Debiased Machine Learning; establishment of valid causal inferences

Tessa Hoyng, 410128
Erasmus University Rotterdam
Erasmus School of Economics
Bachelor thesis: Econometrics & Operations Research
Supervisor: Naghi A.A., Second assessor: Grith M.

July 16, 2018

Abstract

The Debiased Machine Learning (DML) methods are only recently introduced in the field of econometrics. DML can be used with multiple Machine Learning (ML) methods for the estimation of a dataset. ML methods suffer from the regulation and the over-fitting bias when applied to causal inference problems. DML removes these biases through the use of the Neyman-orthogonal restriction and sample splitting. In this paper the causality between parameters established in three different papers, is investigated by applying the DML methods to the datasets. These papers used standard estimation methods to obtain their results. This paper shows that the causality between parameters established in two papers can be held unreliable. The results of this paper display that standard methods perform poor when applied to causal inference problems, especially when there is a large set of covariates used in the estimation of a regression. This displays the relevance of this paper and gives the researchers of the articles a reason to review their results.

1 Introduction

This paper gives insight in new 'modern' methods; debiased machine learning(DML). The main aim of DML is to provide a good framework for estimating and interpreting a low-dimensional parameter, θ_0, in the presence of a high-dimensional nuisance parameter, η_0, with the new generation of non-parametric methods, branded as 'Machine Learning'(ML). To estimate η_0, machine learning methods are used.

This paper will highlight multiple machine learning methods, that are used in the paper on DML by Chernozhukov et al.(2018). Machine learning methods perform well in the prediction performance and reducing variance in the estimation of θ_0. In order to obtain a good prediction function a ML method regularizes the estimates by shrinking the non-important estimates to zero, which results in a few non-zero estimates. The downside of this regularization is that the distribution of the estimated $\tilde{\theta}_0$ gets biased. The distribution of $\tilde{\theta}_0$ additionally suffers from the over-fitting bias, when the whole sample is used to estimate θ_0. To summarize, the estimation of θ_0 suffers from regulation bias and over-fitting bias when you naively plug in the ML estimators of η_0 in the estimation equations of θ_0. DML methods remove these biases by the use of the Neyman-orthogonal moments and the use of cross-fitting in their algorithms.

In the estimation of regressions multiple standard methods/models are applied in the field of economics and econometrics. These standard methods perform poorly if the dataset contains lots of covariates, due to invertible issues in the estimation of the coefficients. In the paper by Athey (2017) multiple benefits about the use of machine learning methods are clarified.
Firstly, machine learning methods make an improvement in estimating a regression when there are multiple covariates, especially when there are a lot of covariates compared to the total of observations in a regression. The machine learning methods have a lot of power in flexibly selecting a functional form for a given dataset.
Secondly, one of the major advantages of applying ML to a dataset is that ML methods provide researchers not only with better performance of the estimations but they also enable them to completely describe the process of the model selection. This is in contrast with common economic approaches where the researcher selects a model based on certain assumptions and estimates it only one time. ML methods build in 'tuning' in their algorithms and therefor estimate and compare many models based on the provided data. This 'tuning' ensures that the model selection is fully described.
Another advantage of using machine learning is that it makes the assumption of unconfoundness in the treatment variable, which means that there is a randomized treatment variable used in order to control the estimation of the effects of the covariates on the variable of interest (Rosenbaum, 2000).
By removing the biases of which machine learning suffers, DML provides methods that have all the benefits of the use of ML and gives accurate results about the causality between the variables. Therefor DML can establish a causal inference between the treatment effect on θ_0, under the assumption of unconfoundness. Finally, DML methods are semi-parametric methods that allow the nuisance parameters to be nonparametric. This means that no parametric assumptions need

to made about the way covariates affect the outcome. Linearity is a parametric assumption, therefor most of the standard tools like OLS do not capture non-linearity.

The main aim of this paper is to give insights in these new 'modern' methods, DML, and show the relevance of the use of DML by applying these relatively new methods to the datasets of three articles and discuss their main results. The focus of this paper is to see if the causal inferences found in the papers hold after applying DML to the datasets. First this paper replicates the results of the papers by performing the standard methods to their dataset. This replication is useful because it could establish if there exists the same causality between the parameters as in the papers. In the case in which the results are compatible, it can be concluded that the standard methods are executed correctly. This allows this paper to compare the results obtained by the standard methods with results obtained by DML. The first article, the Slave Trade and Origins of Mistrust in Africa, examines through which channels the Africa's slave trade could affect the economic development nowadays (Nunn & Wantchekon, 2011). They focus mainly on the causality between slave trade and various measures of trust. This paper contains a very large dataset with a great amount of covariates which can cause standard estimation methods to perform poorly. This makes it relevant to investigate their main results with the DML method. Due to the large dataset, the causal effect between the variable slave trade and the various measures of trust that the paper established differ could from the causal effect established after applying DML to their datasets.
The second paper, Inherited Trust and Growth investigates the causality between trust and economic growth by concentrating on the inherited element of trust and its fluctuations over time on long periods (Algan & Cahuc, 2010). The conclusion of the paper is established by adding multiple covariates to several regressions. Therefore it is interesting to investigate if their main conclusion holds after applying the DML method to their dataset.
Finally the last paper analyses the effect of corporate taxes on investment and entrepreneurship (Djankov et al., 2010). To evaluate and model a tax policy its relevant to understand the causal effect of corporate taxes. Additionally this causal effect is relevant for economic growth. The paper suggest that after controlling for an additional 12 variables, the corporate tax rates lose all their significant influence on investment and entrepreneurship. This paper will investigate their results with the DML method when all the control variables are used. The paper contains a small dataset in terms of observations in combination with 12 covariates which could lead to a different conclusion when using DML. It would be interesting to determine if the causal effect between corporate tax rates and investment or entrepreneurship is also rejected when applying DML to their dataset.

2 Literature

The causality between slave trade and trust in Africa:
The article is an investigation of additional value to previous research, such as the paper by Guiso, Sapienza and Zingales (2004). In this article they display that the behavior of people that move in Italy is influenced by the amount of social capital of the region where the people were born. The article by Nunn and Wantchekon (2011) combines current survey data from

respondents with historical data on slave exports organized by ethnic group. The main question of the article is whether the slave trade in the past has caused a culture to grow of mistrust in Africa. Originally slaves were captured through organized arrest arranged by the state, this caused an environment of insecurity to grow which made individuals turn on each other. They kidnapped and sold their friends and family into slavery. The slave trade took place in or between multiple groups, through local government organizations, between neighbors or in families. Therefore, this paper makes a distinction in five different measures of trust. The measure of trust entails how much trust a respondent has in a certain person, group or organization. The paper sets up a hypothesis that states that in this environment a culture of mistrust developed and this culture can still exist today. The article examines this hypothesis by performing several methods and in the end it makes a distinction between two channels that affect the level of trust: effect of factors internal versus external to the individual.

The causality between inherited trust and economic growth:
The article states that inherited trust of descendants of U.S. immigrants consist of two elements. The country were the descendants originally came from and the arrival time of their forebears. The paper of Tabellini (2010) investigates the role of culture by using historical variables as instruments for estimating the effect of trust on income per capita of European regions. According to Algan and Cahuc controlling for regional fixed effects is now not possible. This is due to the fact that a factor that is not varying over time could be correlated with an instrument and the measures of trust. Thus this article aims to avoid reverse causality through focusing on the inherited element of trust and control for omitted time invariant factors. They do this by taking a measure of trust that varies over longer periods of time, which allows them to investigate the influence of inherited trust on worldwide development during the 20th century. Therefore the article investigates the effect of changes in inherited trust on economic growth, which will be called economic performance in this article.

The causality between corporate tax rates and investment and entrepreneurship:
The paper tries to contribute to the previous literature on the effect of corporate taxes in multiple steps (Djankov et al., 2010). Firstly, they take current data for an extensive sample of countries, these selected countries are representative for all the countries. Almost all cross-country research uses a lot of the OECD countries and therefore do not provide a lot information about global development. Secondly, they construct a new dataset in which corporate tax rates are created that are comparable across countries. Thirdly, they combine new data on entrepreneurship with the aim to construct business enrollment data for a great amount of countries. Fourthly, they provide a separated machinery investment measure for manufacturing and services because corporate income taxes may separately influence investment in various sectors. At the end of the paper, because interest payments are tax-deductible they investigate if corporate taxes encourage debt finances instead of equity finances. The paper performs a great deal of robustness checks. When all their controls are included, all the corporate tax rates lose their significance.

3 Methodology

To obtain our final DML methods, this paper removes the regulation bias and over-fitting biases by using two elements provided in the Chernozhukov et al. (2018) article. The regulation bias is removed by using the Neyman-orthogonal restriction:

$$\partial_\eta E_p \psi(R; \theta_0; \eta_0)[\eta - \eta_0] = 0 \tag{1}$$

Where ψ notes the Neyman orthogonal score, R denotes the random elements, θ_0 the parameter of interest and η_0 the nuisance parameters.

The Neyman orthogonal score is obtained through a moment condition, which makes θ_0 locally insensitive towards the value of the nuisance parameters and therefore allows one to plug-in noisy estimates for η_0. In other words it allows the nuisance parameters in DML to be non-parametric. The combination of the Neyman-orthogonal restriction and the moment condition removes the direct confounding effect between the treatment variable and the covariates.
The over-fitting bias is removed by cross-fitting, where the total sample is split between a main sample and auxiliary samples and these roles are swapped to obtain multiple estimates for θ_0. Before θ_0 can be estimated, the nuisance parameters have to be established. For the estimation of the nuisance parameters the sample is split in multiple subsets. Let $I_i \in$ 1,...,K be an integer indicating the k-th fold sample of size n with the total sample size being equal to N. First the nuisance parameters should be estimated with the chosen ML method on all data, excluding the k-th sample. This results in $\hat{\eta}_{0,k}$ for each k-th fold sample. Now there can be a $\check{\theta}_{0,k}$ obtained for each k-th fold sample through the Neyman-orthogonal restriction. In Chernozhukov et al. (2018) a distinction is made between two different DML algorithms; the DML1 and DML2 algorithms. The difference between the two depends on the implementation of k-fold cross-fitting as displayed in their algorithms.
For the DML1 algorithm the construction of $\check{\theta}_{0,k}$ can been seen as the solution of this equation:

$$E_{n,k}[\psi(R; \check{\theta}_{0,k}; \hat{\eta}_{0,k}] = 0 \tag{2}$$

Where the final estimator for $\tilde{\theta}_0$ can be obtained by averaging over all $\check{\theta}_{0,k}$:

$$\tilde{\theta}_0 = \frac{1}{K} \sum_{k=1}^{K} \check{\theta}_{0,k} \tag{3}$$

For the DML2 algorithm the construction of $\tilde{\theta}_0$ can directly be obtained by solving this equation:

$$\frac{1}{K} \sum_{k=1}^{K} E_{n,k}[\psi(R; \tilde{\theta}_0; \hat{\eta}_{0,k}] = 0 \tag{4}$$

In most cases the DML2 algorithm is better behaved according to Chernozhukov et al. (2018). Therefore the DML2 algorithm is applied to the datasets that are investigated.

Choosing a specific sample partitioning in the cross-fitting algorithm has no impact on estimation results asymptoticly but can be of importance in finite samples. Therefore this paper repeats the construction of the DML estimator H times; the estimates $\tilde{\theta}_0^h$, for h=1,...,H are obtained. The final estimator for θ_0 can be realized by using one of these two methods:

$$\tilde{\theta}_0^{mean} = \frac{1}{H}\sum_{h=1}^{H}\tilde{\theta}_0^h \quad or \quad \tilde{\theta}_0^{median} = median\{\tilde{\theta}_0^h\}_{h=1}^H \tag{5}$$

The median approach is a more robust way of using sample splitting as the obtained estimator and corresponding variance are more robust towards outliers. For this reason the median approach is used in the estimation of for θ_0 in this article.

In DML, machine learning methods are used for estimating the nuisance parameter. There is no general form for a machine learning method, for each type of regression another ML method can be selected. This paper highlights the seven ML methods, as explained in the lectures of Imbens (2015), that are used for estimating our regressions (Ridge, LASSO, Reg.Tree, Boosting, Forest, Neutral Net. and Ensemble). The methodology of these methods is displayed in appendix A. After obtaining the estimates of the nuisance function for six of these methods, this paper adds an extra column ('best') in which the best methods are selected by their out-of-sample prediction performance to estimate the different nuisance functions.

3.1 Estimation of the papers

This section of the methodology was used to analyze the papers. It only contains the part of the results that were investigated with the DML method. For the implementation of the OLS results of the articles, this paper will use a partially linear regression (PLR) following from Chernozhukov et al. (2018):

$$Y = D\theta_0 + g_0(X) + U, \quad E[U|X, D] = 0 \tag{6a}$$

$$D = m_0(X) + V, \quad E[V|X] = 0 \tag{6b}$$

Where Y represents the outcome variable, D represents the variable of interest, namely the policy variable, X represents a vector of covariates and U and V are error terms. The nuisance parameter is given by $\eta_0 = (g_0, m_0)$. The variable D is exogenous conditional on the covariates. In the case that D is really exogenous conditional on the covariates it can be interpreted as a treatment effect or causal parameter in many implementations. Equation 6b checks for confounding between the treatment variable and the controls. This equation is of importance because it establishes the affect of the covariates, X, on the treatment variable D through $m_0(X)$ and on the outcome variable through $g_0(X)$. Therefore this equation needs to be established in order to remove the regulation bias.

The causality between slave trade and trust in Africa:

For the paper by Nunn and Wantchekon (2011) their baseline estimating equation is represented by:

$$trust_{i,e,d,c} = \alpha_c + \beta \cdot Slave\_exports_e + X'_{i,e,d,c}\Gamma + X'_{d,c}\Omega + X'_e\Phi + \epsilon_{i,e,d,c} \tag{7}$$

Where i indexes individuals, e ethnic groups, d districts, and c countries. α_c represent the country fixed effects. $Slave\_exports_e$ measures the total of slaves captured from an ethnic group e. The covariates are given by; $X'_{i,e,d,c}$ which are individual level covariates, $X'_{d,c}$ represents variables to capture the ethnic distribution of the district, X'_e, denotes ethnicity-level covariates. $trust_{i,e,d,c}$ represents the five different measures of trust and $\epsilon_{i,e,d,c}$ is an unobserved error term.

The data of several regressions of the paper will be analyzed with the DML methods. The general policy variable is equal to $Slave\_exports_e$, according to the paper this is an exogenous variable. The outcome variable, Y, represents one of the five different measures of trust, $trust_{i,e,d,c}$.

In the first examined OLS result, the policy variable is a transformation of $slave\_exports_e$; the natural log of one plus $slave\_exports_e$ divided by land area, namely ln_export_area. The covariates are given by individual controls and district controls.

In the second OLS result on which the DML methods are applied, there are two variables of interest; a measure of slave exports based on ethnicity, ln_export_area, and a measure of slave exports based on the location, $loc\_ln\_export\_area$, representing the internal and external channel respectively. Therefore this paper takes one of them as policy variable and places the other one in the covariates. After all the estimations are made with DML, this process is performed again with the roles of the variables switched. The covariates are given by canonical population density, ethnicity-level colonial controls and baseline controls.

In addition, the DML methods are applied to their data of the IV results. The PLR model is extended with an instrument variable Z; this is called the partially linear IV model. In equation 6b the policy variable (D) is replaced by Z. In the IV results the instrument is presented by a determination of the distance from the coastline of a persons ethnic group during the slave trade. The policy variable is represented by ln_export_area. The covariates are given by; reliance on fishing distance to Saharan city, route colonial population density, ethnicity-level colonial controls, individual controls district controls and country fixed effects.

The causality between inherited trust and economic growth:

The paper has the following main regression (Algan & Cahuc, 2010):

$$Y_{ct} = \alpha_0 + \alpha_1 \cdot S_{ct} + \alpha_2 \cdot X_{ct} + F_c + F_t + \epsilon_{ct} \tag{8}$$

Where Y_{ct} represents the income per capita in country c at period t, the covariate S_{ct} measure the average of trust of persons in country c at period t. X_{ct} are characteristics that vary over time of a country. F_c represents country fixed effects and F_t represents period fixed effects. ϵ_{ct} denotes a disturbances, namely an unobserved error term. In order to make sure that the disturbance is uncorrelated with the measure of trust, the article provides the following equation:

$$S_{ct} = \gamma_0 + \gamma_1 \cdot S_{ct-1} + \gamma_2 \cdot X_{ct} + \Phi_c + \Phi_t + \nu_{ct} \tag{9}$$

Where S_{ct-1} represents a country's average of trust of the previous generation in the period, t-1. X_{ct} are time varying characteristics of a country. Φ_c represents country fixed effects and Φ_t represents period fixed effects. ν_{ct} denotes a unobserved error term. In equation 9 it is assumed that the present level of trust of a person is influenced by two elements; multiple factors that are expected to determine economic performance and by the level of trust of the forgoing generations

The general policy variable is represented by S_{ct}, which now, after using equation 9, is a level of inherited trust obtained through previous generations. Between the economic outcome at year T, which indicates the start of a period, and inherited trust a lag is specified. The article takes a standard lag of 25 years. This means that they are only interested in beliefs conducted before the time T-25. Additionally the article assumes a lag of 25 years lies between two generations. The article focuses on inherited trust of second generation, third-generations and fourth generation Americans. These generations were born respectively before the date, T - 25, T - 25 + 25 and T - 25 + 50. The article focuses on inherited trust in the periods 1935-1938 and 2000-2003. These two levels of inherited trust correspond to the trust inherited by the three generations.
The paper shows by multiple analyses and by setting up equation 9 that this policy variable can be treated as exogenous.
For the first OLS result that will be investigated with DML, the policy variable will be represented by the country Africa. The output variable changes between two levels of inherited trust. In the second OLS result that is analyzed with DML there are three regressions in which the dependent variable changes between different levels of inherited trust. The policy variable is represented by trust in home country, which stands for the average level of trust in the home country of the U.S. immigrants calculated from the wave 2000. The covariates are defined by multiple individual characteristics.
For their second OLS result analyzed with DML the policy variable is represented by the change in inherited trust between 1935 and 2000. The dependent variable is represented by a measure of economic performance, defined by the shift in income per capita between 1935 and 2000. The covariates are denoted by controls for economic and political factors and country fixed effects. In the last analyzed OLS result additional controls are added to the covariates of the previous OLS regressions. The policy variable and dependent variable stay the same.

The causality between corporate tax rates and investment and entrepreneurship: The paper does not provide a baseline equation but uses simple OLS regressions to estimate all the effects. They investigate the effect of corporate tax rates on investment and entrepreneurship (Djankov et al., 2010). For both the output variables, investment and entrepreneurship, the article provides two measurements; Gross fixed capital formation and foreign direct investment (FDI) for investment and the total of business formations and the rate of current business enrollments as measure for entrepreneurship. This means that the output variable in the partially linear regression changes between four variables. The policy variable is represented by one of the corporate tax rates; statutory, 1st-year effective or 5-year effective tax rate. The covariates are represented by the additional 12 control variables.

4 Data

The causality between slave trade and trust in Africa:
The researchers of this article collected their current data of respondents from the 2005 Afrobarometer surveys (Nunn & Wantchekon, 2011). In these surveys they have a potential sample of 21,822 individuals. From which they remove 120 respondents immediately. The article uses the figures from Nunn (2008) to obtain the estimates of the total of slaves captured from each ethnic group.

The causality between inherited trust and economic growth:
The paper uses the Madison database that captures the period between 1820-2003 to construct measurement of the economic performance (Algan & Cahuc, 2010). The General Social Survey database (GSS), which captures the period 1972-2004, is used for the trust of individuals in the United States. This database provides additional information on birthplace and the source country of the individuals forefathers since 1977 and the ethnic variable. Furthermore the paper contains observations of over more than 24 countries, from which they display only the countries that contain more than 15 observations in their estimations. The trust in home country is a measurement obtained by the World Values Survey (WVS).

The causality between corporate tax rates and investment and entrepreneurship:
The article collected their data from PricewaterhouseCoopers accountants and tax lawyers (Djankov et al., 2010). The researchers asked them to describe a standard business and fill out its tax return, they called this business TaxPayerCo. This paper uses the data that covers the tax system operative in the fiscal year 2004. 85 countries (Djankov et al., 2002) are included in the sample, which are divided over different continents and have different average income levels in countries. They set up two panels, A and B, that respectively describe the balance sheet and the profit and loss presentation.

5 Replication

In order to discuss the results of the three articles that this paper examines, it is crucial to replicate their obtained results. This paper reproduces for both papers the results in STATA. In all the regressions the missing values or omitted variables are left out while estimating the regressions. Furthermore, categorical variables, like education, are encoded into dummies for each category/level. This means that a lot of variables can be created out of one variable, depending on the number of categories, and this results in a bigger dataset. There is one exception in the paper by Nunn and Wantchekon (2011) where one categorical variable is not encoded into dummies. In a couple of cases there are small differences between the replication tables of this paper and the tables displayed in the articles. This is because this paper rounded off to three decimals. For table 14 there are small deviations in the replication tables that are not due to rounding, all the results are still significant which is in accordance with the article. In table 18 in column five the significant level differs from the one in the paper, the magnitude of the coefficient and standard error are equal to the one in the paper. All the tables are displayed in appendix B.

The causality between slave trade and trust in Africa:

For this paper all the regressions use the same baseline controls, in the appendix is shown in or under the tables which additional controls are provided in each regression. Table 8 investigates the relationship between different measures of the slave exports variable and a measure of trust, namely trust of neighbors. The results show that there is a significant negative effect between slave exports measures and the trust the respondents have in their neighbors. In table 9 the relationship between different levels of trust with a transformed measure of slave export is studied. The results display a significant negative relationship between different measures of trust and the measure for slave exports. Table 10 has the same dependent variables and export variable as table 9, only additional control variables are added. The conclusion is similar to table 9.

In table 11 and 12 the results come from executing multiple IV regressions. Both in table 11 and 12 a determination of the distance from the coastline of a persons ethnic group during the slave trade is used as an instrument in the IV regressions. Again the relationship between different measures of trust with a transformed measure of slave export is studied. Only in table 12 more control variables are added. In both tables a significant negative effect between the different measures of trust and the transformed measure of slave exports, $ln\_exports\_area$, is established. The findings of the article indicates that IV estimates are comparable to the OLS estimates, therefore this suggests that selection into the slave trade, due to omitted variables, is not strongly biasing the OLS estimates. Again the causal effect between slave trade and the measures of trust is confirmed.

In the first two columns of table 13 the relationship between slave trade and nowadays trust in the local government council is investigated. Table 13 controls for three measures of anticipated quality of the local government. In the last three columns of table 13, a is distinction made between the internal channel of trust and the level of trust in others. The table investigates the effect of slave trade obtained through a change in one of these two elements. Finally table 14 performs OLS regressions to test for channels of causality. In the table there are two covariates included which represent the internal channel and external channel. In table14 the additional control variables of table 13 are excluded. The results show that both the internal and external channels influence the measures of trust and the internal channel appears to be of more importance.

The causality between inherited trust and economic growth:

In this paper the specific controls in each regression are displayed in the tables in the appendix. The first result presents the estimates between two periods of inherited trust and 23 countries and individual characteristics(Algan and Cahuc, 2010). In table 15 the estimate of Africa is displayed because this was the most important aspect of the table. Africa has a relatively large estimate compared to the other countries. Therefore Africa can be treated as an outlier and will be eliminated in some regressions as robustness check. The effect between inherited trust and the variable Africa is significantly negatively correlated, which gives reason to believe that inherited trust has a negative effect on developing countries. In table 16 the relationship between different levels of inherited trust and trust immigrants have in their country of origin is examined. The results show that the correlation of inherited trust for period 2000 with the trust in home country is statistically significant, while for inherited trust in the period 1935 there is no significant correlation. The results propose that the effect of home country on the level of inherited trust in 1935 differs from the effect on the level of inherited trust in 2000.

In table 17 the relationship between changes in inherited trust and economic performance over time is investigated. The table displays that in each regression there are different controls. The controls that can be added are economic performance between 1870 and 1930, changes in political organizations between 1930 and 2000 and country fixed effects. For each regression, the relationship between the change in inherited trust and economic performance is significantly correlated. Table 18 displays the reaction of differences in inherited trust with economic performance both between 1935 and 2000 with the same controls as in table 17. Additionally there is controlled for multiple inherited social beliefs. In every regression a different social belief control is added. In the last column the result of the regression in which all controls where added is displayed. For each regression the relationship between change in inherited trust and economic performance remained significantly correlated. This confirms that the effect of a change in inherited trust on economic performance is economically sizable.

The causality between corporate tax rates and investment and entrepreneurship: In the article by Djankov et al. (2010) the relationship between corporate taxes and investment and entrepreneurship is analyzed. Table 19 displays the effect on the two different measures of investment. In table 20 the effect on the two different measures of entrepreneurship is shown. Both these tables have no additional controls. A extensive statistical effect is shown in the results of table 19 of both the effective rates on the dependent variables investment and FDI. Statutory tax rate has a large effect on FDI but there is no effect on investment that is statistical significant. The estimated effects of taxes on the output variables of entrepreneurship are all large and statistically significant.

In table 21 and 22 there are 12 control variables added. After this, the estimates of the corporate income tax variables lose all their significance and their magnitudes drop a lot. Therefore the causal effects established in table 19 and 20 cannot be confirmed.

6 Results

In this section the results of applying DML to the dataset of the paper are presented. In these results the estimates of the ML method, 'best', are reported. The ML methods which are used in this paper are Trees, Forest, Nnet, Boosting, Rlasso and Ensemble. The ML methods that were included in estimation of a coefficient is stated in the most left column of each table. In brackets, next to the used ML methods is stated respectively how many splits and k-fold were used in the estimation. In the case of an estimation of small dataset, this paper used the method Boosting with 5-fold instead of 2-fold to obtain more accurate results. The reason for this is explained in appendix A. For the paper by Nunn and Wantchekon (2011), the DML methods were applied to the dataset of the regressions displayed in table 10, 12 and 14. The DML methods were implemented on the data of the regressions of tables 15, 16, 17 and 18 of the article by Algan and Cahuc (2010). Finally on the data of the table in which the paper by Djankov et al. (2010) adds 12 extra control variables, 21 and 22, were the DML methods implemented. The paper by Nunn and Wantchekon (2011) contains a big dataset. Therefor the DML methods were applied with the use of only two splits and the ML method Trees executed separately with 100 splits as robustness check. Most of the results of the tables of the article by Algan and Cahuc (2010) did not differ much from the original results of the paper. Therefore these results are of less interest to this paper and are displayed in appendix C. The t-test methodology, which is used to test for significance of the estimates, is displayed in appendix D.

The causality between slave trade and trust in Africa:
In the tables 1 up to and including 4 the results of applying the DML methods to the data of the regressions of the paper by Nunn and Wantchekon (2011) are displayed. In table 1 it is clear that the results obtained with the DML methods differ from the results in table 10. The three coefficients in the most right columns are positive while in table 10 these are negative. The absolute magnitude of the coefficients falls a lot between a half or one tenth and for the first two estimation methods they lose all significance, which is in contrast with table 10. Furthermore the standard error is higher relatively to table 10, and even increases when more methods are added. When looking at the result of Trees(100,2) the standard errors are similar to the ones of the article, but the coefficients deviate even more. Therefore this paper does not confirm the negative causal effect of slave trade on the level of trust that has been established in the article, it even shows a positive effect for some measures of trust.

Table 1: Results of applying the DML method on the data of table 3 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
Trees,Forest,Booting(2,2)	-0.015	-0.108	0.057	0.089	0.168
	(0.099)	(0.145)	(0.110)	(0.175)	(0.207)
Trees,Forest,Boosting,					
Ensemble(2,2)	-0.032	-0.204	-0.026	0.100	0.195
	(0.101)	(0.276)	(0.142)	(0.187)	(0.183)
Trees(100,2)	0.143***	0.118***	0.066*	-0.020	0.231***
	(0.027)	(0.031)	(0.034)	(0.030)	(0.031)
number of obs.	16709	16679	15905	16636	16473

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: DML, five different measure of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). The used policy variable is a transformation of *slave_exports*, namely *ln_export_area*. All covariates and controls used in the article are included.

For table 2 the displayed results deviate a lot compared to the results of table 10. For the applied DML methods, with the ML methods Trees, Forest, Boosting, only the estimates for the third regression looks similar to the one of table 10. All the standard errors have four to ten times the magnitude compared to the ones in table 12. The robustness check displays that the standard errors fall in their magnitude but are still bigger than the ones in table 12. The estimates of all the coefficients remain deviating from table 12 and even the third coefficient. All the estimates lose their significance in table 2. When comparing the results of table 1 and 2 it is clear these estimates differ, while in the paper they concluded that there was no difference between the OLS and IV results. Again it seems that there is no general causal effect between slave trade and the levels of trust.

Table 2: Results of applying the DML method on the data of table 6 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
Trees,Forest,Booting(2,2)	0.257	1.976	-0.222	-3.241	-0.659
	(1.733)	(1.828)	(1.415)	(1.587)	(0.495)
Trees(100,2)	0.019	0.751	0.204	-0.142	0.149
	(0.242)	(0.287)	(0.312)	(0.228)	(0.287)
number of obs.	16709	16679	1509	16636	16473

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: DML, five different measures of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). The used policy variable is a transformation of *slave_exports*, namely *ln_export_area*. All covariates and controls used in the article are included.

For table 3 and 4 the estimates of the applied DML methods are almost all negative, except for estimates in the third and fifth column of table 3, which are similar to the results of table 14. In table 3 the standard errors obtained with the DML methods are larger than the standard errors of the article. The robustness check shows that the coefficients deviate even more. An interesting observation is that for the first and second result of table 3 the estimates looks similar to the ones of table 14 only are they positive instead of negative. For table 4 the estimates differ from the ones in table 14 but have the same sign. The standard errors are quite similar to the ones in table 14, this remains after performing the robustness check. Comparing table 3 with 4 it seems that the magnitude of the effect of the internal towards the external channel is of the same magnitude as in the article, only the effect of the internal channel seems to be positive.

The differences between the results of this paper and the results of the paper by Nunn and Wantchekon (2011) can be explained by the fact that DML are semi-parametric methods. DML captures nonlinearities and interactions between the covariates and the outcomes. The standard methods can not capture these interactions because they make the assumption of well established properties in the model. Especially, in case of very big datasets as in this paper, it can be expected that these properties do not hold.

Table 3: Results of applying the DML method on the data of table 10 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
Trees,Forest,Booting(2,2)	-0.188	-0.172	0.141	-0.227	0.186
	(0.148)	(0.167)	(0.092)	(0.222)	(0.154)
Trees(100,2)	0.154***	0.166***	0.003	0.083	0.231***
	(0.027)	(0.029)	(0.032)	(0.029)	(0.031)
number of obs.	15999	15972	15221	15931	14773

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: DML, five different measures of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). The used policy variable is a transformation of *slave_exports*, namely *ln_export_area*. All covariates and controls used in the article are included.

Table 4: Results of applying the DML method on the data of table 10 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
Trees,Forest,Booting(2,2)	-0.016	-0.018	-0.039**	-0.024	-0.024
	(0.018)	(0.017)	(0.018)	(0.017)	(0.017)
Trees(100,2)	-0.041***	-0.019	-0.019	-0.037***	-0.034***
	(0.014)	(0.014)	(0.015)	(0.013)	(0.013)
number of obs.	15999	15972	15221	15931	14773

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: DML, five different measures of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). The used policy variable is a transformation of *slave_exports*, namely *loc_ln_export_area*. All covariates and controls used in the article are included.

The causality between inherited trust and economic growth:
The results of applying the DML to the datasets of the tables of the article by Algan and Cahuc (2010), show that the estimates and standard errors do not deviate much from the article. Only in table 5 the estimate of home country on inherited trust in 1935 is significant, which is in contradiction with the paper. Therefore the conclusion made in table 16, namely that the effect of the variable home country on the level of inherited trust in 2000 and 1935 differs, can be held unreliable. The variable home country seems to have a similar effect on both levels of inherited trust. In the last three tables the standard errors are almost half the size when the DML methods were applied compared to the standard errors of the article. When looking at the estimates of the tables 5, 24 and 25 they are all lower in magnitude than in, respectively, tables 16, 17 and 18. Especially after performing a robustness check by using five fold instead of two.

Table 5: Results of applying the DML method on the data of table 3 of the causality between inherited trust and economic growth

	Inher. trust:		
	in 2000	in 1935	in 2000 4th generation
All methods(20,2)	0.448***	0.416***	0.448***
	(0.060)	(0.050)	(0.083)
All methods(20,5)	0.441***	0.415***	0.465***
	(0.057)	(0.051)	(0.083)
All methods(100,2)	0.440***	0.415***	0.477***
	(0.057)	(0.050)	(0.082)
number of obs.	4491	6535	2095

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: DML. The dependent variable is represented by a level of inherited trust of US immigrants. Additionally controls are multiple countries, age, age(squared), religion, gender, education, income and employment status. This table combines table 1 of the paper were they use dummies for each period with taking both periods separately.

The causality between corporate tax rates and investment and entrepreneurship: For the paper by Djankov et al.(2010) the results of applying the DML method are shown in table 6 and 7. All the estimates obtained by performing the DML method shown in table 6 are higher than the estimates obtained by performing OLS shown in table 21. In the second, third, fourth and fifth column the estimates are now even statistically significant. This is in contradiction with the paper where all their estimates lose their significance. The estimates of table 7 in the first, second, fourth and fifth column are substantially different from the OLS estimates in table 20. When performing OLS all estimated coefficients become of no significance, therefore the article can not confirm a causal effect between corporate tax rates and investment and entrepreneurship. This is not the case in the results where a negative effect between the corporate tax rates and investment and entrepreneurship for 8 out the 12 regressions is established. The standard errors in both tables, in which the DML methods were applied, are similar or smaller than the standard errors of the tables in which OLS was used to estimate the coefficients.

Table 6: Result of applying the DML method to the data of table 5D of the article about the causality between corporate tax rates and investment and entrepreneurship

All methods(100,2),	Investment			Foreign direct investment		
Tax rates:	(1)	(2)	(3)	(4)	(5)	(6)
Statutory corporate	0.029			-0.135**		
	(0.100)			(0.057)		
1st-year effective		-0.151*			-0.152***	
		(0.086)			(0.050)	
5-year effective			-0.265***			-0.084
			(0.084)			(0.053)
number of obs.	61	61	61	61	61	61

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: DML, for the estimation of regression 1-6 three different policy variables, displayed in the first column, were used. The variables Investment and FDI are both obtained from the period 2003-2005. 12 additional controls were included.

Table 7: Result of applying the DML method to the data of table 5D of the article about the causality between corporate tax rates and investment and entrepreneurship

All methods(100,2)	Business density			Entry rate(aver.)		
Tax rates:	(1)	(2)	(3)	(4)	(5)	(6)
Statutory corporate	-0.084*			-0.114*		
	(0.048)			(0.062)		
1st-year effective		-0.173**			-0.182***	
		(0.065)			(0.057)	
5-year effective			-0.092			-0.110
			(0.062)			(0.079)
number of obs.	60	60	60	50	50	50

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: DML, for the estimation of regression 1-6 three different policy variables, displayed in the first column, were used. The variable average entry rate is obtained from the period 2000-2005 12 additional controls were included.

7 Conclusion

The findings of applying the DML methods to the dataset of the paper by Nunn and Wantchekon (2011), show that the results presented by the paper should be questioned. In this paper, it is concluded that there is no negative causal effect between different measures of the total of slaves captured from an ethnic group and different measures of trust. After performing a robustness check the effect seems to be positive for some measures of trust. The assumption made in the paper about their executed IV regressions is also rejected. The DML results show that there is no reason to assume the IV and OLS estimates are compatible. With this stated, the robustness check of the paper by means of performing an IV, in order to investigate if selection into slave trade is biasing the OLS estimates, can not be held reliable. Therefore this paper does not establish a causal negative effect between slave trade and the different measures of trust. Finally the results obtained by DML do not confirm the conclusions of the paper about effects of the internal and external channel. According to the DML findings there is no causal relationship established between the two channels and the levels of trust. The results point out that the internal channel estimates are sufficiently larger than those of the external channel, which is in line with the paper.
In conclusion, almost all drawn conclusions of the paper are unreliable. The paper works with a great amount of observations and lots of covariates and it could be very beneficial for the content of the article to use the DML methods to review their results.

For the paper by Algan and Cahuc (2010), most of their OLS results can be held reliable. Even after performing a robustness check by using more fold, this paper can not question most of their results. Only the assumption the article makes about difference between the effect of home country on the level of inherited trust in 2000 and 1935 can be questioned. Furthermore in all the DML results, in case of a small sample size, the standard errors obtained with DML are lower. This shows the power of the prediction performance of the DML methods. Therefore using the DML methods as estimation methods could be a very valuable addition to the content of the paper.

The results of implementing the DML methods to the paper by Djankov et al.(2010) are very interesting. The article stated that after taking all the 12 control variables in consideration the effect of all the corporate tax rates on the investment and entrepreneurship is lost. This in contrast with the result of this paper, for 8 out of the 12 regressions the estimates obtained with DML are significant. This demonstrates the existence of a causal relationship even with the 12 additional controls. The estimates look similar to the ones in the paper when no control variables were added. This is due to the fact that DML shrinks the estimates of non-important covariates towards zero. This causes the important estimates, in this case the corporate tax rates, to remain significant. The overall conclusion of the article is not rejected. However the article loses credibility on their conclusions about the effect of corporate taxes by performing their robustness checks while this was not necessary. It would be a great addition to the paper to implement the DML methods on their data in order to show the significant causality between corporate tax rates and investment and entrepreneurship with multiple controls.

The results of the three papers show the relevance of using DML as estimation methods in research that uses new constructed datasets. The reliability of conclusions of already published papers can also be checked by implementing the DML methods on the data. The three articles that are analyzed by this paper have different contents, which indicates how DML implementation can be of importance to a wide field of studies. The DML methods are very 'new' methods and could be very useful complementary methods for investigating causal inferences between parameters. Especially because DML are semi-parametric methods, which allow the nuisance parameter to be non parametric. Therefore it can capture nonlinearities and other interactions between variables, which is not possible when standard methods are used. This paper wants to emphasize the great benefits that can be obtained by applying the DML methods. Estimating and establishing valid relationships between the treatment variable and the dependent variable is one of its strengths. This is especially significant in more complex situations such as when there are a lot of covariates relatively to the number of observations or in the situation of big datasets.

8 Discussion

Based on the dataset of the article by Nunn and Wantchekon (2011) only two sample splits were used in the DML estimation of the parameters. The computational time of one estimation was very long due to the large dataset. Therefore increasing the number of splits was not possible due to time limitation. However the robustness check still showed that the results of the paper can be seen as unreliable. It would be very interesting to continue the research on DML performance in combination with the big data of this paper in order to draw more solid conclusions. In this paper a couple of ML methods were not used in the DML algorithm due to time limitations. Additionally in table 14 two policy variables can be found, which are now estimated separately. It would be an interesting addition to write an algorithm for DML that could work with more than one policy variable.
The ML method Neural Net. was excluded from some estimations because it caused the DML algorithm to default. The goal of a learning algorithm is to find the function that lies underneath the data. For a neural network is it important to determine all the patterns in the data from the beginning. If the method learns a new pattern further into the training, it fails to learn the right function and therefore is unable to set its weights in the right way (Elman, 1993). The Neural Net. method is highly sensitive towards the input data and for a small sample size Neural Net. is unable to determine all the patterns. For the first two papers, in which big data sets were used, Neural Net. also caused a default in the DML method. In some data there are multiple dummies created out of one category variable. When a couple of dummies are removed, the method Neural Net. works. This paper believes that due to multiple dummies the Neural Net. algorithm gets stuck in a local minimum. This paper chose to use the exact amount of variables as used in the papers and eliminate the method Neural Net. in the cases it caused DML to default. For the second paper there was only one result of interest to be found. In some estimations there were a lot of country dummies but less other 'raw' covariates, which makes OLS a more reliable method to use.

References

- Algan, Y., & Cahuc, P. (2010). *Inherited trust and growth.* American Economic Review, 100(5), 2060-92.
- Athey, S. (2017). *The Impact of Machine Learning on Economics.* In Economics of Artificial Intelligence. University of Chicago Press.pr
- Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). *Sparse models and methods for optimal instruments with an application to eminent domain.* Econometrica, 80(6), 2369-2429.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W.,& Robins, J. (2018). *Double/debiased machine learning for treatment and structural parameters.* The Econometrics Journal, 21(1), C1-C68.
- Djankov, S., Ganser, T., McLiesh, C., Ramalho, R., & Shleifer, A. (2010). *The effect of corporate taxes on investment and entrepreneurship.* American Economic Journal: Macroeconomics, 2(3), 31-64.
- Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2002). The regulation of entry. The quarterly Journal of economics, 117(1), 1-37.
- Elman, J. L. (1993). *Learning and development in neural networks: The importance of starting small.* Cognition, 48(1), 71-99.
- Friedman, J. H. (2002). *Stochastic gradient boosting.* Computational Statistics & Data Analysis, 38(4), 367-378.
- Guiso, L., Sapienza, P., & Zingales, L. (2004). *The role of social capital in financial development.* American economic review, 94(3), 526-556.
- Ho, J. (2016). *Machine Learning for Causal Inference: An Application to Air Quality Impacts on House Prices.*
- Imbens G.(2015, july 18).*Introduction to Supervised ML Concepts and Algorithms.*Retrieved from: `http://www.nber.org/econometrics_minicourse_2015/`
- Imbens G.(2015, july18).*Prediction, Classification, and Applications.*Retrieved from: `http://www.nber.org/econometrics_minicourse_2015/`
- Kononenko, I. (2001). *Machine learning for medical diagnosis: history, state of the art and perspective.* Artificial Intelligence in medicine, 23(1), 89-109.
- Nunn, N. (2008). *The long-term effects of Africa's slave trades.* The Quarterly Journal of Economics, 123(1), 139-176.
- Nunn, N., & Wantchekon, L. (2011). *The slave trade and the origins of mistrust in Africa.* American Economic Review, 101(7), 3221-52.

- Rosenbaum, P. R. (2002). *Observational studies.* In Observational studies (pp. 1-17). Springer, New York, NY.
- Tabellini, G. (2010). *Culture and institutions: economic development in the regions of Europe. Journal of the European Economic association*, 8(4), 677-716.
- Widrow, B., Greenblatt, A., Kim, Y., & Park, D. (2013). *The No-Prop algorithm: A new learning algorithm for multilayer neural networks.* Neural Networks, 37, 182-188.

9 Appendix

9.1 Appendix A

Machine learning methods:

In the case that a penalty parameter is chosen through cross-validation. The model should be estimated with ML for multiple subsample, i =1,...,k, on all data excluding the k-th cross-validation with all possible values of λ, call this $g(b, \lambda)$. For each λ the squared errors can be added up over the cross-validation sample to get:

$$Q(\lambda) = \sum_{k=1}^{K} \sum_{i:I_i=k} (Y_i - g(b, \lambda))^2 \tag{10}$$

The λ that minimizes equation 10, represents the penalty term. After the penalty term is selected, this can be incorporated in the given ML method over the full sample.

Ridge:

The ridge method estimates the β of a regression model in the following way:

$$\hat{\beta}_{ridge} = (X'X + \lambda \cdot I_k)^{-1}(X'Y) \tag{11}$$

Where λ is added as a penalty term in the denominator. Due to the added penalty term the matrix can be inverted and the estimates are shrank towards zero. When all the covariates are uncorrelated the coefficients of the regression model are shrank by a factor of $1/(1+\lambda)$.

Least absolute selection and shrinkage operator (RLasso):

$$\min_{\beta} \sum_{i=1}^{N} (Y_i - X_i\beta)^2 + \lambda\|\beta\|_1 \tag{12}$$

with the L_1 norm being equal to:

$$\min\|\beta\|_p = \sum_{k=1}^{K} (|\beta|_p)^{(\frac{1}{p})} \tag{13}$$

The interpretation of these equations is that the LASSO minimizes the Least Squares but adds a penalty term with the form of L_1 to it. The penalty term is selected through cross-validation. This means that LASSO shrinks the OLS estimates that are not important to zero. Lasso doesn't shrink the important estimates too much, therefor the effects of the important covariates can be discussed. In the paper by Chernozhukov et al.(2018) they select the L_1 penalty term based on a selection rule stated in Belloni et al. (2012). This paper also incorporates this penalty term.

Regression trees(Trees):

The main idea is to divide the covariate space into multiple subspaces. The partitioning is sequential, which means that one covariate is split at the time. This means that for covariate k and for threshold t, the data can be split depending on whether: $X_{i,k} \leq$ t versus $X_{i,k} >$ t. Over these two samples the average over the outcomes can be obtained; $\overline{Y}_{left}$ and $\overline{Y}_{right}$ respectively. For these outcomes the estimation of the sum of squared deviations can be made:

$$Q(g_{k,t}) = \sum_{i=1}^{N} (Y_i - g_{k,t})^2 = \sum_{i=1}^{N} (Y_i - \overline{Y})^2 \tag{14}$$

Where $g_{k,t}$ defines a function that gives the outcome as $\overline{Y}_{left}$ or $\overline{Y}_{right}$, depending on the chosen covariate and threshold. Finally the covariate k^* and the threshold t^* can be found that solve:

$$(k^*, t^*) = \operatorname*{argmin}_{k,t} Q(g_{k,t}(.)) \tag{15}$$

The algorithm keeps splitting the covariates if the sum of deviations is reduced. This leads to the biggest improvement in the objective function. A penalty term is added through cross-validation for the number of leaves.

Boosting:

This method makes use of 'weak learners' to get a good predictor for the regression function. Weak learners are simple ways to estimate a regression function. Weak learnings can be obtained from multiple methods (trees, kernels etc.). At first a residual relative to the first weak learning ϵ_{1i} should be defined, for example obtained from the first split in the regression tree. This residual can be applied to a new dataset (X$|\epsilon_{1i}$). From the new dataset a second residual (weak learner) can be defined. Again this residual can be applied to a new dataset (X$|\epsilon_{2i}$). After proceeding this many times, an additive approximation of the regression function is obtained. The complexity of the algorithm can be controlled by changing the weak learner. In this paper we select the 'weak learner' based on a subsamples of the dataset and this 'weak leaner' is used to construct the new model. The more k-fold subsamples are taking the more randomness in the process of boosting. For samples of a smaller size it is very beneficial to consider taking more k-fold subsamples (Friedman, 2002).

Random Forest(Forest):

The key feature of the random forest is that this method does not look at all the regressors to decide on which covariate (with threshold) it splits, like the regression tree. The method selects randomly a set of regressors, named the bootstrap sample. In this set of regressors the algorithm searches for the optimal covariate and threshold to split on. If some leaves have more observations than the set minimum, N_{min}, the algorithm keeps taking a new random bootstrap sample. Finally the trees are averaged over the bootstrap samples.

Normal trees are very discreet, either two observations are in the same leave or not. The random forest method results by averaging over these bootstrap samples in a more smooth prediction function. This leads that for random forest, when choosing a relatively small bootstrap sample to the overall sample, asymptotic normality may hold. Thus for a particular value of x there can be a prediction obtained with the associated interval.

Neural networks(Nnet.):

This methods models the relationship between X_i and Y_i through hidden layer(s) of Z_i, with Z_i having M elements. The layers have the function of, $Z_{i,m}$=$\sigma(\alpha_{0m} + \alpha_{1m}^{'}X_i)$, for m =1,....,M being the number of nodes in a layer. The relation between the layers and the outcomes are given by:

$$Y_i = \beta_0 + \beta_1' Z_i + \epsilon_i \tag{16}$$

Equation 16 shows that the Y_i are linear in a number of transformations of the original covariates. The parameters α_m and β are fitted by minimizing: $\sum_{i=1}^{N}(Y_i - g(X_i, \alpha, \beta))^2$. This can be seen as minimizing the output response of the network and the desired response (Widrow et al, 2013). In each layer weights can be attached to each node, the Neural Net. algorithm changes the weights iteratively towards the chosen decay parameter to find a minimum. This method is very flexible and allows that there can be very complicated structure between the covariates. At the other hand can the estimation be hard. Especially in small sample sizes Neural net. performs poor (Chernozhukov et al., 2018).

Ensemble methods:

The main idea is to combine estimators obtained by different (previous) ML methods to establish a better estimator. The method takes the average of all these predictors to get a better estimator. Therefor the method tries to select a set of weights that are non-zero for multiple methods. If the algorithm has many methods to choose form, it can be beneficial to regularize this issue by adding a LASSO-type penalty term. This penalty term shrinks the weights towards zero, keeping non-zero weights on only a few models.

9.2 Appendix B

In this appendix the replication of the tables of the articles are displayed. The significance of the estimates is determent by the p-value in STATA.

Table 8: Replication of table 1 of the article about the causality between slave trade and trust in Africa

	Slave exp.	Exp./ area	Exp./ historical pop.	ln(1+ exp.)	ln(1+ exp./ area)	ln(1+ exp. / historical pop)
Exports(exp.)	-0.001	-0.019	-0.531	-0.037	-0.159	-0.743
	(0.000)	(0.005)	(0.0.147)	(0.014)	(0.034)	(0.187)
number of obs.	20027	20027	17644	20027	20027	17644
R^2	0.16	0.16	0.15	0.15	0.16	0.15

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS. Additional controls are: baseline and district controls and country fixed effects.

Table 9: Replication of table 2 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
ln_export_area	-0.133***	-0.159***	-0.111***	-0.144***	-0.097***
	(0,365)	(0,343)	(0,021)	(0,032)	(0,028)
number of obs.	20062	20027	19733	19952	19765
R^2	0.13	0.16	0.20	0.14	0.11

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS, five different measures of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). Additional controls are: baseline and district controls and country fixed effects.

Table 10: Replication of table 3 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
ln_export_area	-0.178***	-0.202***	-0.128***	-0.188***	-0.115***
	(0.032)	(0.031)	(0.021)	(-0.032)	(0.030)
number of obs.	16709	16679	15905	16636	16473
R^2	0.13	0.16	0.21	0.16	0.12

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS, five different measures of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). Additional controls are: individual, ethnicity measure colonial and district controls, country fixed effects and colonial population density.

Table 11: Replication of table 5 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
2nd stage: Depend. var. is a respondents trust					
ln_export_area	-0.190***	-0.245***	-0.221***	-0.250***	-0.174**
	(0.067)	(0.071)	(0.060)	(0.088)	(0.081)
R^2	0.13	0.16	0.2	0.15	0.12
1st stage: Depend. var. is ln_export_area					
Instrument:	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
number of obs.	16709	16679	15905	16636	16473
R^2	0.81	0.81	0.81	0.81	0.81

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: IV, five different measures of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). Additional controls are: individual, ethnicity measure colonial and district controls, country fixed effects and colonial population density. A determination of the distance from the coastline of a persons ethnic group during the slave trade is used as an instrument.

Table 12: Replication of table 6 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
2nd stage: Depend. var. is a respondents trust:					
ln_export_area	-0.172**	-0.272***	-0.262***	-0.254**	-0.189*
	(0.076)	(0.088)	(0.075)	(0.109)	(0.103)
R^2	0.13	0.16	0.20	0.15	0.12
1st stage: Depend. var. is ln_export_area:	-.002***	0.002***	-0.001***	-0.002***	-0.002***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Instrument:	yes	yes	yes	yes	yes
Reliance on fishing	yes	yes	yes	yes	yes
Distances to Saharan	yes	yes	yes	yes	yes
number of obs.	16709	16679	1509	16636	16473
R^2	0.81	0.81	0.81	0.81	0.81

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: IV, five different measures of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). Additional controls are: individual, ethnicity measure colonial and district controls, country fixed effects and colonial population density. A determination of the distance from the coastline of a persons ethnic group during the slave trade is used as an instrument.

Table 13: Replication of table 9 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	local council		intergroup (1)	intergroup (2)	intergroup (3)
slave_exports based on ethnicity	-0.072***	-0.070***	-0.102***	-0.120***	-0.098***
	(0.019)	(0.019)	(0.028)	(0.027)	(0.029)
slave_exports measure (aver.) surrounded by other ethnicities in the same place:			-0.037	-0.063**	-0.091***
			(0.027)	(0.030)	(0.040)
number of obs.	12827	12203	9673	12513	15999
R^2	0.37	0,37	0.12	0.12	0.12

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS. Types of intergroup trust: with town(1), within district(2) and within province(3). Additional controls are: baseline, ethnicity measure colonial controls, country fixed effects, colonial population density, council trustworthiness fixed effect and five public goods fixed effects.

Table 14: Replication of table 10 of the article about the causality between slave trade and trust in Africa

	Measures of trust:				
	(1)	(2)	(3)	(4)	(5)
slave_exports based on location	-0.058***	-0.041**	-0.068***	-0.039*	-0.047**
	(0.020)	(0.019)	(0.017)	(0.022)	(0.024)
slave_exports based on ethnicity	-0.155***	-0.182***	-0.100***	-0.169***	-0.090***
	(0.029)	(0.029)	(0.023)	(0.033)	(0.030)
number of obs.	15999	15972	15221	15931	14773
R^2	0.13	0.16	0.21	0.15	0.12

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS, five different measures of trust are used as dependent variable: trust of relatives(1), neighbors(2), local council(3),intra-group trust(4) and intergroup trust(5). Additional controls are: baseline, ethnicity measure colonial controls, country fixed effects and colonial population density.

Table 15: Replication of table 1 of the article about the causality between inherited trust and economic growth

	Inher. trust in 1935		Inher. trust in 2000	
	Coeff.	Stand. error	Coeff.	Stand. error
Africa, together	-0.231***	(0.004)	-0.243***	(0.007)
Other countries	yes		yes	
number of observations:			11026	
R^2			0.1005	
Africa, separately	-0.229	(0.003)	-0.292	(0.007)
Other countries	yes		yes	
number of obs.	6535		4491	
R^2	0.103		0.091	

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS. The dependent variable is represented by a level of inherited trust of US immigrants. Additionally controls are multiple countries, age, age(squared), religion, gender, education, income and employment status. Table 1 of the paper uses dummies for each period (Africa, together), in this table the effect of the covariates on the periods is estimated separately (Africa, separately).

Table 16: Replication of table 3 of the article about the causality between inherited trust and economic growth

	Inher. trust in 2000	Inher. trust. in 1935	Inher. trust in 2000 4th generation
trustwvs2000	0.462***	0.419	0.461**
	(0.142)	(0.268)	(0.211)
gender(men)	0.029***	0.018	0.022
	(0.012)	(0.018)	(0.016)
age	0.005***	0.003***	-0.003***
	(0.000)	(0.000)	(0.003)
ageedu	0.037***	0.035***	0.044
	(0.004)	(0.003)	(0.007)
incomegood	0.010***	0.016***	0.011***
	(0.003)	(0.002)	(0.003)
unemployed	-0.077**	-0.200	-0.05
	(0.037)	(0.039)	(0.044)
catho	-0.003	0.064***	-0.031
	(0.028)	(0.031)	(0.033)
employed	-0.006	0.015	-0.031
	(0.025)	(0.024)	(0.030)
pro	-0.006	0.002	-0.016
	(0.021)	(0.026)	(0.029)
number of obs.	4491	6535	2095
R^2	0.077	0.08	0.07

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS. The dependent variable is represented by the level of inherited trust of US immigrant. Ageedu stands for education. Additionally controls are multiple countries.

Table 17: Replication of table 6 of the article about the causality between inherited trust and economic growth

	Income per capita:				
Inher. trust	41,007.70***	27,332.62***	31,198.48***	28,280.15***	23,930.95***
	(6041.575)	(7179.626)	(7231.026)	(7350.495)	(6181.2)
Initial income pc (1870-1930)		2.939***	2.177*	2.815**	2.658***
		(1.038)	(1.142)	(1.028)	(0.864)
Political organizations (1930-2000)				-149.348	-103.156
				(89.412)	(75.19)
country fixed effects	yes	yes	yes	yes	yes
number of obs.	48	48	46	46	46
R^2	0.84	0.88	0.88	0.87	0.88

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS. The dependent variable is represented by the economic performance in 1935 and 2000. The change in inherited trust is also from the period 1935-2000. Additionally controls are multiple countries.

Table 18: Replication of table 11 of the article about the causality between inherited trust and economic growth

	Income per capita :				
Inher. trust	24,139.51***	32,133.31***	29,875.68	28,397.90***	31,429.61**
	(7586.259)	(10949.69)	(8423.307)	(7196.214)	(10778.15)
Political organizations (1930-2000)	-173.623*	-134.571**	-120.212*	-67.851	-27.050
	(87.944)	(95.97)	(113.746)	(105.481)	(121.794)
Initial income pc (1870-1930)	2.763**	2.693**	2.937**	2.337**	2.169*
	(0.995)	(1.076)	(1.086)	(1.066)	(1.096)
Inher. beliefs against work and sources of success	15,557.01				19,921.46*
	(10012.16)				(8402.547)
Inher. confidence in business		-6,472.757			-10,179
		(13249.42)			(12657.09)
Inher. beliefs against government intervention			-1,086.057		-1,391.677
			(2533.698)		(2375.704)
Inher. beliefs against gender division of work				11,860.37	13,506.7
				(8567.715)	(8402.547)
country fixed effects	yes	yes	yes	yes	yes
number of obs.	46	46	46	46	46
R^2	0.88	0.87	0.87	0.88	0.91

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS. The dependent variable is represented by the economic performance in 1935 and 2000. The change in inherited trust is also from the period 1935-2000. Additionally controls are multiple countries.

Table 19: Replication of table 5 of the article about the causality between corporate tax rates and investment and entrepreneurship:

Tax rates:	Investment			Foreign direct investment		
Statutory corporate	-0.072			-0.195***		
	(0.076)			(0.046)		
1st-year effective		-0.217***			-0.226***	
		(0.074)			(0.045	
5-year effective			-0.249***			-0.223***
			(0.080)			(0.050)
number of obs.	85	85	85	84	84	84
R^2	0.01	0.09	0.01	0.18	0.23	0.20

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. The variables investment and FDI are both obtained from the period 2003-2005. Estimation method: OLS.

Table 20: Replication of table 5 of the article about the causality between corporate tax rates and investment and entrepreneurship:

Tax rates:	Business density			Entry rate(aver.)		
Statutory corporate	-0.153**			-0.127**		
	(0.063)			(0.060)		
1st-year effective		-0.193**			-0.137**	
		(0.062)			(0.057)	
5-year effective			-0.200***			-0.136**
			(0.068)			(0.061)
number of obs.	80	80	80	62	62	62
R^2	0,07	0.11	0.10	0.07	0.09	0.08

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. The variable average entry rate is obtained from the period 2000-2005. Estimation method: OLS.

Table 21: Replication of table 5D of the article about the causality between corporate tax rates and investment and entrepreneurship:

Tax rates:	Investment			Foreign direct investment		
Statutory corporate	-0.064			-0.030		
	(0.098)			(0.066)		
1st-year effective		-0.117			-0.100	
		(0.106))			(0.071)	
5-year effective			-0.189			-0.095
			(0.118)			(0.081)
Additional controls:						
Alternative taxes	-0.527	-0.450	-0.396	-0.065	0.003	0.001
	(0.1881)	(0.200)	(0.203)	(0.127)	(0.134)	(0.138)
PIT top MR	0.139	0.144	0.144	-0.047	-0.038	-0.044
	(0.05)	(0.049)	(0.047)	(0.034)	(0.033)	(0.032)
VAT and sales taxes	0.038	0.018	-0.002	0.003	-0.013	-0.016
	(0.065)	(0.067)	(0.068)	(0.044)	(0.045)	(0.047)
log(total of payments)	-0.024	0.149	0.185	-0.185	0.020)	-0.074
	(0.987)	(0.998)	(0.968)	(0.667)	(0.667)	(0.660)
log gross domestic product per capita 2003	-0.052	-0.258	-0.221	0.201	0.068	0.121
	(0.900)	(0.880)	(0.865)	(0.608)	(0.588)	(0.590)
Severity of employment	-0.083	-0.081	-0.091	-0.036	-0.037	-0.040
	(0.040)	(0.039)	(0.04)	(0.027)	(0.026)	(0.027)
Property right index	-0.124	-0.110	-0.101	-0.084	-0.068	-0.071
	(0.064)	(0.066)	(0.065)	(0.043)	(0.044)	(0.044)
Procedures to begin a business	-0.004	0.028	0.043	0.040	0.083	0.065
	(0.222)	(0.222)	(0.216)	(0.150)	(0.148)	(0.148)
EWF	0.767	0.723	0.397	2.218	2.064	2.019
	(0.976)	(0.916)	(0.947)	(0.659)	(0.612)	(0.645)
Seign. 2004	0.384	0.336	0.319	0.036	-0.007	0.003
	(0.158)	(0.164)	(0.160)	(0.106)	(0.109)	(0.109)
Tax evasion	0.301	0.297	0.299	1.040	1.002	1.036
	(0.798)	(0.785)	(0.772)	(0.539)	(0.525)	(0.526)
Inflation (1995-2004)	-0.066	-0.070	-0.078	0.049	0.043	0.044
	(0.029)	(0.288)	(0.028)	(0.019)	(0.019)	(0.019)
number of obs.	61	61	61	61	61	61
R^2	0.47	0.48	0.49	0.46	0.48	0.47

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. The variables Investment and FDI are both obtained from the period 2003-2005. The variable seign. stands for seignorage. MR is short for marginal rate. Estimation method: OLS.

Table 22: Replication of table 5D of article about the causality between corporate tax rates and investment and entrepreneurship:

Tax rates:	Business density			Entry rate(aver.)		
Statutory corporate	-0.034			-0.029		
	(0.083)			(0.086)		
1st-year effective		-0.068			-0.083	
		(0.092)			(0.094)	
5-year effective			-0.070			-0.133
			(0.103)			(0.103)
Additional controls:						
Alternative taxes	-0.312	-0.263	-0.262	0.012	0.083	0.124
	(0.156)	(0.171)	(0.175)	(0.155)	(0.175)	(0.176)
PIT top MR	-0.053	-0.048	-0.052	0.006	0.016	0.016
	(0.043)	(0.043)	(0.041)	(0.054)	(0.054)	(0.052)
VAT and sales taxes	0.025	0.015	0.012	0.065	0.043	0.022
	(0.054)	(0.056)	(0.058)	(0.102)	(0.104)	(0.105)
log(total of payments)	-0.362	-0.269	-0.319	-2.145	-1.978	-1.940
	(0.811)	(0.821)	(0.810)	(0.845)	(0.850)	(0.823)
log gross domestic product per capita 2003	1.833	1.711	1.747	-0.402	-0.535	-0.551
	(0.737)	(0.727)	(0.724)	(0.795)	(0.794)	(0.782)
Severity of employment	-0.019	-0.019	-0.021	-0.012	-0.010	-0.015
	(0.033)	(0.033)	(0.034)	(0.033)	(0.033)	(0.033)
Property right index	-0.045	-0.035	-0.038	-0.062	-0.041	-0.029
	(0.053)	(0.056)	(0.055)	(0.062)	(0.066)	(0.065)
Procedures to begin a business	-0.100	-0.081	-0.090	-0.179	-0.118	-0.101
	(0.182)	(0.182)	(0.181)	(0.223)	(0.231)	(0.222)
EWF	1.645	1.530	1.506	1.410	1.337	1.126
	(1.002)	(0.968)	(1.008)	(0.809)	(0.742)	(0.767)
Seign. 2004	-0.087	-0.115	-0.110	-0.191	-0.252	-0.265
	(0.129)	(0.135)	(0.134)	(0.170)	(0.183)	(0.177)
Tax evasion	-0.627	-0.624	-0.606	0.530	0.459	0.448
	(0.655)	(0.648)	(0.647)	(0.715)	(0.712)	(0.699)
Inflation(average) (1995-2004)	-0.006	-0.006	-0.006	0.026	0.026	0.025
	(0.027)	(0.026)	(0.027)	(0.025)	(0.024)	(0.024)
number of obs.	60	60	60	60	60	60
R^2	0.47	0.47	0.47	0.41	0.42	0.44

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. The variable average entry rate is obtained from the period 2000-2005. The variable seign. stands for seignorage. MR is short for marginal rate. Estimation method: OLS

9.3 Appendix C

Table 23: Results of applying the DML method to the data of table 1 of the causality between inherited trust and economic growth

Africa, separately	Inher. trust in 1935 (1)	Inher. trust in 2000 (2)
All methods(excl. NNet)(50,2)	-0.279***	-0.246***
	(0.039)	(0.036)
number of obs.	6535	4491

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: DML. This table combines table 1 of the paper were they use dummies for each period with taking both periods separately.

Table 24: Results of applying the DML method on the data of table 6 of the causality between inherited trust and economic growth

	Income per capita:				
All methods(excl.NNet)(100,2)	33463,75***	18100***	27600***	19700**	15400***
	(7563,398)	(3630)	(4050)	(7480)	(3220)
All methods(excl.Nnet)(100,5)	34300***	24100***	22900***	22000***	22500***
	(7240)	(3490)	(5320)	(4030)	(3830)
number of obs.	48	48	46	46	46

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS. The dependent variable is represented by the economic performance in 1935 and 2000. Additionally controls are multiple countries.

Table 25: Results of applying the DML method on the data of table 11 of the causality between inherited trust and economic growth

	Income per capita:				
All methods(excl.NNet)(100,2)	21600***	32900***	25800***	25700***	21900***
	(4030)	(6140)	(5970)	(5420)	(5610)
All methods(excl.Nnet)(100,5)	25600***	25200***	19900***	20900***	23100***
	(4930)	(5030)	(4780)	(4290)	(5300)
number of obs.	46	46	46	46	46

Note: *,**,*** significant at the 1,5 and 10 percent level respectively. Estimation method: OLS. The dependent variable is represented by the economic performance in 1935 and 2000. Additionally controls are multiple countries.

9.4 Appendix D

To test for significance of the estimated coefficient, $\hat{\beta}$, of a regression this paper uses a standard t-test that is given by:

$$t = \frac{\hat{\beta} - \beta_0}{\sigma} \sim \mathcal{N}(0, 1) \tag{17}$$

Where $\hat{\beta}$ represents the DML estimates, β_0 represents the value under the null hypothesis, σ represents the standard deviation of $\hat{\beta}$. In this paper the value of the β_0 is zero, for the results section. The critical value for the two sided t statistic is represented by t(1-$\alpha/2$, n-k), with 1-$\alpha/2$ being the significance level and n represent the number of observations and k the number of regressors. When n is large the t-test converges to a z-test. Then the critical values for 10%, 5% and 1% significance level are approximately 1.645, 1.960 and 2.576 respectively.