Do Latin Classes Make Schoolchildren Perform Better? Evidence from French Administrative Data

Anthony KUYU

May 2023

Abstract

Using panel data from the French Education Ministry, this paper assesses the claim that studying ancient languages, such as Latin or Ancient Greek, has benefits beyond any granted by a better knowledge of these languages, such as improved understanding and mastery of languages rooted in them, and/or stronger logical thinking and reasoning abilities. Without conducting a randomized controlled trial, drawing a causal inference becomes complicated due to various confounding factors. To estimate a potential causal effect of taking latin classes, I rely upon a matching identification strategy, where I match treated and control observations on covariates selected through a rigorous variable selection technique, LASSO. I find a modestly sized treatment effect; however, a placebo test reveals that even with a relatively large number of controls, some confounding remains — I discuss the methodological implications of that finding. Given that the treatment effect estimate I find is not very large and is relatively close to the placebo estimate, I tentatively conclude that this treatment effect estimate should likely be interpreted as an upper bound for any potential causal effect..

Acknowledgements

First, I would like to address my thanks to my advisor, Clément de Chaisemartin, for his always helpful and relevant advice and guidance. This work owes a lot to his ideas, remarks, and noticing of some of my mistakes. I also thank Carlo Barone for kindly accepting to be on my jury.

As the writing of this thesis marks the end of this program, I want to thank the Department of Economics of Sciences Po, as well as the administrative staff, for allowing me to benefit from these difficult but rich years.

I also of course have to thank my classmates. It was a true privilege to be among such a smart and kind bunch — I have no doubt that, whatever path each one chooses, it will be a successful one.

I would like to save some specific words for the close friends I made on the way: Dávid, Hugo, Ruijing, Jeteesha, Romain and Nicolas. Be it grinding through problem sets, preparing presentations, studying for exams in ”the War Room”, or having a beer on a nice day, I will truly cherish the moments we spent together — and, of course, the ones to come.

I want to thank Halle, for supporting me all the way through. That being said, without her this paper may have been finished a lot sooner...

Finally, I want to thank my parents, for supporting me all the way through my (overdrawn) studies and, more importantly, always believing in the soundness of my ambitions, and my ability to achieve them.

Of course, all errors are mine.

Contents

1 Introduction 4

2 Context and descriptive statistics 6

3 Data 10

4 Identification strategy 11

4.1 Model selection 11

4.2 Multiple imputation 14

4.3 Selected model 16

4.4 Nearest neighbor matching 20

4.5 Estimation of the average treatment effect on the treated (ATT) 22

4.6 Rubin's rules for pooling estimates 22

5 Results 23

6 Discussion 26

7 Conclusion 28

1 Introduction

A common trope of the discourse on education in France is the debate on the importance of learning ancient languages, namely Latin and Ancient Greek. Some argue the focus on these languages is evidence of inept conservatism, of inability to move on from the past; that if we are to teach children languages, Python or C++ would be more apt choices (though that particular argument may recede with the recent advances of AI). The opposite side generally answers with two arguments: the first is essentially that conservatism is good, in this matter at least; the second, on which I focus, is that learning ancient languages has important benefits for the development of other skills and acquisition of other forms of knowledge.

In particular, it is claimed, learning Latin or Ancient Greek will provide schoolchildren with deeper understanding of French, and leave them with a richer vocabulary and generally a much better command of the language. Some also claim that, as Latin and Ancient Greek are (supposedly) ”logical” languages, learning one of them improves pupils’ logical thinking abilities and would make them more capable at mathematics and other problem-solving tasks.

Unfortunately, as is the case for many policy-related claims, not much evidence is presented to support these latter arguments. Empirically, as I will show and as others have shown (Gasq, Touahir (2015)), students who study ancient languages outperform on average those who do not. But, of course, causality doesn’t follow from this observation: it also happens that children who study Latin or Ancient Greek are disproportionately coming from a socially privileged background, and are disproportionately pupils who showed good school performance before starting to study the language.

This study assesses whether the association is causal. For this purpose, I use panel data obtained from the French Ministry of Education. As they followed approximately 35 000 students entering middle school in 2007, they collected a wealth of information on them — individual characteristics, social background, family characteristics, classes taken... — and also tested their performance on several occasions.

The existence of this data provides me with the opportunity to observe the variation on children's' test results depending on whether they took Latin classes or not, while controlling for several important confounding variables. A large literature exists on the determinants of school performance: the intuitively unsurprising consensus is that the main factors contributing to it are cognitive ability (Giofrè et al., 2017), psychological characteristics (Lee and Stankov, 2018), parental involvement (Castro et al., 2015) or families' socioeconomic status (as shown, for instance, by analysis of PISA results by the OECD). One could expect the choice to study Latin to be associated with a similar set of factors, leading to an important confounding problem for identification.

I first use LASSO to select the most relevant variables to be included as controls in the analysis. I then match pupils who study latin and pupils who do not on this set of controls, and I estimate average treatment effect on the treated (ATT). I find a relatively modest positive effect; however, a placebo test shows that, despite controlling on a relatively large set of potential confounding factors, some confounding remains as I find a smaller but non-negligible effect that shouldn't appear. The estimate seems likely to be biased upwards, and should probably be treated as an upper bound for the true causal effect, if there is any.

The main contribution of this paper is addressing a gap in the research, as very few studies have attempted to answer this causal inference question outside of the United States context and, to the best of my knowledge, none in France. Experiments conducted in the United States generally find large positive effects of studying Latin on performance in English for native English speakers, and more ambiguous effects for other cognitive skills. Bracke, Bradshaw (2020) reviews a century of available literature and finds that "while the collated data do provide significant evidence for the beneficial impact of learning Latin on the L1 development of English native speakers, evidence for an impact on MFL [modern foreign language] and cognitive development is less substantial". They note however that most of the studies are old and often of questionable quality, as they tend not to provide enough information on their methods for proper assessment.

The seminal study on this topic is Masciantonio (1977), which documents a series of experiments conducted in the US in the 1970s, and finds they all showed large effects when it comes to mastery of English. Unfortunately, as observed by Bracke, Bradshaw (2020), the methods and results are quite succinctly presented. In particular, it appears from the descriptions that in most if not all experiments, the assignment of pupils to treatment and control group was not conducted at random, leaving room for selection effects impacting the estimates. The results I find significantly diverge from those suggested by existing literature. This discrepancy could be explained either by the methodological weaknesses within that literature or by pedagogical differences in the teaching of Latin between France and the United States.

The second contribution of this paper is a methodological one, as it provides an assessment of how far some of the traditional causal inference methods developed by econometricians can take us when it comes to cases with important confoundings. While matching is a well-known empirical strategy, it is here associated with more recent methodological advances, such as the use of penalized regression for variable selection; it is also associated with sophisticated techniques from the statistical literature in order to deal with the problem of missing values. My results tend to show that econometric methods which rely on independence conditional on some controls to identify a causal effect may be insufficient, at least in some contexts, to fully take care of the confounding problem, even when reinforced by sophisticated selection methods.

2 Context and descriptive statistics

French middle school lasts four years (barring cases when pupils have to retake a year) and starts when children generally are 11-12 years old, at a level called *sixième*, and ends when they are 14-15 years old, at a level called *troisième*. At the end of their *troisième*, French pupils are to take a national examination, *le brevet des collèges*, in order to graduate and proceed to high school.

Middle schoolers have to take classes in two foreign languages, the first generally being English, and the second being chosen among many options — the most common ones being Spanish and German. From the second year of middle school onwards (*cinquième*) they have the possibility of taking optional courses, Latin being one of the most common ones.

Students who choose Latin are expected to follow the courses for the three remaining years of middle school, and can keep following them in high school, if option is available in the school they attend.

Latin (and to a much lesser extent Ancient Greek) teaching has been the subject of several debates (and policy reforms) in the last decades — the full details of which are beyond the scope of this paper. One of the salient points of these debates is the fact that socially and economically privileged students are over-represented among the pupils who study Latin, and socially and economically underprivileged students are under-represented. As one can intuit, and as the evidence shows, children who study Latin also tend to perform much better than those who don't. The following figures show that the percentage of pupils who study Latin varies with social background.

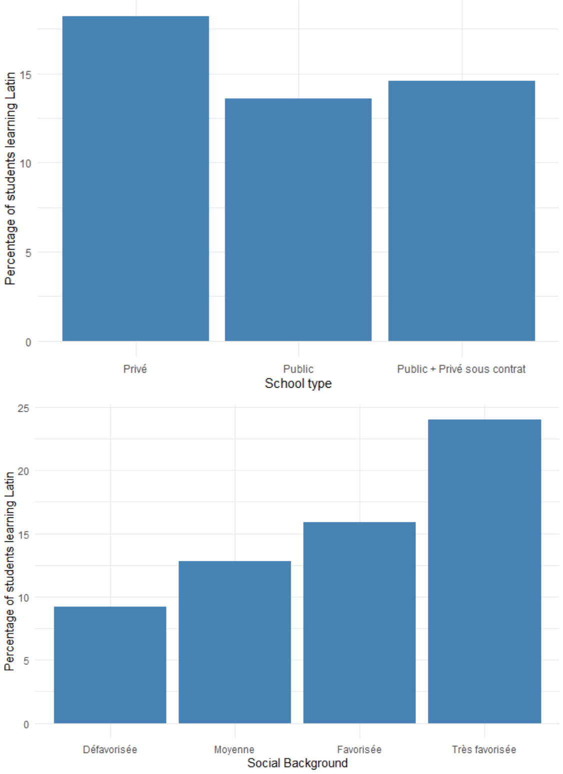

Figure 1: Percentage of pupils studying Latin per social background and school type (Ministry of Education data)

One can see that private school pupils are much more likely to take Latin classes, and that the likelihood of taking Latin classes increases as the background of the pupil is more socially favorable. It is well known that French private schools tend to welcome more privileged students (Tavan, 2005).

It is also well-established empirically that social background and school performance are correlated. The following graph shows some more evidence of this observation by showing the correlation between test results and the (log-)income of the household. The correlation is not strong ($R^2 = 0.13$), as income is not a perfect proxy for all variables that may define "social background" and many other factors come into play, but it is significant and illustrates the issue.

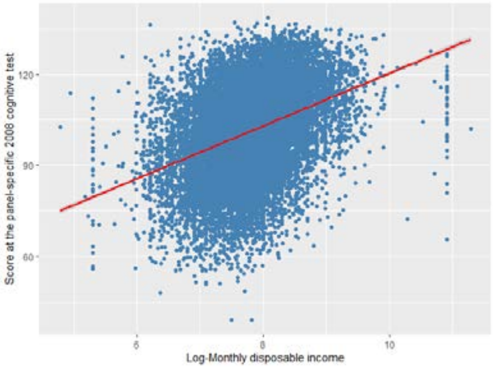

Figure 2: Plot of global 2008 cognitive test performance on log-scale monthly disposable income

Table 1: Regression analysis of test performance (2008 cognitive test) and log-monthly income

Variable	**Estimate**	**Std. Error**	**t value**
(Intercept)	31.5747***	1.1235	28.10
log monthly income	8.8957***	0.1433	62.08
$Pr(>\|t\|)$			
(Intercept)	$< 2 \times 10^{-16}$		
log monthly income	$< 2 \times 10^{-16}$		

Building upon the same figure, one can show that pupils who take Latin classes tend to be students who come from higher-income households, and they have stronger test performances.

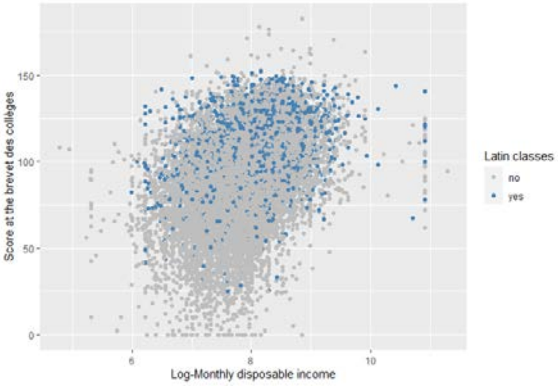

Figure 3: Plot of *brevet* performance on log-scale monthly disposable income, separating pupils who took Latin classes and those who didn't

This graph shows test results of the *brevet des collèges*, which as explained above is taken after students start having the opportunity to take Latin classes. So the disparity observed may be in part caused by a potential causal effect of Latin. However, the graph below shows there is a large disparity even without a causal effect:

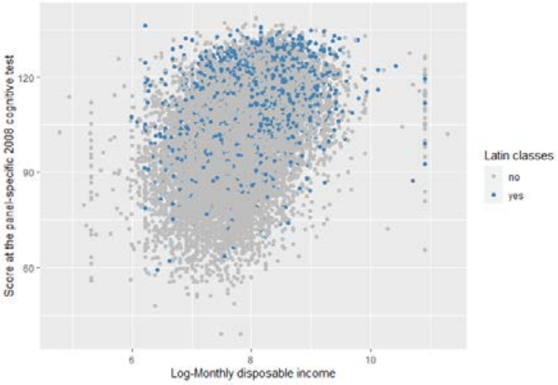

Figure 4: Plot of global 2008 cognitive test performance on log-scale monthly disposable income, separating pupils who took Latin classes and those who didn't

Indeed, this plot shows test results at the cognitive evaluation conducted in 2008, at the end of *sixième*, that is before pupils could start Latin classes. Therefore, the disparity we observe here cannot be caused by Latin. This all shows that the choice to take Latin, social background and test performance are correlated with each other, which causes an important identification problem, without mentioning the many other variables that probably come into play.

3 Data

The dataset I use is provided by the *Direction de l'évaluation, de la prospective et de la performance* (DEPP), which is the service in charge of producing data and analysis in the French Ministry of Education. It is built upon a panel of pupils who entered middle school in 2007. 35 000 pupils (out of the 760 000 who entered middle school on that year) were randomly drawn into a balanced sample, constructed in order to ensure representativity. In total, the dataset contains 694 variables.

Information on the pupils was collected in several ways by the DEPP in order to construct the dataset, including:

1. collection of basic information on the children's schooling years (such as the classes taken, the type of school(s) attended...);

2. tests on pupils' cognitive and conative skills, which were conducted in 2008, 2011 and 2012;

3. comprehensive surveys sent to the families of all children selected into the panel in 2008 and 2011, to have a better understanding of their social and familial background.

In practice, I work with a subset of this dataset. I filter the observations of pupils who did not participate in the 2007 national evaluation, the 2008 panel-specific evaluation and the 2011 panel-specific evaluation; I filter as well the observations for which parents didn't respond to the 2008 family survey. This reduces the number of observations to 20 743.

I also remove many variables from the dataset. I select out the variables with more than 50 percent missing values. I also select out the variables that I know I am not going to use as controls, which are most of the variables collected by the 2011 survey, most of the basic information variables pertaining to the school year 2009-2010 and onwards, and any variable pertaining to the year 2012. Indeed, as I intend to test the methodology against a placebo, that is 2008 cognitive test results obtained before the start of Latin classes, I do not add as controls student characteristics from the school years coming after Latin classes start. These post-2008 test characteristics being obviously strongly correlated with pre-2008 test characteristics, it seems unlikely that this filtering would deprive me of much relevant information. This reduces the number of variables to 279.

As will be detailed in the next section, my main outcome variable for this analysis is test results of the 2011 cognitive evaluation for the panel.

4 Identification strategy

The main challenge when trying to assess the potential causal effect of Latin is, as is often the case, the various confounding variables that come into play. As seen above, both the treatment (Latin) and the outcome (test scores) are linked to individual characteristics such as social background, and they are presumably both correlated to psychological and cognitive characteristics - any estimate obtained from a method that doesn't take this into account cannot be interpreted as causal.

To uncover a potential causal effect, I rely on a strategy of matching: I test whether, conditional on some covariates, there is a difference between the test results of students who took Latin and students who did not. I therefore need a set of covariates that satisfies the conditional independence condition, that is that, conditional on that set, the outcome Y is independent from the treatment choice D:

$$Y \perp\!\!\!\perp D | X \tag{1}$$

4.1 Model selection

The first challenge is to select the covariates to control for: a sparse (or parsimonious) model is necessary in order to be able to interpret results. One could rely on intuition, theory and empirical literature to choose covariates that are likely to explain both test scores and the choice of Latin. However, there are more sophisticated and data-driven methods for model selection, particularly applicable in cases where one has access to hundreds of covariates related to individual characteristics.

Following Belloni, Chernozhukov, Hansen (2013) (BCH), I implement a method called post-double-selection. This simple and intuitive method is performed in three steps:

1. find variables X_D that are significant predictors of the treatment variable D;

2. find variables X_Y that are significant predictors of the outcome variable Y;

3. select as controls the union set of the variables found in steps 1 and 2.

In order to perform the first two steps, following the recommandations of BCH, I use LASSO, which is a regression method that allows for variable selection by adding a penalty on the absolute size of the coefficient estimates to the mean squared error minimization problem. Formally the LASSO estimator solves the following problem:

$$\hat{\beta}_{lasso}(\lambda) = \arg\min \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 + \frac{\lambda}{n}\sum_{j=1}^{p}\psi_j|\beta_j| \tag{2}$$

Where λ is a tuning parameter that controls the overall penalty level, and ψ is a predictor-specific penalty.

For LASSO to be useful, it needs to be model selection consistent, which means that the true model is selected with probability approaching 1 as $n \to \infty$. The condition that is sufficient and almost necessary for LASSO to be model selection consistent (the "irrepresentable condition", Zhao and Yu (2006); Meinshausen and Buhlmann (2006)) applies strong constraints on the degree of correlation between predictors in the true model and predictors outside the model (Ahrens, Hansen, Schaffer (2020)). Therefore, following Zou (2006), I use adaptive LASSO, which relaxes these constraints and where ψ is $1/|\beta_{0,j}|$, with $\beta_{0,j}$ being an initial estimator obtained by an OLS regression.

Adaptive LASSO is useful when one's outcome variable is continuous. However, in this case, I need to select predictors for both the final outcome (test scores) and the treatment (Latin classes). The treatment variable being binary, classical regression analysis isn't the most appropriate choice. I use logistic LASSO to deal with the binary variable. The estimator of logistic LASSO maximizes log-likelihood, as a traditional logistic regression does, but with an added penalty to the optimization problem:

$$\hat{\beta}_{logitlasso}(\lambda) = \arg\max \frac{1}{n}\sum_{i=1}^{n}\left[y_i(\beta_0 + x_i'\beta) - log(1 + e^{\beta_0 + x_i'\beta})\right] - \frac{\lambda}{N}||\beta||_1 \tag{3}$$

Where λ is again a tuning parameter that controls the overall penalty level. Speaking of which, for both LASSO and logistic LASSO, model selection is performed with the value of λ that minimizes Bayesian Information Criterion (BIC). BIC is defined as:

$$BIC(\lambda) = n\log(\hat{\sigma}^2(\lambda)) + df(\lambda)\log(n) \tag{4}$$

Where $\hat{\sigma}^2(\lambda) = n^{-1}\sum_{i=1}^{n}\hat{\epsilon_i}^2$, $\hat{\epsilon_i}$ are the residuals, and $df(\lambda)$ is the effective degrees of freedom, which measures model complexity. In practice, this is simply the number of regressors: Zou et al. (2007) show that the number of coefficients estimated to be non-zero is an unbiased and consistent estimate of $df(\lambda)$ for the LASSO.

Conceptually, selecting a value of λ to minimize BIC is, in essence, solving a basic optimization problem. Increasing the number of regressors will reduce the mean square residual, which reduces BIC (and improves accuracy of prediction), but it will also increase the degrees of freedom, which increases BIC: the point is to find an optimum between accuracy and sparsity of the model.

A weakness of BIC, pointed out by Chen and Chen (2008), is that its use is likely to lead to overselection of variables. Indeed, it is built upon the assumption that all models have the same prior probability; there are more possible combinations of covariates when the number of covariates rises, and therefore, for any s < p/2, there are more possible models with s + 1 covariates than models with s covariates, and therefore it is more likely a model with s + 1 covariates is going to be chosen. To solve this shortcoming, the aforementioned paper proposed the Extended Bayesian Information Criterion (EBIC):

$$EBIC(\lambda) = n\log(\hat{\sigma}^2(\lambda)) + df(\lambda)\log(n) + 2\zeta df(\lambda)\log(p) \tag{5}$$

EBIC adds a penalty on the size of the model, which increases with the number of predictors, preventing EBIC from selecting too many parameters. ζ is an additional parameter that controls the size of the penalty; its value must be in [0, 1]. Here, following Chen and Chen (2008), $\zeta = 1 - \frac{log(n)}{2log(p)}$.

4.2 Multiple imputation

Some data are missing in the dataset. The variables with a high (above 50 percent) missing rate were also not relevant for the analysis and were dropped. After data cleaning, 279 variables remain, among which 17 have missing values. The average percentage of missing value in the entire dataset is very low, about 0.29 percent. One could think this is negligible, but there are two reasons why it isn't. The first is that most of the variables with missing values are related to cognitive evaluations passed by students, which makes it likely they are useful as controls. Besides, a naive complete-case analysis (that is, removing all observations with any missing value) becomes very quickly wasteful, as even a low number of missing values can be spread all over the dataset. Case in point, a complete-case analysis in this situation would have meant removing 5879 observations out of 20743, close to 30 percent.

To deal with this problem, I use multiple imputation (Rubin 1987b; Rubin 1996), as, along with maximum likelihood approaches, it is the most robust method to deal with incompleteness in most cases (Schafer, Graham, 2002). The principle of multiple imputation is that, instead of creating one variable using one method or another and pretending it is real, several likely possibilities are generated, in order to maintain some of the uncertainty surrounding the missing value. This means that, in practice, several different datasets are created, each with a different value for each missing variable in the original dataset.

The main benefit of multiple imputation is that it solves a common problem with other imputation methods, which is that the standard errors of estimates obtained are not going to take into account the uncertainty surrounding the missing values, and therefore are going to be smaller than they should be in order to properly reflect the degree of uncertainty we should have regarding the estimates (van Buuren, 2018). The pooled standard errors obtained after analysis takes into account both the variability within each imputed dataset and the variability across all imputed datasets (see section 4.6 for more details).

I create five imputed datasets, using predictive mean matching (Rubin (1986), Little (1988)):

- the missing value is predicted through a linear main effect model conditional on all other variables;

- among complete cases, five potential "donors" with predicted values as close as possible to the missing one are selected;
- one "donor" is randomly selected among the five candidates, and its observed value replaces the missing value.

More formally, an imputed value $\dot{y_j}$ is computed as follows. For each pair of observation (i, j), where i denotes observations with the observed target value, and j denotes observations with the missing target value, a distance metric is computed using Bayesian parameter draws:

$$\dot{\eta}(i,j) = |X_i\hat{\beta} - X_j\dot{\beta}| \tag{6}$$

Where $\hat{\beta}$ is a regression weight, while $\dot{\beta}$ is a randomly drawn value from the posterior distribution of β. These two vectors are constructed as follows:

$$\hat{\beta} = VX'_{obs}y_{obs} \tag{7}$$

Where $V = (S + diag(S)\kappa)^{-1}$, $S = X'_{obs}X_{obs}$, κ is a ridge parameter set to a very small value (here $\kappa = 0.0001$) to avoid problems arising from singular matrices, X_{obs} is the $n_1 \cdot q$ matrix containing the values of covariates for all observations where the target value y is observed, and y_{obs} is the $n_1 \cdot 1$ vector of observed data in the target variable y.

$$\dot{\beta} = \hat{\beta} + \dot{\sigma}\dot{z}V^{-1/2} \tag{8}$$

Where $\dot{\sigma} = (y_{obs} - X_{obs}\hat{\beta})'(y_{obs} - X_{obs}\hat{\beta})/\dot{g}$, $\dot{g} \sim \chi^2_\nu$ $(\nu = n_1 - q)$, $\dot{z} \sim N(0,1)$.

Once the distance $\dot{\eta}(i,j)$ is computed for all couples, a set Z_j is constructed for all missing values, each containing five donors, such that $\sum \dot{\eta}(i,j)$ is minimized. Then the imputed value $\dot{y_j}$ is randomly selected among Z_j.

This method has several advantages. One of the main ones is the fact that imputations are based on values observed elsewhere, so they are realistic: assuming a clean dataset, it is impossible for this method of imputation to produce a meaningless value (such as, for instance, a negative value for an intrinsically positive variable). This method rests upon two assumptions:

- ignorability: missing values are Missing At Random, which means the probability of the value being missing is independent from the value itself. This assumption is reasonable in this case: the variables with missing values were mostly test result items, which were to be collected/reported by the *Education nationale* without specific involvement from the parents or pupils.;
- the linear main effect model may be inadequate in the presence of strong nonlinear relationships.

4.3 Selected model

After imputations, I implement the post-double-selection on each dataset, and use a simple majority criterion in case different models are selected across imputed datasets. This majority method is admittedly not the most sophisticated approach (Wood, White, and Royston (2008)). Ideally, performing LASSO across all stacked datasets (as the MI-LASSO method by Chen and Wang (2013)) would have been the most robust, but given the low percentage of missing values, and therefore the low variability between imputed datasets, the gains would likely not justify the substantial computational demands. Table 1 and Table 2 show the selected variables.

Table 2: Variables selected as significant predictors of 2011 cognitive test scores

Selected variables positively associated with test results	Selected variables negatively associated with test results
Self-efficacy score: self-regulation	Self-efficacy score: social skills
Self-efficacy score: schoolwork	Subscore in French at the 2007 national testing: arithmetics
Score in French at the 2007 national testing	Age when entering sixième
Subscore in French at the 2007 national testing: ability to write a text	Learning Spanish as a second language
Subscore in French at the 2007 national testing: ability to transform a text	Attending professional discovery classes
Subscore in French at the 2007 national testing: ability to recognize words	Leaving in an urban zone with 50K to 100K inhabitants
Subscore in French at the 2007 national testing: ability to interpret	Leaving in an urban zone with 200K to 2M inhabitants
Subscore in French at the 2007 national testing: ability to analyze	Living in the following départements: Seine-Saint-Denis, Nord-Pas-de-Calais
Subscore in French at the 2007 national testing: ability to observe and look for information	School enrolled in ”ambition réussite” network
Subscore in French at the 2007 national testing: ability to select and link words and sentences	Non-working mother
Score in mathematics at the 2007 national testing	Unemployed mother
Mother has a graduate degree	Mother has never worked
Mother has an undergraduate degree	Mother born in a foreign country
More than 200 books in the house	Less than 30 books in the house
No repeated year	Parents regularly attend sports games
Parents think their child was a ”good” or ”excellent” student in elementary school	Father is a skilled workman
Child does not benefit from free tutoring at school	SEGPA (separated groupings of pupils with important difficulties)
Parents think their child is a ”good” or ”excellent” student in middle school	Repeated year
No television in the room of the child	Urvan area with special difficulties
Private school	Turkish mother
Chef de famille ”professions libérales”	Morrocan father
Father is:	
a corporate executive	
an engineer	
”Classe européenne” or ”classe internationale”	
	Studying in the following administrative académies:
Studying in the following administrative académies:	Lille
Besançon	Aix
Caen	Amiens
Clermont-F	Nice
Lyon	Montpellier
Poitiers	Strasbourg
Rennes	Créteil
Nantes	Versailles
Orléans	Martinique
Rouen	La Réunion
	Guadeloupe

Table 3: Variables selected as significant predictors of taking Latin classes

Selected variables negatively associated with test results	Selected variables negatively associated with Latin
Self-efficacy score: schoolwork	Parents think their child had ”some” difficulties in elementary school
Score in French at the 2007 national testing	Parents think their child has ”some” difficulties in middle school
Score in mathematics at the 2007 national testing	Parents think Professional *baccalauréat* is the best path for employment
Spanish as a second foreign language	
German as a second foreign language	
Mother has a graduate degree	
Child does not benefit from free tutoring at school	
Parents think their child is an ”excellent” student in middle school	
Parents think a college degree is the best path for employment	
Head of household is a teacher	
General *cinquième*	

These results are mostly what one would have expected. Past test scores positively predict test scores; the level of education and the occupation of parents, attending a private school, and having many books at one's home do as well. Meanwhile, variables like having a non-working mother, living in certain impoverished *départements*, such as Seine-Saint-Denis or Nord-Pas-de-Calais, or having a foreign-born mother are negatively associated with test scores. One intriguing selection is the fact that higher score in arithmetics in 2008 would be associated with *lower* overall score in 2011, while a higher score in mathematics overall in 2008 is logically associated with higher scores in 2011.

When it comes to the variables that were selected to influence taking Latin classes, results are again mostly what one would have expected. Pupils who are more capable in French and maths, and have higher self-efficacy when it comes to schoolwork, are more likely to enroll in Latin classes; mothers with graduate degrees are more likely to steer their children toward a more elitist path; parents with more prestigious and elitist ambitions for their children, as well as parents who have more faith in their children's abilities, are more likely to encourage them to take Latin... Some results might seem more surprising; for instance, it's not immediately apparent why taking Spanish or German as a second foreign language would make a student more likely to take Latin classes.

4.4 Nearest neighbor matching

I impute five times a subset of the dataset containing the controls and relevant outcome variables, following the same procedure as described above. Then I use one-to-one nearest neighbor matching with replacement on each dataset: each treated observation is matched with the untreated observation that is the closest to it with regards to the control variables. The measure of "distance" between treated and untreated observations is computed using Mahalanobis distance. The distance between two observations x_1 and x_2 (treated and untreated, respectively) is the following:

$$d(x_1, x_2) = \sqrt{(x_1 - x_2)' S^{-1} (x_1 - x_2)} \tag{9}$$

Where x_1 and x_2 are vectors including all selected variables for one observation, and S is the covariance matrix for all selected variables. This latter term allows to have a measure of distance that is invariant to the degree of correlation between the variables.

This method provides satisfactory balance for the data. Balance refers to the distribution of covariates or confounding variables between treatment and control groups being similar. Achieving balance is crucial because it helps ensure that any observed differences in outcomes between the treatment and control groups are primarily due to the treatment effect and not confounding factors.

Mean differences and plots of distribution are commonly used to assess balance. Mean differences compare the average values of covariates between the treatment and control groups. If the mean differences across covariates are small or negligible, it suggests that the treatment and control groups are well-balanced concerning those variables. Plots of distribution provide a visual representation of the covariate distributions for both groups. Ideally, the distributions should overlap and show similar shapes and spread. Substantial differences in the distributions between the groups would indicate an imbalance in the covariates.

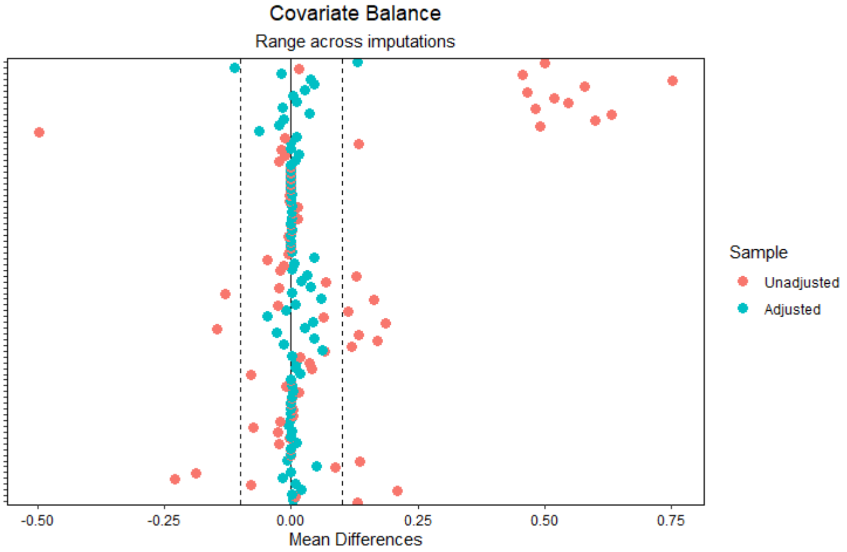

Figure 5: Balance measure - mean differences

As depicted in Figure 5, the overall balance appears satisfactory, as the absolute values of most mean differences do not exceed 0.1, with many being very close to zero. Further individual analysis of the covariates generally shows good balance, although it isn't perfect across the board (see appendix 1).

4.5 Estimation of the average treatment effect on the treated (ATT)

The final step is to estimate a treatment effect. I choose as an estimand the average treatment effect on the treated, ATT, as I am interested in the impact of Latin classes on those who take Latin classes. I use the method of G-computation (Snowden, Rose, and Mortimer 2011) (or regression estimation (Schafer and Kang 2008)) :

1. using the matched dataset, fit a simple OLS regression: $Y_i = \alpha D_i + \beta X_i + u_i$. The incorporation of covariates into the outcome model after matching, the usefulness of which may not appear obvious, serves multiple purposes: enhancing the accuracy of the effect estimate, mitigating bias caused by remaining imbalances post-matching; it allows for a more robust estimate, which remains consistent if either the matching process sufficiently reduces covariate imbalances or if the regression model is well specified. Robust standard errors are used;

2. for each treated unit, compute the outcome value predicted by the model if $D_i = 1$ and if $D_i = 0$. So for each observation we have $\widehat{Y(1)}_i|D_i = 1$ and $\widehat{Y(0)}_i|D_i = 1$;

3. compute the average of the two cases: $E(\widehat{Y(1)}|D = 1)$ and $E(\widehat{Y(0)}|D = 1)$;

4. $\widehat{ATT} = E(\widehat{Y(1)}|D = 1) - E(\widehat{Y(0)}|D = 1) = E(\widehat{Y(1)} - \widehat{Y(0)}|D = 1)$.

As the estimated effect is a function of the coefficient estimates, the Delta method allows to compute its standard error.

4.6 Rubin's rules for pooling estimates

As the estimation procedure described above is performed within each imputed datasets, to obtain the final results I must pool the estimates, standard errors and p-values following Rubin's rules, which are the following:

The pooled estimate is simply the mean of all estimates:

$$\overline{\theta} = \frac{1}{m} \sum_{i=1}^{m} \theta_i \tag{10}$$

Where m is the number of imputations (so here, m = 5). The total variance is a linear function of the variance between imputations and the variance within imputations:

$$V_T = V_W + V_B + \frac{V_B}{m} \tag{11}$$

Where $V_B = \frac{\sum_{i=1}^{m}(\theta_i - \overline{\theta})^2}{m-1}$ and $V_W = \frac{1}{m}\sum_{i=1}^{m} SE_i^2$. Of course, the pooled standard error is $SE_P = \sqrt{V_T}$. One can verify from the pooled standard errors (see Table 4 in section 5) and the individual standard error for each dataset (see Table 7 and 8 in appendix 3) that the pooled standard errors is larger than the individual standard errors.

Significance testing is performed by computing the pooled Wald value:

$$WALD_P = \frac{\overline{\theta} - \theta_0}{SE_P} \tag{12}$$

The pooled Wald value follows a t-distribution. The degrees of freedom are computed by the following formula:

$$\frac{A * B}{A + B} \tag{13}$$

Where $A = \frac{m-1}{\lambda^2}$ and $B = \frac{(n-k)+1}{(n-k)+3} * (n-k)(1-\lambda)$. n being the sample size of the imputed dataset, k the number of parameters to fit, and λ a measure of the fraction of the variance due to missingness, equal to $\frac{V_B + \frac{V_B}{m}}{V_T}$. The degrees of freedom allow to compute the t-statistic, which will then allow to compute confidence intervals:

$$CI = \overline{\theta} \pm t_{df,1-\alpha/2} * SE_P \tag{14}$$

5 Results

Following this method, I compute the results detailed below. Note that all outcome variables are standardized, so the results are expressed as fractions of a standard deviation:

- ATT of studying latin on the results of a panel-specific cognitive test in 2011;
- For a placebo comparison, I also calculate the ATT of studying Latin on the results of a panel-specific cognitive test from 2008, which is before the pupils started studying

Latin. If the methodology has indeed taken care of confounding, one would expect this ATT to be equal or very close to zero;

Table 4 shows the results. I also try to implement the same method, using additional controls: the ones that were selected for a minority of imputed datasets during the model selection procedure. For additional comparison, I try to use the G-computation procedure, without using any controls nor matching. This test allows us to see how large the estimated average treatment effect on the treated would be without the matching and controls used and, therefore, gives us an indication of how useful this procedure is.

I also perform the analysis without matching, only relying on the G-computation, to see how the results differ, if at all. I do the same in reverse, matching and performing the G-computation without any control in the regression.

Table 4: Results for the initial analysis, an analysis with additional controls and comparison analysis with no matching, then no controls at all, and matching then no controls included in the regression

	Initial analysis		Extra controls		No matching, controls		No matching, no controls		Matching, no controls	
	Treatment	Placebo	Treatment	Placebo	Treatment	Placebo	Treatment	Placebo	Treatment	Placebo
Estimate	0.1586***	0.0862***	0.1695***	0.0984***	0.1529***	0.0984***	0.8418***	0.7613***	0.1672***	0.0937***
Std. Errror	0.0149	0.0173	0.0152	0.0171	0.0103	0.0171	0.0169	0.0180	0.0245	0.0275
Conf. Interval	0.1293 - 0.1879	0.052 - 0.1201	0.1395 - 0.1994	0.0648 - 0.1321	0.1326725 0.1730541	0.0648 - 0.1321	0.8086 - 0.875	0.726 - 0.7966	0.1192 - 0.2152	0.0395 - 0.1479

6 Discussion

As Table 4 shows, I find a statistically significant placebo estimate of about 9 percent of a standard deviation, while the treatment effect estimate found is about 16 percent of a standard deviation. The fact that the placebo estimate is non-null indicates that the matching strategy was not fully successful in eliminating omitted variable bias, and that despite the large set of controls, there are still some small but non-negligible differences between the matched treated observations and control observations. However, one should conclude that if any effect of latin exists, it is probably even smaller than the already small effect size found: the placebo ATT is positive and, from an empirical viewpoint, any uncontrolled for differences between treated and untreated observations are likely to be in favor of treated observations. I therefore conclude that the treatment effect estimate I find should probably be considered as an upper bound for the true causal effect. This hypothesis is also supported by the test of estimating the ATT without matching and without controls, reported in Table 4. Even with all confounding factors, there isn't much difference between the estimated mean difference before studying Latin and after studying it, and the size of that difference is similar to the size observed with controls included, which implies a small or null causal effect. The difference observed between the treatment estimate and the placebo estimate may result from a causal effect of Latin. However, it could also be the consequence of an unseen mechanism that widens the ability gap between the treated and untreated groups over time. This highlights the importance of various placebo tests when one uses for causal inference a method that assumes independence, such as matching, as opposed to methods that exploit likely cases of independence or natural experiments. Interestingly, one can note that both adding new regressors and not matching at all and relying only on the regression yield similar results, and that matching and then regressing without controls also doesn't have much effect, meaning that at least in this case matching and G-computation with correctly selected regressors work equivalently.

The inability to completely eliminate confounding could be attributed to an inappropriate methodological choice. However, a plausible explanation could be that the set of covariates I use fails to satisfy the conditional independence assumption. Since this set was selected through a data-driven method, and not simply through *a priori* model-building, it seems to follow that the data did not have all the "right" covariates in the first place. There are empirical reasons to believe this is a good explanation. The "elephant in the room" factor I could not control for due to lack of data is IQ. The association between IQ and school performances, as well as educational and professional success in adulthood, is one of the most robust and regularly replicated results in psychology (Deary et al. (2007), Roth et al. (2015)...). IQ would without much doubt be correlated with the national evaluation test scores I use as controls, but this correlation would most likely be imperfect - therefore, test scores represent imperfect proxy for IQ which would, in all likelihood, be a stronger control. I also have imperfect controls for conscientiousness, which psychological research shows is a good predictor of schooling, educational and professional outcomes (e.g., Lounsbury, Sundstrom, Loveland, Gibson, 2003; Preckel, Holling, Vock, 2006; Trautwein, Ludtke, Roberts, Schnyder, Niggli, 2009), with the conative skills tests results. Maybe the problem is, simply, that no amount of care in selecting controls can compensate for the fact of not having the right ones available in the first place.

The differences in results observed between experiments in the US and these results could be explained by methodological weaknesses of the studies in the US. They could also be explained by different pedagogical methods. Indeed, Masciantonio (1977) argues the difference in results they observe from studies dating from the first half of the 20th century and the second half is going from a pedagogy focused on grammar and translation to a pedagogy based on an "oral-aural", "multisensory" approach. If they are right about this, it may be part of the explanation, as the French pedagogy when it comes to Latin is much closer to the "grammar-translation" end.

This paper should serve as starting point for further research. Ideally, a proper randomized controlled trial would be conducted in France, with a large number of elementary classrooms split at random into three groups: one following Latin classes with a pedagogy closer to that used in the US; one following Latin classes with a classic French pedagogy; and one with no Latin classes at all. Failing that, new observational studies should either wait for a natural experiment opportunity, or conduct a rigorous observation and testing of pupils over several years, combining econometric, sociological and psychometric knowledge in order to construct *ad hoc* the right data to serve as controls, in order to validate or overturn my results.

From education policymakers' point of view, these results make it seem unlikely that policy efforts encouraging Latin uptake would have a noticeable causal effect on school performance. While Latin teaching is certainly worthwhile on its own, investing in it more in the hope of improving performance would probably not pass a sensible cost-benefit test, unless, possibly, there are important changes in pedagogy.

7 Conclusion

This paper has attempted to assess the causal impact of Latin on middle schoolers' test performances, in the French context. Without a randomized experiment or natural experiments to exploit, this endeavour presents an important challenge. Despite employing state-of-the-art methods and advances such as control selection with penalized regressions, and opting for multiple imputation over complete-case analysis, the conducted placebo test reveals residual confounding. The results obtained after different tests tend to show that Latin classes do not have an important causal effect, if they have any at all. This work also informs us on the limits of matching methods for causal inference, and why they should be used with prudence.

Bibliography

Ahrens, Achim, et al. "Lassopack: Model Selection and Prediction with Regularized Regression in Stata." The Stata Journal: Promoting Communications on Statistics and Stata, vol. 20, no. 1, Mar. 2020, pp. 176–235

Belloni, A., et al. "Inference on Treatment Effects after Selection among High-Dimensional Controls." The Review of Economic Studies, vol. 81, no. 2, 24 Nov. 2013, pp. 608–650

Bracke, Evelien, and Ceri Bradshaw. "The Impact of Learning Latin on School Pupils: A Review of Existing Data." The Language Learning Journal, vol. 48, no. 2, 15 Nov. 2017, pp. 226–236

Castro, María, et al. “Parental Involvement on Student Academic Achievement: A Meta-Analysis.” Educational Research Review, vol. 14, no. 1, Feb. 2015, pp. 33–46

Chen, J., and Z. Chen. “Extended Bayesian Information Criteria for Model Selection with Large Model Spaces.” Biometrika, vol. 95, no. 3, 1 Sept. 2008, pp. 759–771

Chen, Qixuan, and Sijian Wang. “Variable Selection for Multiply-Imputed Data with Application to Dioxin Exposure Study.” Statistics in Medicine, vol. 32, no. 21, 25 Mar. 2013, pp. 3646–3659

Deary, Ian J., et al. “Intelligence and Educational Achievement.” Intelligence, vol. 35, no. 1, 2007, pp. 13–21

Gasq, Paul-Olivier, and Mustapha Touahir. Le Latin Au Collège : Un Choix Lié à l’Origine Sociale et Au Niveau Scolaire Des Élèves En Fin de Sixième. 2015

Giofrè, David, et al. “The Relationship between Intelligence, Working Memory, Academic Self-Esteem, and Academic Achievement.” Journal of Cognitive Psychology, vol. 29, no. 6, 2 Apr. 2017, pp. 731–747

Lee, Jihyun, and Lazar Stankov. “Non-Cognitive Predictors of Academic Achievement: Evidence from TIMSS and PISA.” Learning and Individual Differences, vol. 65, July 2018, pp. 50–64

Little, Roderick J. A. “Missing-Data Adjustments in Large Surveys.” Journal of Business Economic Statistics, vol. 6, no. 3, July 1988, p. 287

Lounsbury, John W, et al. “Intelligence, “Big Five” Personality Traits, and Work Drive as Predictors of Course Grade.” Personality and Individual Differences, vol. 35, no. 6, Oct. 2003, pp. 1231–1239

Masciantonio, Rudolph. “Tangible Benefits of the Study of Latin: A Review of Research.”

Foreign Language Annals, vol. 10, no. 4, Sept. 1977, pp. 375–382

Meinshausen, Nicolai, and Peter Bühlmann. "High-Dimensional Graphs and Variable Selection with the Lasso." The Annals of Statistics, vol. 34, no. 3, June 2006, pp. 1436–1462

OECD. PISA 2018 Results (Volume II). PISA, OECD, 3 Dec. 2019

Preckel, Franzis, et al. "Academic Underachievement: Relationship with Cognitive Motivation, Achievement Motivation, and Conscientiousness." Psychology in the Schools, vol. 43, no. 3, 2006, pp. 401–411

Roth, Bettina, et al. "Intelligence and School Grades: A Meta-Analysis." Intelligence, vol. 53, Nov. 2015, pp. 118–137

Rubin, Donald. Multiple Imputation for Nonresponse in Surveys. Wiley Series in Probability and Statistics, Hoboken, NJ, USA, John Wiley Sons, Inc., 9 June 1987

Rubin, Donald B. "Multiple Imputation after 18+ Years." Journal of the American Statistical Association, vol. 91, no. 434, 1 June 1996, pp. 473–473

Rubin, Donald B. "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations." Journal of Business Economic Statistics, vol. 4, no. 1, Jan. 1986, p. 87

Schafer, Joseph L., and John W. Graham. "Missing Data: Our View of the State of the Art." Psychological Methods, vol. 7, no. 2, 2002, pp. 147–177

Schafer, Joseph L., and Joseph Kang. "Average Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example." Psychological Methods, vol. 13, no. 4, 2008, pp. 279–313

Snowden, Jonathan M., et al. "Implementation of G-Computation on a Simulated Data

Set: Demonstration of a Causal Inference Technique." American Journal of Epidemiology, vol. 173, no. 7, 16 Mar. 2011, pp. 731–738

Tavan, Chloé. "École Publique, École Privée." Revue Française de Sociologie, vol. 45, no. 1, 2004, p. 133

Trautwein, Ulrich, et al. "Different Forces, Same Consequence: Conscientiousness and Competence Beliefs Are Independent Predictors of Academic Effort and Achievement." Journal of Personality and Social Psychology, vol. 97, no. 6, 2009, pp. 1115–1128

van Buuren, Stef. Flexible Imputation of Missing Data, Second Edition. Second edition. — Boca Raton, Florida : CRC Press, [2019] —, Chapman and Hall/CRC, 17 July 2018

Wood, Angela M., et al. "How Should Variable Selection Be Performed with Multiply Imputed Data?" Statistics in Medicine, vol. 27, no. 17, 30 July 2008, pp. 3227–3246

Zhao, Peng, and Bin Yu. "On Model Selection Consistency of Lasso." The Journal of Machine Learning Research, vol. 7, no. 90, 1 Dec. 2006, pp. 2541–2563

Zou, Hui, et al. "On the "Degrees of Freedom" of the Lasso." The Annals of Statistics, vol. 35, no. 5, Oct. 2007, pp. 2173–2192

Zou, Hui. "The Adaptive Lasso and Its Oracle Properties." Journal of the American Statistical Association, vol. 101, no. 476, 1 Dec. 2006, pp. 1418–1429

Appendix 1 - Balance plots for all controls, before and after matching

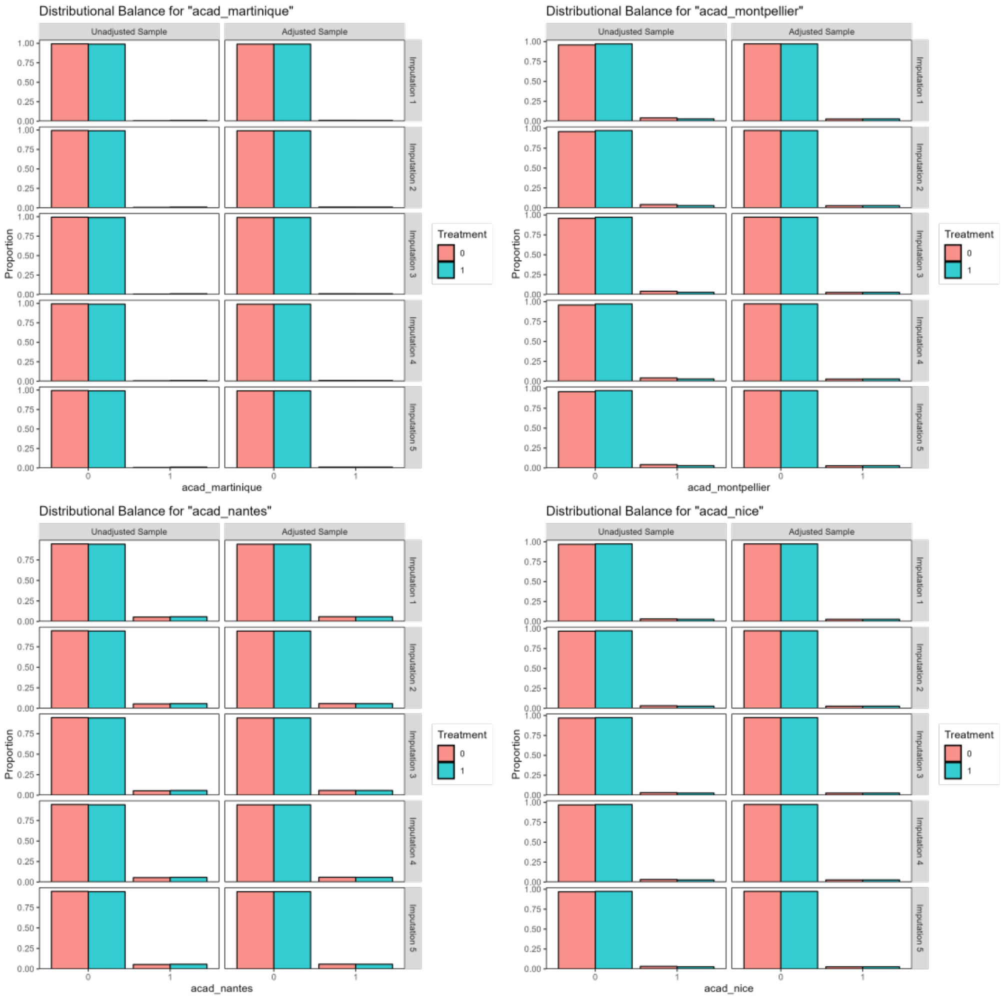

Figure 6: Balance measures on several covariates

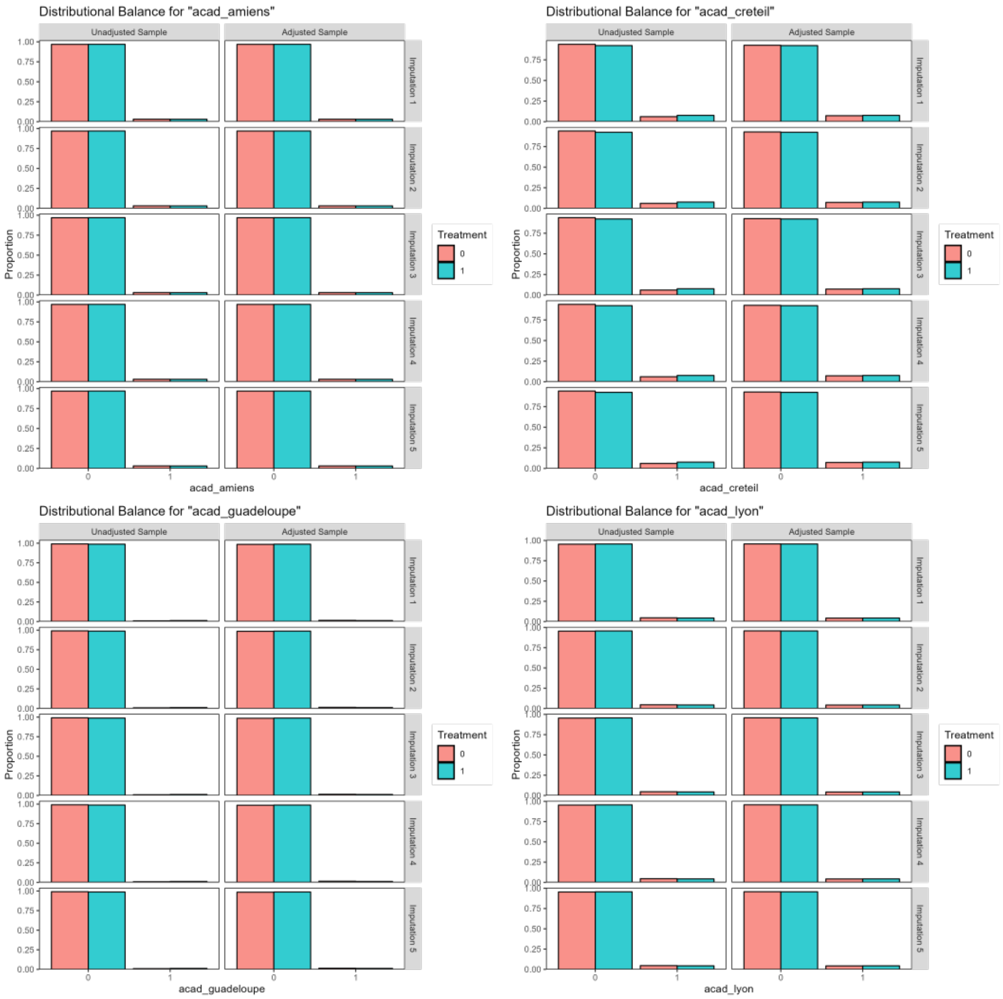

Figure 7: Balance measures on several covariates

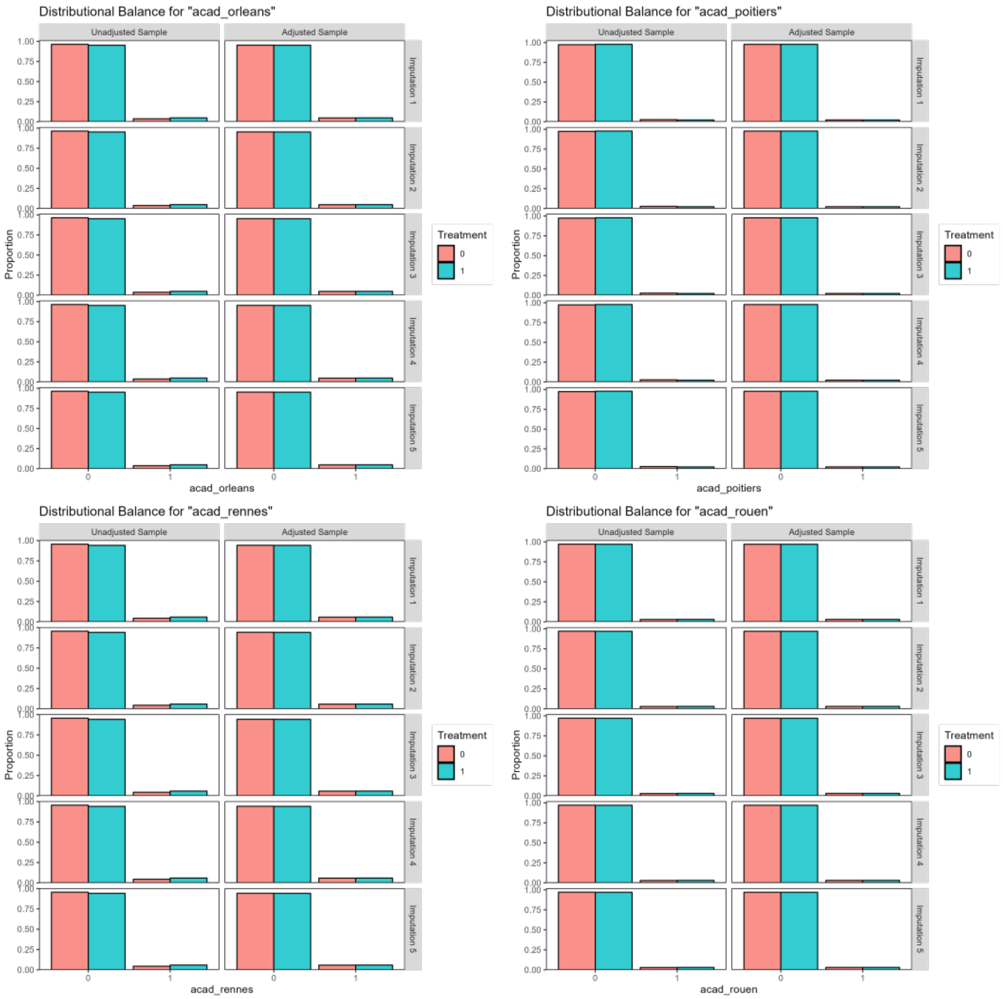

Figure 8: Balance measures on several covariates

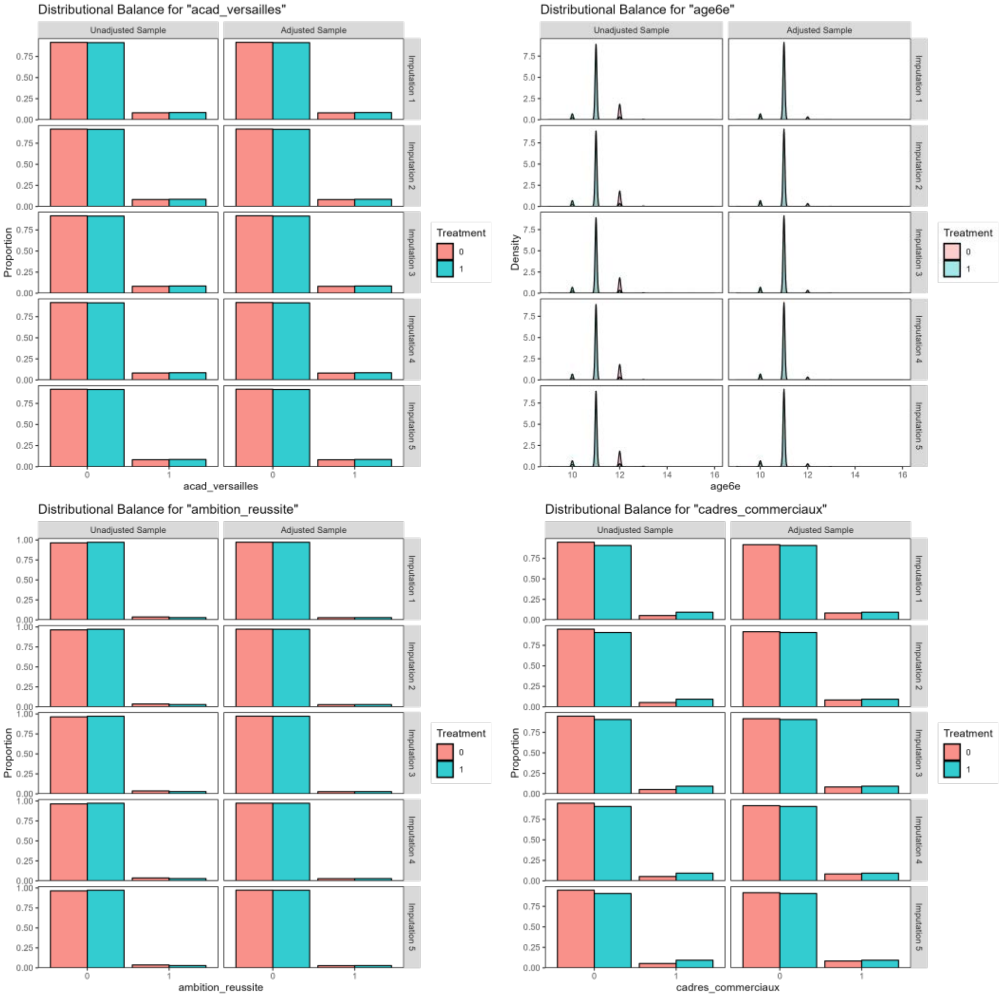

Figure 9: Balance measures on several covariates

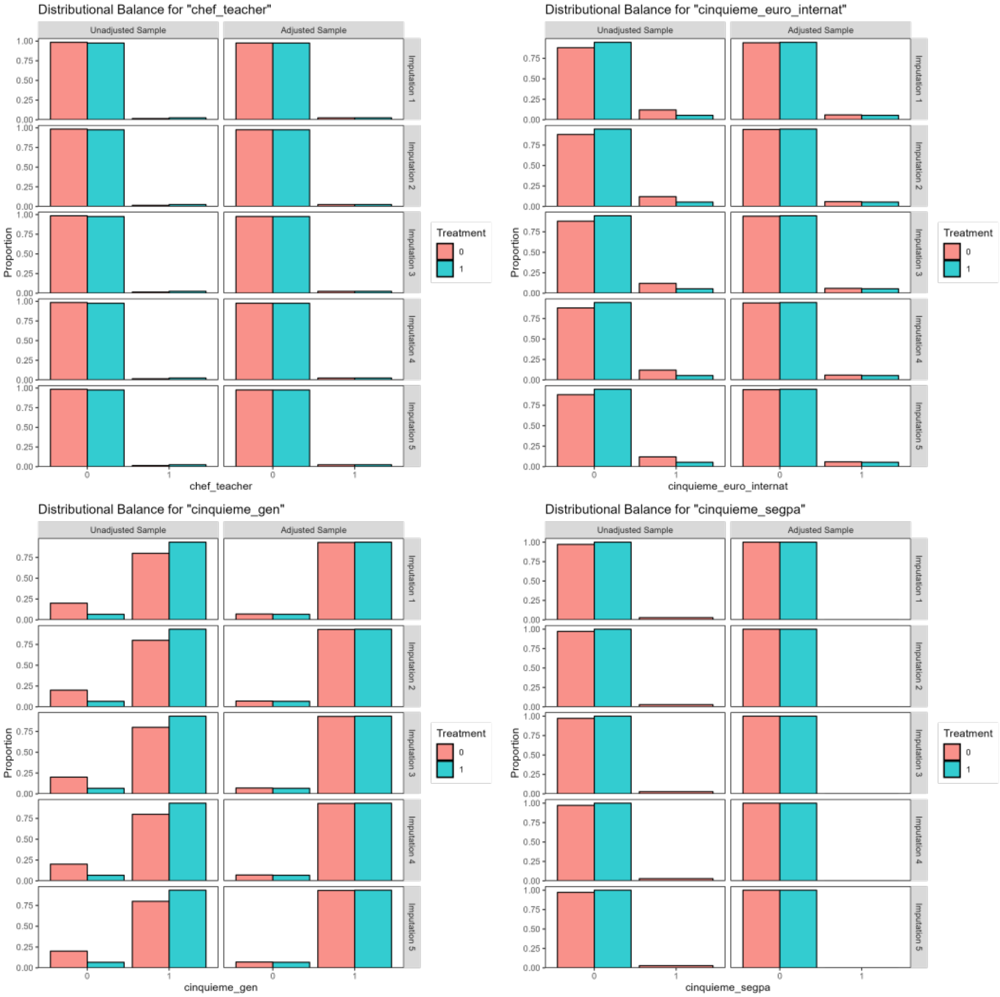

Figure 10: Balance measures on several covariates

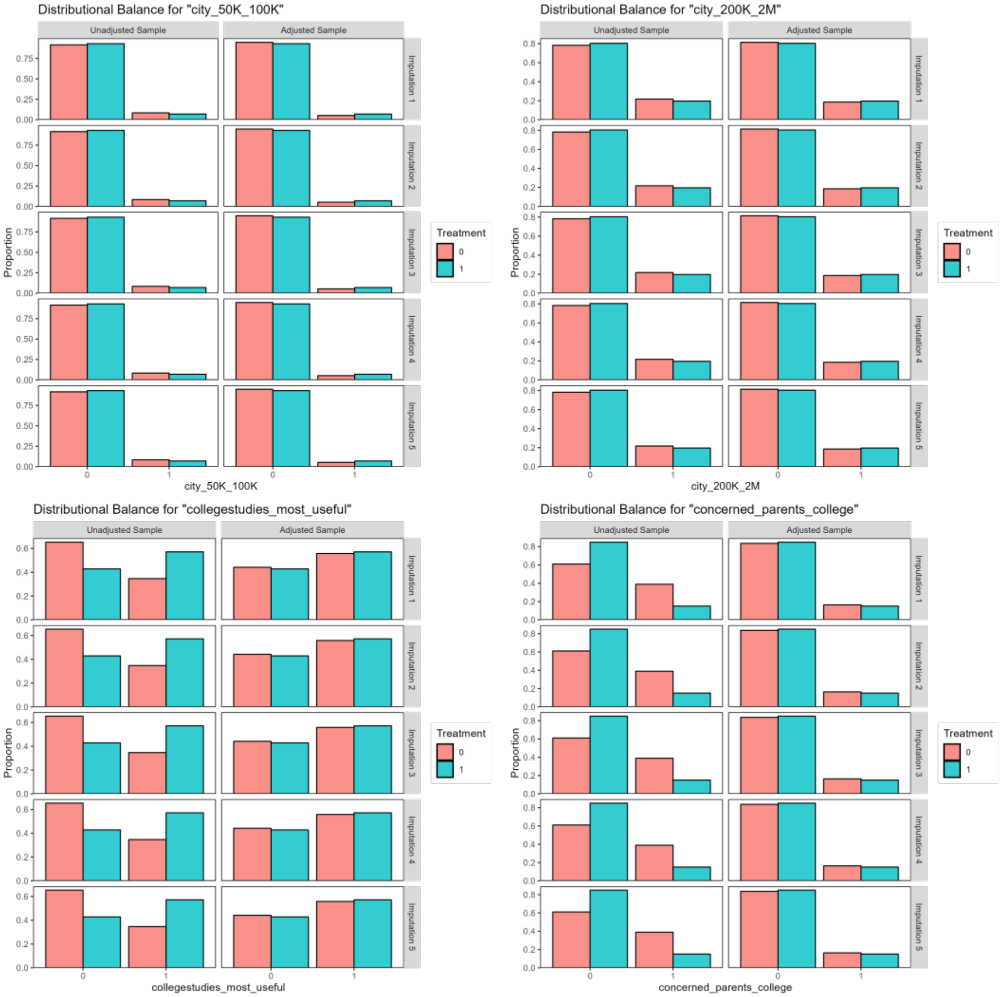

Figure 11: Balance measures on several covariates

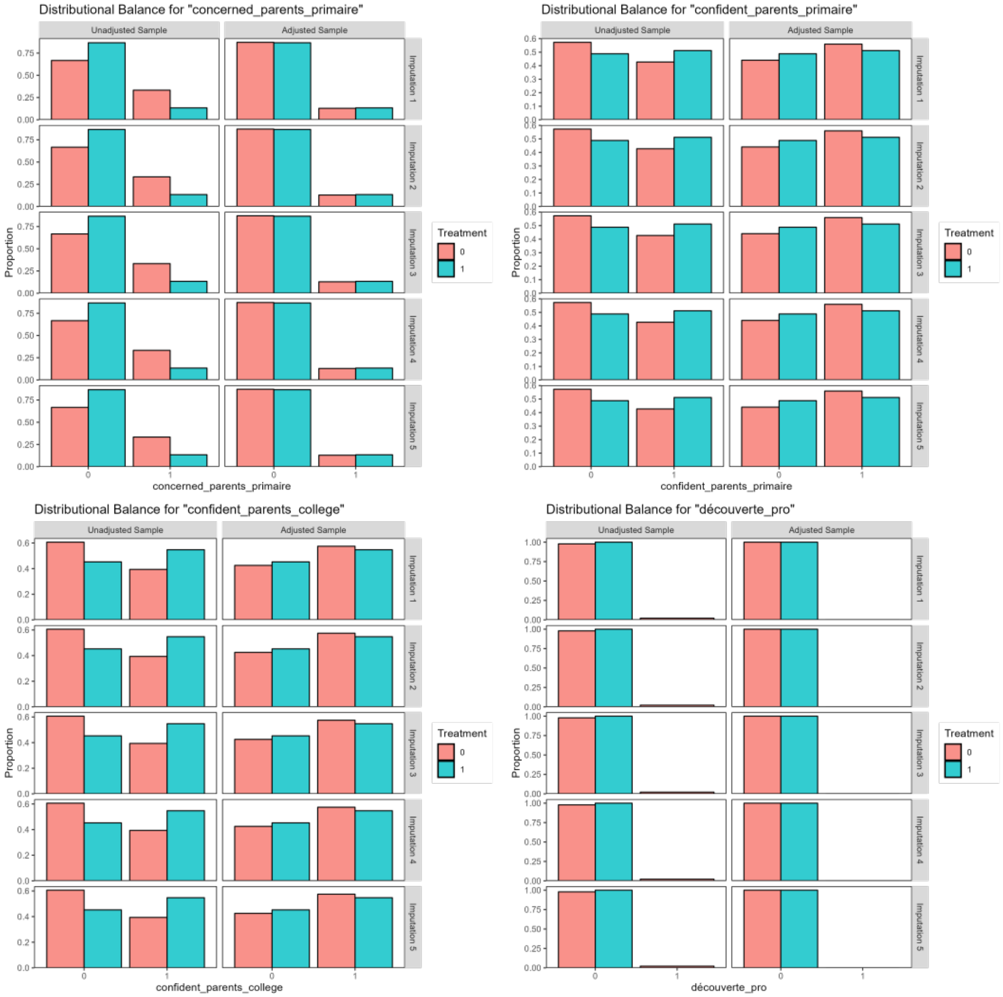

Figure 12: Balance measures on several covariates

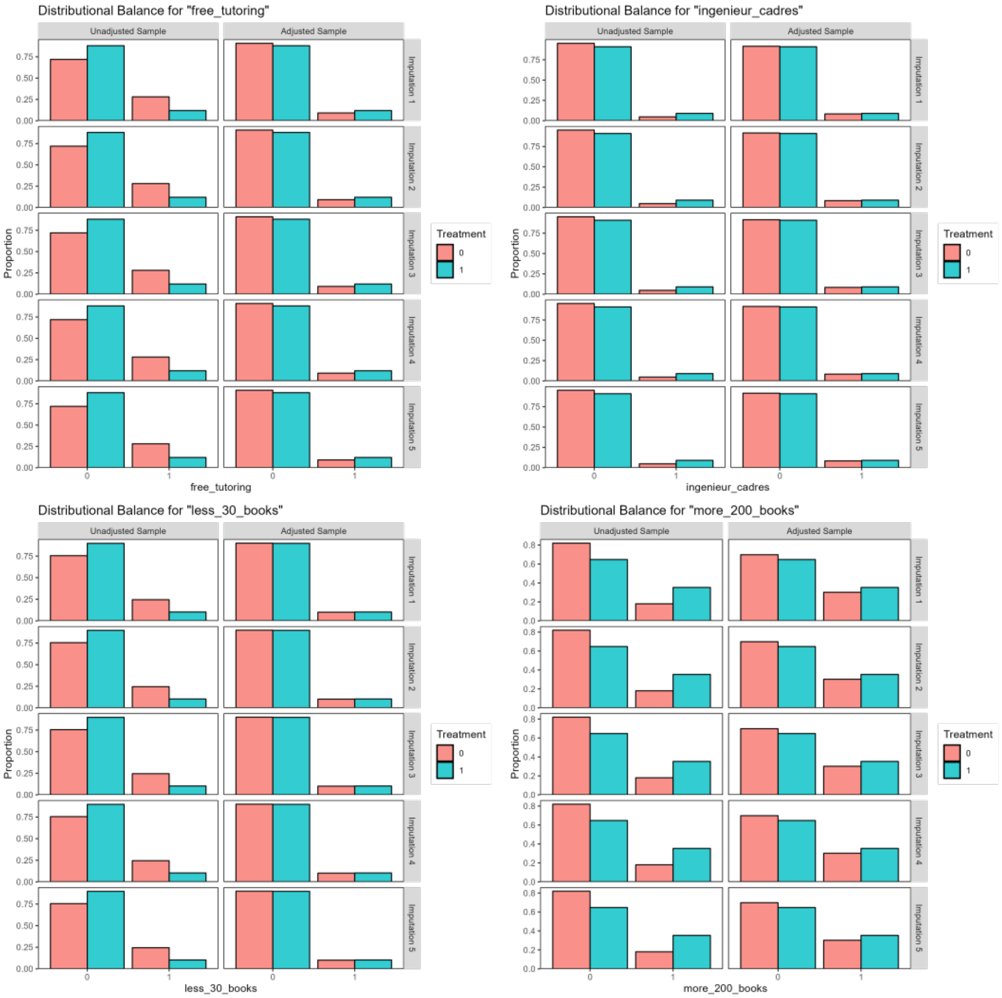

Figure 13: Balance measures on several covariates

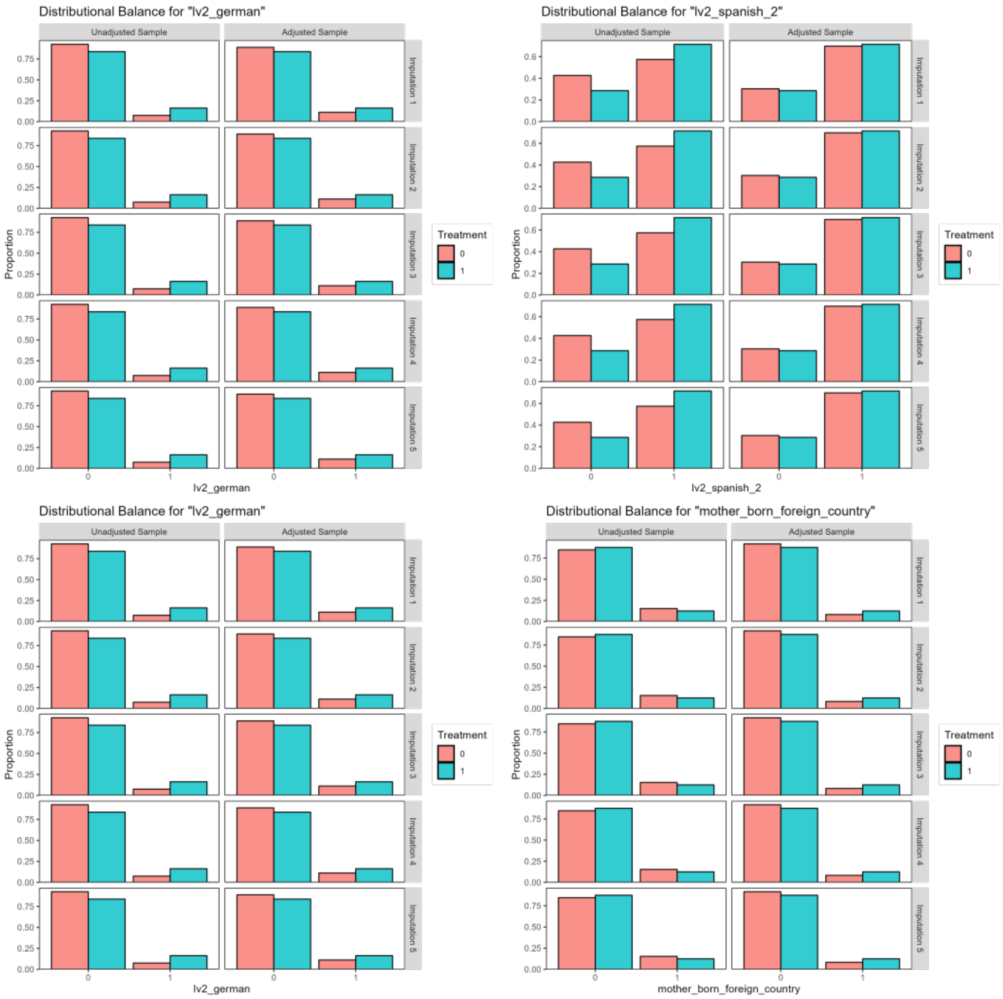

Figure 14: Balance measures on several covariates

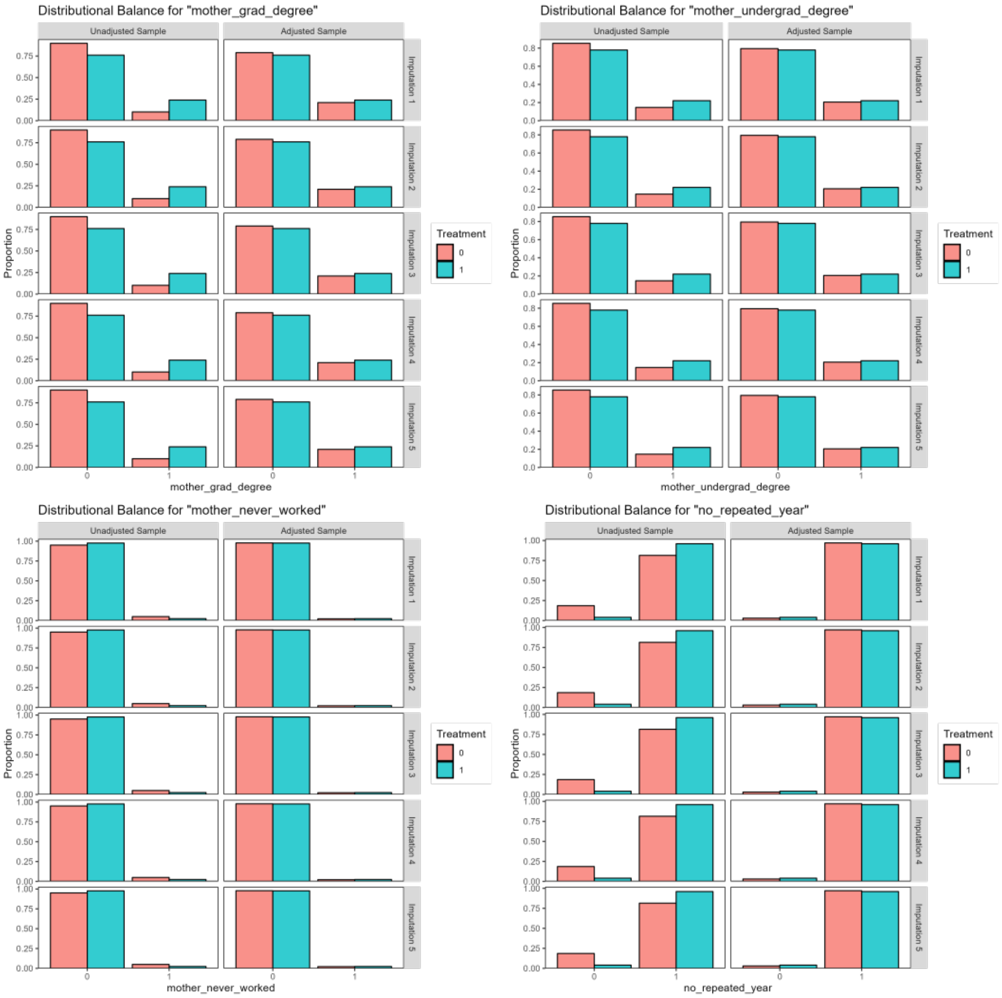

Figure 15: Balance measures on several covariates

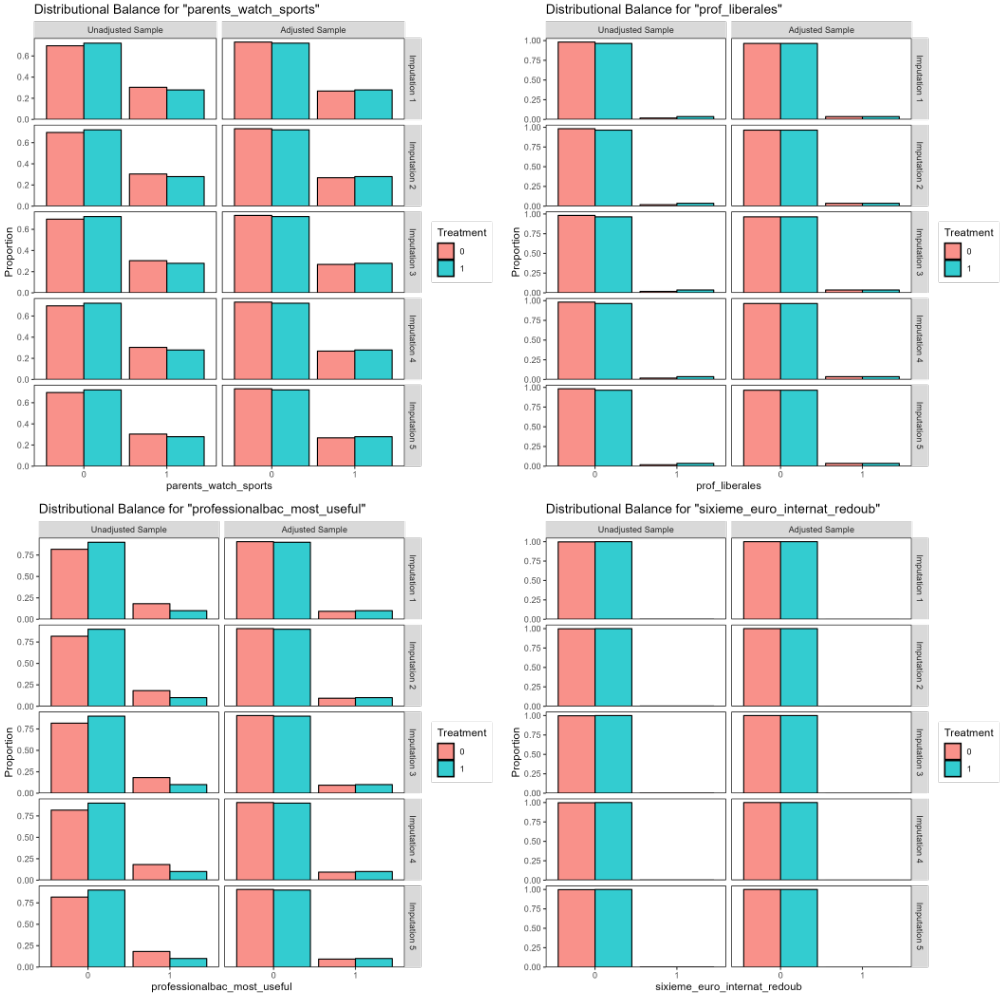

Figure 16: Balance measures on several covariates

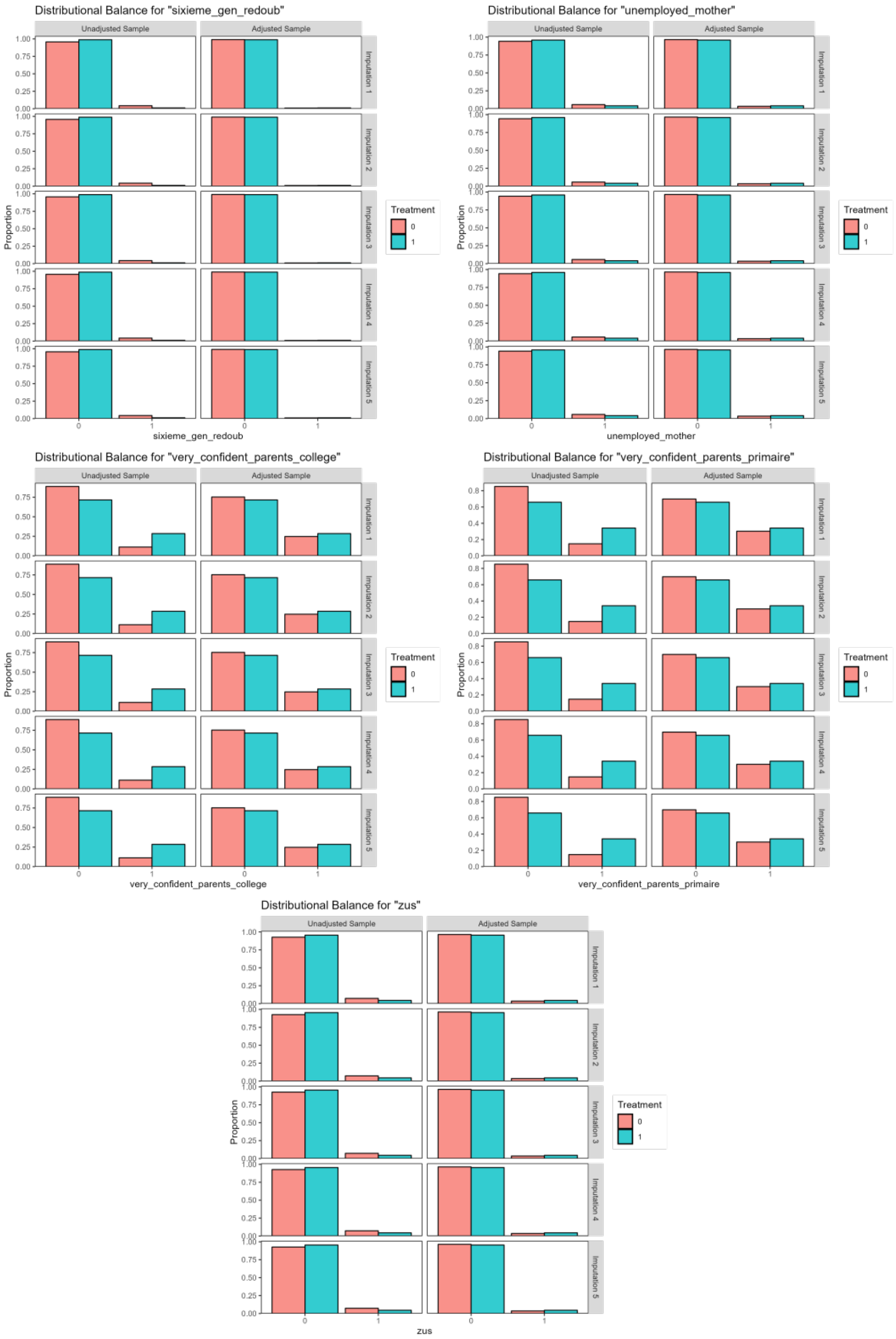

Figure 17: Balance measures on several covariates

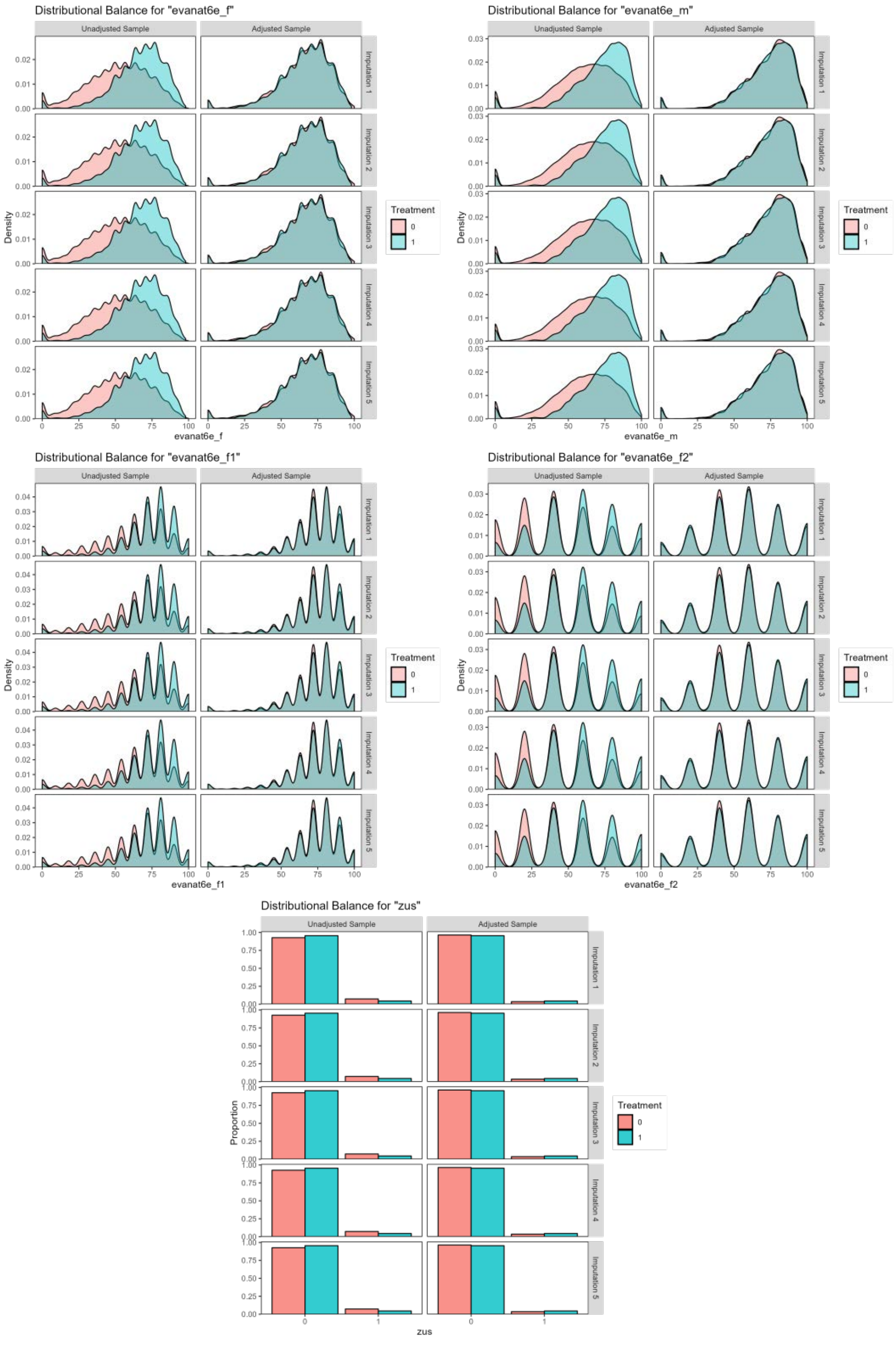

Figure 18: Balance measures on several covariates

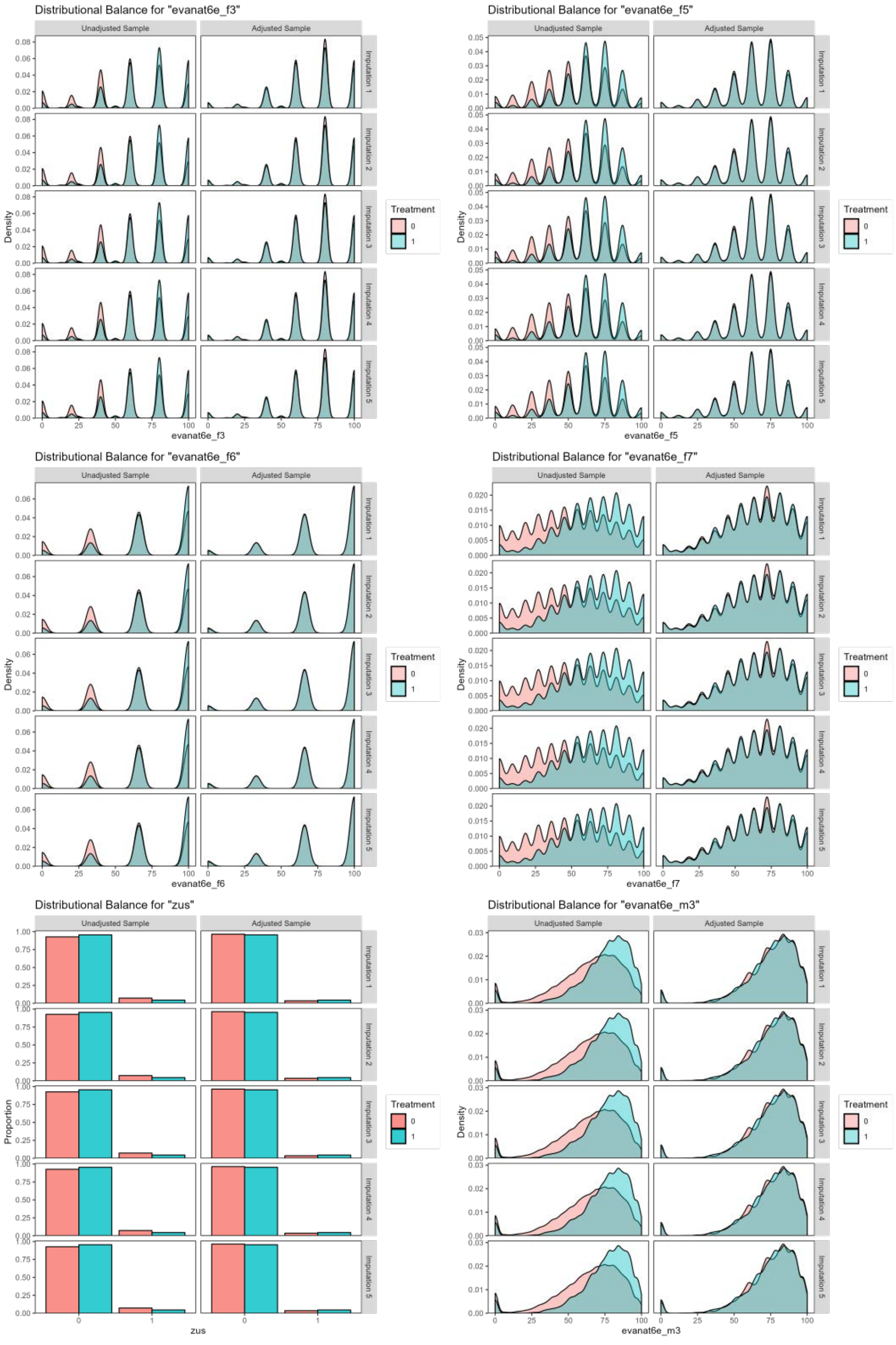

Figure 19: Balance measures on several covariates

Appendix 2 - Sample results of LASSO and logistic LASSO for model selection

Table 5: Results of LASSO to select the best predictors of test scores, first dataset

Selected	Lasso	Post-est OLS
sep_scol_eva2008	0.0310876	0.0656352
sep_social_e 2008	-0.0739613	-0.1027865
sep_autoregu 2008	0.2033553	0.2005588
evanat6e_f	0.0079252	0.0042606
evanat6e_f1	0.0030075	0.0038399
evanat6e_f2	0.0020017	0.0025649
evanat6e_f3	0.0010928	0.0020637
evanat6e_f5	0.0062906	0.0063488
evanat6e_f6	0.0014018	0.0019632
evanat6e_f7	0.0032331	0.0037112
evanat6e_m	0.0583098	0.0548425
evanat6e_m3	-0.0160704	-0.0162233
age6e	-0.3829720	-0.3376220
optob14		
24	-0.0356769	-0.1090170
43	-0.2599594	-0.4804062
tuetab1		
5	-0.0182337	-0.0770933
7	-0.0148976	-0.0518833
2.sex		
	-0.0199017	-0.1068420
acad1		
3	0.0302075	0.0830461
5	0.0765653	0.1037751
6	0.0676630	0.1169158
9	-0.0594870	-0.0514502
10	0.0437978	0.0726061
13	0.0353188	0.0380243
14	0.0576657	0.0558367
16	0.0033232	0.0346522
17	0.1481330	0.1554348
18	0.1697058	0.1843521
19	0.0030652	0.0276983
20	-0.0399770	-0.0684821
21	0.1096174	0.1331971
23	-0.0285229	-0.0745951
dep_resid		
62	-0.0156863	-0.0497336
93	-0.0519780	-0.1495320
2.rar1		
	0.0430267	0.1437177
2.a11m		
	-0.0583983	-0.0319294
3.a12m		
	-0.0429542	-0.0755014

a13m		
1	0.0316922	0.0399362
2	-0.0780837	-0.0503649
99.a17p		
	-0.0117329	-0.0839223
12.a20p		
	0.0132782	0.1181520
a20m		
11	0.0543207	0.0990478
12	0.1167639	0.1382991
3.a21m		
	-0.0470073	-0.1047838
a37		
1	-0.1915327	-0.1718022
4	0.0525718	0.1171657
3.a41a		
	-0.0588383	-0.1123075
2.b6		
	0.1029561	0.1565676
9.b7d		
	0.0830401	0.1778638
b12		
3	0.1578722	0.1933661
4	0.3241660	0.3664687
2.c10a		
	0.0048699	0.0073857
9.c10b		
	-0.0299735	-0.0667011
c18		
3	0.1898068	0.2010601
4	0.3575784	0.3675146
2.c25		
	0.0623795	0.1031209
3.c1		
	0.1820917	0.1892743
31.pcschef		
	0.0651569	0.0120847
pcspere		
37	0.0246032	0.0455881
38	0.1190058	0.0952419
61	-0.0378758	-0.0361564

2.rrs1	0.0122663	0.1248464
acad2		
2	-0.0896560	-0.1181480
11	-0.0392573	-0.0978020
15	-0.0039594	-0.0573875
24	-0.0607217	-0.0674666
25	-0.0349691	-0.0767562
28	-0.5629154	-0.5621609
31	-0.2377577	-0.2867155
32	-0.2617947	-0.3461830
clas2		
500	-0.5452380	-0.6724596
501	0.0183724	0.0621304
504	0.0942009	0.1149293
601	-0.2592813	-0.2622146
604	-0.3575935	-0.6411469
2.zus2		
9.nationaismereregr	-0.0455176	-0.0971490
2.nationaispereregr	-0.0427492	-0.1251592
8.natscoleleveregr	0.0059573	0.0745128

Table 6: Results of logistic LASSO to select the best predictor of taking Latin classes, first dataset

Selected	**Logistic Lasso**	**Post logit**
sep_scol_eva2008	0.1845941	0.3332949
evanat6e_f	0.0126543	0.0144855
evanat6e_m	0.0014065	0.0017156
optob13		
21	0.6959040	1.9091678
24	0.3000874	1.1604659
12.a20m	0.2330192	0.5221625
2.b12	-0.2876098	-0.6983586
2.c10a	0.2032026	0.7209618
c18		
2	-0.2670848	-0.4766533
4	0.2349952	0.3630196
c22		
3	-0.1597517	-0.7674072
6	0.1480102	0.2757718
42.pcschef	0.1395085	0.8876091
501.clas2	0.5264116	1.2219886
_cons	-3.5472902	-5.5605444

Appendix 3 - Summary of treatment effect estimate per dataset, main analysis

Table 7: Estimates per dataset - treatment

Term	**Contrast**	**Estimate**	**Std. Error**	**z**	**Pr(> $|z|$)**	**2.5%**	**97.5%**
latin	1 - 0	0.157	0.0142	11	<0.001	0.129	0.185
latin	1 - 0	0.163	0.0144	11.3	<0.001	0.135	0.191
latin	1 - 0	0.153	0.0143	10.7	<0.001	0.125	0.181
latin	1 - 0	0.158	0.0141	11.2	<0.001	0.13	0.186
latin	1 - 0	0.162	0.0143	11.3	<0.001	0.134	0.19

Table 8: Estimates per dataset - placebo

Term	**Contrast**	**Estimate**	**Std. Error**	**z**	**Pr(> $|z|$)**	**2.5%**	**97.5%**
latin	1 - 0	0.0788	0.0153	5.14	<0.001	0.0488	0.109
latin	1 - 0	0.0878	0.0162	5.41	<0.001	0.056	0.12
latin	1 - 0	0.0825	0.016	5.15	<0.001	0.0511	0.114
latin	1 - 0	0.0867	0.0163	5.32	<0.001	0.0548	0.119
latin	1 - 0	0.0955	0.0158	6.05	<0.001	0.0646	0.126