593048) | (0.00525491) |

Table 6.9: Table 2, third coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.0877548 | -0.0664323 | -0.07274705 | -0.04926524 | -0.07957221 | -0.05289904 |
| se(median) | (0.01435736) | (0.01050077) | (0.01523571) | (0.02221263) | (0.01198115) | (0.0156239) |
| se | (0.00673095) | (0.00636709) | (0.00650715) | (0.00566661) | (0.00653045) | (0.00573267) |

Table 6.10: Table 2, fourth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.04535312 | -0.03716246 | -0.03421165 | -0.02753794 | -0.03568167 | -0.02814612 |
| se(median) | (0.00843355) | (0.00909245) | (0.00782993) | (0.01171105) | (0.00640457) | (0.00727244) |
| se | (0.00650025) | (0.00633803) | (0.00626319) | (0.00548852) | (0.00629326) | (0.00554612) |

Table 6.11: Table 2, fifth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.11652189 | -0.10582163 | -0.10878998 | -0.10402451 | -0.09988534 | -0.11859325 | -0.11 |
| se(median) | (0.01596533) | (0.00906277) | (0.01175008) | (0.01680798) | (0.01150798) | (0.02200657) | (0.01 |
| se | (0.01369739) | (0.00852786) | (0.00861878) | (0.00991226) | (0.00847351) | (0.01379957) | (0.01 |

Table 6.12: Table 3, first coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.13465394 | -0.09422649 | -0.10238293 | -0.09527674 | -0.09886842 | -0.12908491 | -0.12 |
| se(median) | (0.02941527) | (0.00995863) | (0.01904519) | (0.03789186) | (0.01028399) | (0.03490648) | (0.01 |
| se | (0.01401861) | (0.00830526) | (0.0084391) | (0.00979719) | (0.00834091) | (0.01390955) | (0.01 |

Table 6.13: Table 3, second coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.02478993 | -0.13495362 | -0.12594918 | -0.11815808 | -0.12642125 | -0.03758452 | -0.03 |
| se(median) | (0.06331441) | (0.01599806) | (0.0495778) | (0.08764491) | (0.01023058) | (0.08609905) | (0.01 |
| se | (0.01533147) | (0.0093245) | (0.00940991) | (0.01080381) | (0.00932055) | (0.01540926) | (0.01 |

Table 6.14: Table 3, third coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.11638922 | -0.09814093 | -0.10651819 | -0.09745635 | -0.11374028 | -0.11248066 | -0.112 |
| se(median) | (0.0187342) | (0.00939597) | (0.01076418) | (0.01969294) | (0.01455671) | (0.01787421) | (0.014 |
| se | (0.0142022) | (0.00849277) | (0.00850906) | (0.00996404) | (0.00843556) | (0.01405143) | (0.014 |

Table 6.15: Table 3, fourth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.05480684 | -0.05507594 | -0.05086515 | -0.04925065 | -0.0554197 | -0.04811934 | -0.04 |
| se(median) | (0.01416996) | (0.01155669) | (0.0089661) | (0.01313429) | (0.00880441) | (0.01534275) | (0.01 |
| se | (0.01411555) | (0.00821851) | (0.00840971) | (0.0097735) | (0.00827112) | (0.01395591) | (0.01 |

Table 6.16: Table 3, fifth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | 0.13618663 | -0.06425625 | -0.07334738 | -0.08866565 | -0.07858426 | 0.02220043 | 0.022 |
| se(median) | (0.12220518) | (0.02635881) | (0.06073071) | (0.16974629) | (0.02212151) | (0.1128902) | (0.03 |
| se | (0.03914856) | (0.02027391) | (0.01925347) | (0.02399267) | (0.01901928) | (0.03930884) | (0.03 |

Table 6.17: Table 5, first coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | 0.03722486 | -0.07320716 | -0.1100403 | -0.12798514 | -0.14098404 | -0.0639221 | -0.06 |
| se(median) | (0.07662663) | (0.05050106) | (0.04672788) | (0.11774651) | (0.05346359) | (0.08126695) | (0.04 |
| se | (0.0404734) | (0.02098214) | (0.01987183) | (0.02463932) | (0.02044064) | (0.04029677) | (0.04 |

Table 6.18: Table 5, second coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| treatment effect | -0.21384984 | -0.23685679 | -0.23166309 | -0.22072877 | -0.26229246 | -0.24060037 | -0.24 |
| se(median) | (0.04792731) | (0.02628372) | (0.03031193) | (0.0293841) | (0.03574659) | (0.0528265) | (0.04 |
| se | (0.04514844) | (0.02395111) | (0.02214822) | (0.02692254) | (0.02206652) | (0.04543225) | (0.04 |

Table 6.19: Table 5, third coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -0.02549455 | -0.13663212 | -0.14570247 | -0.14080609 | -0.17351854 | -0.07812691 |
| se(median) | (0.07877358) | (0.02489664) | (0.05131923) | (0.09254135) | (0.03838218) | (0.08915756) |
| se | (0.04018401) | (0.02176613) | (0.02000672) | (0.02439728) | (0.02032763) | (0.0402062) |

Table 6.20: Table 5, fourth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | 0.13058854 | -0.02202019 | -0.02076911 | -0.02641777 | -0.05211717 | 0.08789482 |
| se(median) | (0.1007223) | (0.02062728) | (0.07441574) | (0.13798567) | (0.03702185) | (0.13383597) |
| se | (0.04146115) | (0.02051053) | (0.0202633) | (0.02497105) | (0.02032195) | (0.04149346) |

Table 6.21: Table 5, fifth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | 0.40236773 | -0.06425625 | -0.07334738 | -0.08866565 | -0.07858426 | 0.26179748 |
| se(median) | (0.2606956) | (0.02635881) | (0.17986954) | (0.42144875) | (0.02212151) | (0.34888843) |
| se | (0.05070659) | (0.02027391) | (0.01925347) | (0.02399267) | (0.01901928) | (0.049054) |

Table 6.22: Table 6, first coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | 11.52532819 | 1.75446522 | 1.6575228 | 0.79139092 | 1.51157496 | 7.18542578 |
| se(median) | (70.9144325) | (3.62561436) | (4.90498023) | (8.64306373) | (2.08165376) | (29.9908797 |
| se | (70.60402508) | (3.57851215) | (2.9641446) | (1.00982887) | (2.06356113) | (29.3770510 |

Table 6.23: Table 6, second coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | 5.29543087 | 1.71525366 | 1.75737683 | 0.7860938 | 2.11175163 | 4.35342407 |
| se(median) | (22.6414384) | (5.51598434) | (5.44762816) | (4.41527545) | (7.06002298) | (17.1034309) |
| se | (22.52296281) | (5.49735459) | (5.12266744) | (1.64805136) | (7.04726343) | (16.8367418 |

Table 6.24: Table 6, third coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | -2.38343249 | -0.30713467 | -0.26018581 | 0.41030996 | -0.69509715 | -0.46581525 |
| se(median) | (20.52015949) | (2.4375484) | (1.77159855) | (2.82242721) | (2.26413538) | (9.00245297) |
| se | (20.47720361) | (2.38745944) | (1.74240872) | (2.09282645) | (2.20629127) | (8.95980232) |

Table 6.25: Table 6, fourth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble |
|---|---|---|---|---|---|---|
| treatment effect | 8.42181488 | 3.73120469 | 1.14084894 | 0.45349398 | 0.6700401 | 6.10871932 |
| se(median) | (79.73407906) | (22.4833501) | (4.22610616) | (7.27079846) | (2.63191989) | (40.515158 |
| se | (79.62880609) | (22.28878628) | (2.55118798) | (2.07616676) | (1.76729642) | (40.110970 |

Table 6.26: Table 6, fifth coefficient

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | -0.0157 | -0.0076 | 0.0314 | -0.0094 | 0.0447 | 0.0146 | 0.0158 |
| se(median) | (0.0204) | (0.0302) | (0.0307) | (0.0544) | (0.0401) | (0.0379) | (0.0394) |
| se | (0.0194) | (0.0234) | (0.0227) | (0.0503) | (0.023) | (0.036) | (0.0384) |

Table 6.27: Europe: Table 2, first first-column coefficient of Appendix 2: 1500

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | -0.0147 | -0.0271 | 0.0144 | -0.0879 | -0.0035 | -0.0442 | -0.0444 |
| se(median) | (0.0206) | (0.0683) | (0.053) | (0.0823) | (0.077) | (0.0833) | (0.0693) |
| se | (0.0188) | (0.0259) | (0.0299) | (0.058) | (0.0336) | (0.0435) | (0.0471) |

Table 6.28: Europe: Table 2, second first-column coefficient of Appendix 2: 1600

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | 7e-04 | -0.0208 | 0.0171 | 0.0021 | -0.0145 | -0.0264 | -0.0139 |
| se(median) | (0.0264) | (0.0381) | (0.04) | (0.0409) | (0.0314) | (0.0472) | (0.0442) |
| se | (0.0225) | (0.022) | (0.0137) | (0.0304) | (0.0283) | (0.0293) | (0.0273) |

Table 6.29: Europe: Table 2, third first-column coefficient of Appendix 2: 1750

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | 0.0388 | -0.0363 | 0.0094 | 0.0234 | 0.0108 | -0.0038 | 0.0115 |
| se(median) | (0.064) | (0.0634) | (0.0437) | (0.0563) | (0.0608) | (0.0764) | (0.0716) |
| se | (0.0422) | (0.0333) | (0.0417) | (0.055) | (0.0539) | (0.0567) | (0.0604) |

Table 6.30: Europe: Table 2, fourth first-column coefficient of Appendix 2: 1850

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | 0.021121 | 0.019704 | 0.01913 | 0.019175 | 0.019831 | 0.020094 | 0.021145 |
| se(median) | (0.003039) | (0.003262) | (0.005277) | (0.00436) | (0.006402) | (0.006175) | (0.006897) |
| se | (0.002887) | (0.003209) | (0.004778) | (0.003853) | (0.006365) | (0.005886) | (0.006876) |

Table 6.31: Europe: Table 2, last first-column coefficient of Appendix 2:structured via logsept2002atradexatlantictrader

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | -0.5897 | 0.3096 | -1.1242 | -0.8291 | -0.9536 | -0.8548 | -0.5968 |
| se(median) | (1.0136) | (1.5487) | (0.5948) | (0.7669) | (1.2438) | (0.7601) | (0.7468) |
| se | (0.8876) | (0.8623) | (0.4827) | (0.581) | (0.5977) | (0.6165) | (0.7179) |

Table 6.32: Europe: Table 2, first second-column coefficient of Appendix 2: 1500

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | -0.0311 | 0.0106 | -0.0053 | 0.0072 | -0.0045 | -0.0051 | 0.0014 |
| se(median) | (0.0332) | (0.0475) | (0.0349) | (0.0483) | (0.0328) | (0.0377) | (0.0408) |
| se | (0.0194) | (0.0454) | (0.0344) | (0.041) | (0.0317) | (0.0369) | (0.04) |

Table 6.33: Europe: Table 2, second second-column coefficient of Appendix 2: 1600

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | -0.0305 | 0.0151 | 0.0129 | 0.0205 | 0.0102 | 0.006 | 0.0129 |
| se(median) | (0.0348) | (0.0476) | (0.0399) | (0.0597) | (0.036) | (0.0438) | (0.0425) |
| se | (0.0191) | (0.047) | (0.0367) | (0.0449) | (0.0334) | (0.0388) | (0.0422) |

Table 6.34: Europe: Table 2, third second-column coefficient of Appendix 2: 1750

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | best |
|---|---|---|---|---|---|---|---|
| Median ATE | -0.06 | 2.47 | 0.7 | 0.13 | 0.1 | 0.13 | 1.87 |
| se(median) | (1.29) | (2.22) | (1.04) | (1.45) | (1.49) | (0.31) | (1.37) |
| se | (0.04) | (0.81) | (0.79) | (0.19) | (0.1) | (0.23) | (1.04) |

Table 6.35: Europe: Table 2, fourth second-column coefficient of Appendix 2: 1850

| | RLasso | Trees | Boosting | Forest | Nnet | Ensemble | be |
|---|---|---|---|---|---|---|---|
| Median ATE | 0.485101 | 0.389417 | 0.786693 | 0.607283 | 0.614922 | 0.648031 | 0.614894 |
| se(median) | (0.111889) | (0.321794) | (0.250833) | (0.213111) | (0.199119) | (0.231644) | (0.222732) |
| se | (0.092226) | (0.090021) | (0.17141) | (0.15254) | (0.180094) | (0.185439) | (0.180105) |

Table 6.36: Europe: Table 2, last second-column coefficient of Appendix 2:structured via logsept2002atradexatlantictrader

# Bibliography

[1] Daron Acemoglu, Simon Johnson, and James Robinson. The rise of europe: Atlantic trade, institutional change, and economic growth. *American economic review*, 95(3): 546–579, 2005.

[2] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650, 2014.

[3] Alexandre Belloni, Victor Chernozhukov, Ivan Fernández-Val, and Christian Hansen. Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85 (1):233–298, 2017.

[4] Alberto Bisin and Thierry Verdier. "beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics*, 115(3):955–988, 2000.

[5] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

[6] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[7] A. K. Geim and H. A. M. S. ter Tisha. Detection of earth rotation with a diamagnetically levitating gyroscope. *Physica B: Condensed Matter*, 294–295:736–739, 2001. doi: 10. 1016/S0921-4526(00)00753-5.

[8] Luigi Guiso, Paola Sapienza, and Luigi Zingales. Social capital as good culture. *Journal of the European Economic Association*, 6(2-3):295–320, 2008.

[9] Stephen Knack and Philip Keefer. Does social capital have an economic payoff? a cross-country investigation. *The Quarterly journal of economics*, 112(4):1251–1288, 1997.

[10] Jarrod Moss, Kenneth Kotovsky, and Jonathan Cagan. The role of functionality in the mental representations of engineering students: Some differences in the early stages of expertise. *Cognitive Science*, 30(1):65–93, 2006.

[11] Nathan Nunn and Leonard Wantchekon. The slave trade and the origins of mistrust in africa. *American Economic Review*, 101(7):3221–52, 2011.

[12] David Rindos, Robert L Carneiro, Eugene Cooper, Paul Drechsel, Robert C Dunnell, RF Ellen, CJMR Gullick, Robert A Hackenberg, John Hartung, John H Kunkel, et al. Darwinian selection, symbolic variation, and the evolution of culture [and comments and reply]. *Current anthropology*, 26(1):65–88, 1985.

[13] Guido Tabellini. Culture and institutions: economic development in the regions of europe. *Journal of the European Economic association*, 8(4):677–716, 2010.

[14] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[15] Aad Van Der Vaart. On differentiable functionals. *The Annals of Statistics*, pages 178–204, 1991.

# Appendix part 2

## Nunn (2011)

Table 1: Table 1- OLS Estimates of the Determinants of Trust in Neighbors

| Dependent variable: Trust of neighbors | Slave exports (thousands) (1) | Exports/ area (2) | Exports/ historical pop (3) | ln(1+exports) (4) | ln(1+exports/area) (5) | ln(1+exports/historical pop) (6) |
|---|---|---|---|---|---|---|
| Estimated coeffient | -0.00068 | -0.019 | -0.531 | -0.037 | -0.159 | -0.743 |
| std adj. clust | 0.00014 | 0.005 | 0.147 | 0.014 | 0.034 | 0.187 |
| std adj. 2-clust | 0.00015 | 0.005 | 0.147 | 0.014 | 0.034 | 0.187 |
| std adj. spatial | 0.00013 | 0.005 | 0.165 | 0.015 | 0.034 | 0.212 |
| Individual controls | Yes | Yes | Yes | Yes | Yes | Yes |
| District controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 20,027 | 20,027 | 17,644 | 20,027 | 20,027 | 17,644 |
| Number of ethnicities | 185 | 185 | 157 | 185 | 185 | 157 |
| Number of districts | 1,257 | 1,257 | 1,214 | 1,257 | 1,257 | 1,214 |
| R^2 | 0.16 | 0.16 | 0.15 | 0.15 | 0.16 | 0.15 |

Table 2: Table 2 OLS Estimates of the Determinants of the Trust of Others

| | Trust of relatives (1) | Trust of neighbors (2) | Trust of local council (3) | Intra group trust (4) | Inter-group trust (5) |
|---|---|---|---|---|---|
| ln(1+exports/area) | -0.133 | -0.159 | -0.111 | -0.144 | -0.097 |
| std adj. 2-clust | 0.037 | 0.034 | 0.021 | 0.032 | 0.028 |
| Individual controls | Yes | Yes | Yes | Yes | Yes |
| District controls | Yes | Yes | Yes | Yes | Yes |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 20,062 | 20,027 | 19,733 | 19,952 | 19,765 |
| Number of ethnicity clusters | 185 | 185 | 185 | 185 | 185 |
| Number of district clusters | 1,257 | 1,257 | 1,283 | 1,257 | 1,255 |
| R^2 | 0.13 | 0.16 | 0.20 | 0.14 | 0.11 |

Table 3: Table 3 OLS Estimates of the Determinants of the Trust of Others, with Additional Controls

| | Trust of relatives | Trust of neighbors | Trust of local council | Intra group trust | Inter-group trust |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| ln(1+exports/area) | -0.178 | -0.202 | -0.129 | -0.188 | -0.115 |
| std adj. 2-clust | 0.032 | 0.031 | 0.022 | 0.033 | 0.030 |
| Colonial population density | Yes | Yes | Yes | Yes | Yes |
| Ethnicity-level colonial controls | Yes | Yes | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes | Yes | Yes |
| District controls | Yes | Yes | Yes | Yes | Yes |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 16,709 | 16,679 | 15,905 | 16,636 | 16,473 |
| Number of ethnicity clusters | 147 | 147 | 146 | 147 | 147 |
| Number of district clusters | 1,187 | 1,187 | 1,194 | 1,186 | 1,184 |
| R^2 | 0.13 | 0.16 | 0.21 | 0.16 | 0.12 |

Table 4: Table 5 IV Estimates of the Effect of the Slave Trade on Trust

| | Trust of relatives | Trust of neighbors | Trust of local council | Intragroup trust | Intergroup trust |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Second stage: Dependent variable is an individual's trust | | | | | |
| ln(1+exports/area) | -0.190 | -0.245 | -0.221 | -0.251 | -0.174 |
| std adj. 2-clust | 0.067 | 0.070 | 0.060 | 0.088 | 0.080 |
| Hausman test (p-value) | 0.88 | 0.53 | 0.09 | 0.44 | 0.41 |
| R^2 | 0.13 | 0.16 | 0.20 | 0.15 | 0.12 |
| First stage: Dependent variable is ln(1+exports/area) | | | | | |
| Historical distance of ethnic group from coast | -0.0014 | -0.0014 | -0.0014 | -0.0014 | -0.0014 |
| std | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 |
| Colonial population density | Yes | Yes | Yes | Yes | Yes |
| Ethnicity-level colonial controls | Yes | Yes | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes | Yes | Yes |
| District controls | Yes | Yes | Yes | Yes | Yes |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 16,709 | 16,679 | 15,905 | 16,636 | 16,473 |
| Number of (ethnicity) clusters | 147/1,187 | 147/1,187 | 146/1,194 | 147/1,186 | 147/1,184 |
| F-stat of excl. instrument | 26.9 | 26.8 | 27.4 | 27.1 | 27.0 |
| R^2 | 0.81 | 0.81 | 0.81 | 0.81 | 0.81 |

Table 5: Table 6 IV Estimates of the Effect of the Slave Trade on Trust, with Additional Controls

| | Trust of relatives | Trust of neighbors | Trust of local council | Intragroup trust | Intergroup trust |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Second stage: Dependent variable is an individual's trust | | | | | |
| ln(1+exports/area) | -0.172 | -0.271 | -0.262 | -0.254 | -0.189 |
| std adj. 2-clust | 0.076 | 0.088 | 0.075 | 0.109 | 0.103 |
| Hausman test (p-value) | 0.98 | 0.42 | 0.05 | 0.53 | 0.44 |
| R^2 | 0.13 | 0.16 | 0.20 | 0.15 | 0.12 |
| First stage: Dependent variable is ln(1+exports/area) | | | | | |
| Historical distance of ethnic group from coast | -0.0015 | -0.0015 | -0.0015 | -0.0015 | -0.0015 |
| std | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 |
| Colonial population density | Yes | Yes | Yes | Yes | Yes |
| Ethnicity-level colonial controls | Yes | Yes | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes | Yes | Yes |
| District controls | Yes | Yes | Yes | Yes | Yes |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 16,709 | 16,679 | 15,905 | 16,636 | 16,473 |
| Number of (ethnicity) clusters | 147/1,187 | 147/1,187 | 146/1,194 | 147/1,186 | 147/1,184 |
| F-stat of excl. instrument | 21.7 | 21.6 | 22.2 | 21.8 | 21.6 |
| R^2 | 0.81 | 0.81 | 0.81 | 0.81 | 0.81 |

Table 7: Table 10

| Variable of interest<br>ln(1+exports) (1*)<br>loc_ln(1+exports) (2*) | Trust of relatives<br>(1) | Trust of neighbors<br>(2) | Trust of local council<br>(3) | Intragroup trust<br>(4) |
|---|---|---|---|---|
| Estimated coefficient (1*) | -0.155 | -0.182 | -0.100 | -0.169 |
| std | 0.029 | 0.029 | 0.023 | 0.033 |
| Estimated coefficient (2*) | -0.045 | -0.045 | -0.045 | -0.043 |
| std | 0.014 | 0.016 | 0.018 | 0.018 |
| Colonial population density | Yes | Yes | Yes | Yes |
| Ethnicity-level colonial controls | Yes | Yes | Yes | Yes |
| Baseline controls | Yes | Yes | Yes | Yes |
| Country fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 15,999 | 15,972 | 15,221 | 15,931 |
| Number of clusters | 146/269 | 146/269 | 145/272 | 146/269 |
| R^2 | 0.13 | 0.16 | 0.20 | 0.16 |

Table 6: Table 9

| | Trust of local council | | Intergroup trust | | |
|---|---|---|---|---|---|
| Variable of interest<br>ln(1+exports) (1*)<br>townvill (2*) | (1) | (2) | (3) | (4) | (5) |
| Estimated coefficient (1*) | -0.072 | -0.070 | -0.102 | -0.120 | -0.098 |
| std | 0.019 | 0.019 | 0.028 | 0.027 | 0.029 |
| Estimated coefficient (2*) | | | -0.037 | -0.063 | -0.091 |
| std | | | 0.029 | 0.030 | 0.035 |
| Council trustworthiness fixed effects | Yes | Yes | No | No | No |
| Five public goods fixed effects | No | Yes | No | No | No |
| Colonial population density | Yes | Yes | Yes | Yes | Yes |
| Ethnicity-level colonial controls | Yes | Yes | Yes | Yes | Yes |
| Baseline controls | Yes | Yes | Yes | Yes | Yes |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 12,827 | 12,203 | 9,673 | 12,513 | 15,999 |
| Number of clusters | 146/1,172 | 145/1,130 | 147/725 | 147/737 | 147/1,127 |
| R^2 | 0.37 | 0.37 | 0.12 | 0.12 | 0.12 |

## Acemoglu (2005)

Table 8: Table 2

| | Panel, 1300-1850, unweighted (5) | Panel, 1300-1850 unweighted (10) |
|---|---|---|
| Potential for Atlantic trade | 0.055 | 0.75 |
| x 1500 | (0.026) | (0.87) |
| Potential for Atlantic trade | 0.0495 | 0.94 |
| x 1600 | (0.028) | (0.94) |
| Potential for Atlantic trade | 0.071 | 2.01 |
| x 1700 | (0.028) | (0.94) |
| Potential for Atlantic trade | 0.073 | 2.60 |
| x 1750 | (0.028 | (0.94) |
| Potential for Atlantic trade | 0.110 | 3.76 |
| x 1800 | (0.028) | (0.94) |
| Potential for Atlantic trade | 0.115 | 4.67 |
| x 1850 | (0.028) | (0.94) |
| obs/R^2 | 0.82/192 | 0.83/192 |
| Panel B | | |
| Potential for Atlantic trade | 0.016 | 0.62 |
| x volume of Atlantic trade | (0.0034) | (0.11) |
| obs/R^2 | 0.81/192 | 0.82/192 |

Table 9: Table 3

| | Panel, 1300-1850, unweighted (5) | Panel, 1300-1850 unweighted (10) |
|---|---|---|
| Potential for Atlantic trade | 0.16 | 3.42 |
| x 1600 | (0.07) | (2.21) |
| Potential for Atlantic trade | 0.21 | 6.32 |
| x 1700 | (0.07) | (2.21) |
| Potential for Atlantic trade | 0.18 | 8.06 |
| x 1800 | (0.07) | (2.21) |
| obs/R^2 | 0.96/96 | 0.96/96 |
| Panel B | | |
| Potential for Atlantic trade | 0.047 | 2.22 |
| x volume of Atlantic trade | (0.018) | (0.58) |
| obs/R^2 | 0.96/96 | 0.96/96 |

| | Panel, 1300-1850, controlling for religion | Panel 1300 to 1850, controlling for wars | Panel 1300 to 1850, controlling for Roman heritage | Panel 1300 to 1820, controlling for latitude |
|---|---|---|---|---|
| Panel A | | | | |
| Atlantic trader dummy x volume of Atlantic trade | 0.013 | 0.011 | 0.011 | 0.011 |
| | (0.002) | (0.003) | (0.002) | (0.002) |
| R^2/obs | 0.89/192 | 0.89/176 | 0.89/192 | 0.89/192 |
| Panel C | | | | |
| Coastline to area x volume of Atlantic trade | 0.79 | 0.76 | 0.75 | 0.78 |
| | (0.08) | (0.08) | (0.07) | (0.07) |
| R^2/obs | 0.93/192 | 0.93/176 | 0.92/192 | 0.93/192 |