# Re-Estimating the Effect of Skill Specificity on Immigration Attitudes: A Double Machine Learning Approach

University of California, Los Angeles

Master of Arts in Quantitative Economics

Author: Nathan Kim

Faculty Advisor: Denis Chetverikov

December 2nd, 2022

## Table of Contents

## 1. Abstract

In this paper, I extend the research conducted in the publication titled "Skill Specificity and Attitudes toward Immigration" (Pardos-Prado & Xena, 2019). While other researchers have conjectured that labor supply may be the cause for anti-immigration sentiment, Pardos-Prado and Xena explore the effect of labor demand on immigration attitudes. The authors utilize traditional parametric and semiparametric econometric methods to estimate the causal effect of labor demand on immigration attitudes through various observational study designs. I will only focus on analyzing the results the authors obtain under the difference-in-differences (DiD) setting. In their original article, the authors study how skill specificity and labor demand jointly affect immigration attitudes which will be further discussed in the next section. To assess their theory, Pardos-Prado and Xena use a labor market reform in Germany as the treatment in their quasi-experiment. They claim that this reform would have helped unemployed individuals by allowing them to obtain reemployment more easily. The results from the authors' paper imply that the labor market reform in Germany lowered existing anti-immigration sentiment. To check the robustness of Pardos-Prado and Xena's (2019) results, I re-estimate the average treatment effect on the treated (ATT) using a double/debiased machine learning (DML) approach (Chernozhukov et al., 2018) which is more appropriate for the high dimensional data the authors use to compute the semiparametric difference-in-differences (SDiD) estimate (Abadie, 2005). I also consider even more highly dimensional data by using polynomial transformed covariates as well as additional covariates in the data set provided by the authors.

## 2. Introduction

In this section, I will give a brief overview of the question that Pardos-Prado and Xena (2019) attempt to answer in their original article. As discussed by the authors, earlier theories on the formation of immigration attitudes argue that opposition to immigration stems from labor supply. Since past studies either resulted in findings that conflict with these theories or were conducted using unrealistic assumptions, the authors propose that labor demand may form immigration attitudes instead. More specifically, the authors suggest skill specificity and labor demand as the joint cause for immigration attitudes.

Pardos-Prado and Xena (2019) use the works of Iversen and Soskice (2001) and Kambourov and Manovskii (2009) to define skill specificity "as high investments in occupation-specific human capital and low job transferability" (p. 287). According to Pardos-Prado and Xena (2019), skill specificity and labor demand relate to immigration attitudes as individuals in occupations that use very specific skills and are not as demanded by employers may lack the confidence in their ability to become reemployed. The authors further explain that since a future increase in net migration rates would lead to a more competitive labor market, these individuals would then be opposed to immigration in the present (Pardos-Prado & Xena, 2019). Therefore, using skill specificity and labor demand as the causal factors of immigration attitudes allow the authors to study the effect of what they call "potential competition" on immigration attitudes (Pardos-Prado & Xena, 2019, p. 289). This concept differs from the traditional theories that claim that anti-immigration sentiment is caused by a rise in foreign workers or "actual competition" (Pardos-Prado & Xena, 2019, p. 289). To test their newly proposed theory, the authors examine whether skill specificity and labor demand influence immigration attitudes under a DiD setting.

To carry out their quasi-experiment, Pardos-Prado and Xena (2019) use survey responses between 1999 and 2014 contained in the German Socio-Economic Panel (SOEP) (German Institute for Economic Research, 2022). The authors use a series of policy interventions called the Hartz reforms as the treatment for their study (Pardos-Prado & Xena, 2019). These reforms were put into effect across three consecutive years from 2003 to 2005. The authors explain that the reforms were enacted to combat the high unemployment rates in Germany by creating more favorable labor market conditions for unemployed individuals. Pardos-Prado and Xena (2019) state that they “interpret the Hartz reforms as a policy package incentivizing job demand, lowering legal and educational bars to access less regulated employment, and incentivizing transferability within and across occupations”, thereby countering the skill specificity present among the labor force and raising the demand for labor (p. 299). Since the policy intervention was meant to assist those that were unemployed, the authors designate individuals in the panel data that experienced unemployment sometime during 2003 to 2005 as the treatment group. All other individuals in the panel data are considered as the control group.  Therefore, individuals that were unemployed during that three-year period would have been exposed to the improved labor market conditions provided by the Hartz reforms which included higher labor demand and a buffer against skill specificity. In this way, the authors use individuals’ unemployment statuses to decide who was treated by the policy intervention. Based on the authors' new theory, individuals treated by the policy intervention should have become less opposed to immigration. To estimate the ATT, Pardos-Prado and Xena use both parametric and semiparametric econometric methods.

# 3. Traditional Difference-in-Differences Estimators

## 3.1. Linear Difference-in-Differences Estimator

To estimate the ATT, the authors first use the generalized DiD model since the German SOEP data set contains observations for more than two years and the Hartz reforms were enacted across three years (Pardos-Prado & Xena, 2019). Moreover, this model allows individuals to become treated and untreated throughout the years the policies are put in place. In fact, in the treatment group, some individuals are unemployed during the entirety of 2003 to 2005 while some are only unemployed in certain years within that period. The generalized DiD model has the following form:

$$Y_{it} = \alpha + \beta D_{it} + \sum_t \delta_t T_t + \sum_i \gamma_i E_i + u_{it}\,.$$

Based on the authors' paper and code, $Y_{it}$ represents immigration concern and has a range of values that include $1$, $2$, and $3$ which represent "Not concerned at all," "Somewhat concerned," and "Very concerned," respectively (Pardos-Prado & Xena, 2019, p. 295). $D_{it}$ is a dummy variable that equals $1$ if for a given year, the individual is unemployed, and the year is between 2003 and 2005 ($0$ otherwise). $T_t$ and $E_i$ are the year and individual fixed effects, respectively. Finally, $u_{it}$ is the error term. In this regression model, $\beta$ would represent the ATT.

A key condition of the DiD model is the so-called parallel trends assumption which stipulates that the changes in the dependent variable, immigration attitudes in this case, in both the treatment and control groups would have resulted in parallel trends had the treatment (policy intervention) never occurred (Abadie, 2005). However, Abadie (2005) states that systematic variations in the covariates of the treatment and control groups, which could be a result of selection bias, can cause the parallel trends assumption to not hold. In the case of the paper by Pardos-Prado and Xena (2019), I believe it is reasonable to assume that there are covariates that may make individuals more likely to be unemployed and utilize the policies of the Hartz reforms. To handle this issue, Abadie (2005) explains that one could add a vector of covariates as shown below:

$$Y_{it} = \alpha + \beta D_{it} + \sum_t \delta_t T_t + \sum_i \gamma_i E_i + \xi X_i + u_{it}.$$

In the above regression model, $X_i$ is a vector of covariates. Since Meyer (1995) argues that controlling for covariates in the DiD model as in above may result in model misspecification errors, Abadie (2005) presents the SDiD estimator. Pardos-Prado and Xena (2019) do not estimate the linear DiD model with covariates and control for covariates using the SDiD estimator instead.

## 3.2. Semiparametric Difference-in-Differences Estimator

Abadie (2005) uses the potential outcomes framework (Heckman, 1990; Rubin, 1974) to denote $Y^1(t)$and $Y^0(t)$ as the potential outcome at time $t$ had the individual been treated, and the potential outcome also at time $t$ had the individual not been treated, respectively. Abadie (2005) provides the following semiparametric estimator for the ATT:

$$\mathrm{E}[\mathrm{Y}^1(1) - \mathrm{Y}^0(1)|\mathrm{D} = 1] = \mathrm{E}\left[\frac{\mathrm{Y}(1) - \mathrm{Y}(0)}{\mathrm{P}(\mathrm{D} = 1)} \cdot \frac{\mathrm{D} - \mathrm{P}(\mathrm{D} = 1|\mathrm{X})}{1 - \mathrm{P}(\mathrm{D} = 1|\mathrm{X})}\right]$$

where $Y^1(1)$ is the potential outcome after the treatment takes place had the individual been treated, $Y^0(1)$ is the potential outcome after the treatment takes place had the individual not been treated, $Y(1)$ is the observed outcome of the individual after the treatment takes place, $Y(0)$ is the observed outcome of the individual before the treatment takes place, $D$ is a dummy variable which equals $1$ if the individual is in the treatment group, and $X$ is a vector of covariates. Abadie (2005) weighs the change in the dependent variables between the after-treatment and before-treatment periods using the propensity score, $P(D = 1|X)$, which rectifies the scenario where a disproportion in the covariates of the treatment and control groups may cause the parallel trends assumption to be untrue. The propensity score is traditionally estimated using classical nonparametric estimators such as kernel estimators. Pardos-Prado and Xena (2019) calculate their SDiD estimate through the STATA function **absdid**(), developed by Houngbedji (2016), which estimates the propensity score using the linear probability model. The covariates that the authors use include income levels by quintiles, a dummy variable for manual work, gender, age, a dummy variable for having a temporary work contract, skill specificity, and labor demand (Pardos-Prado & Xena, 2019). These variables will be further discussed in Section 5. For more detailed proofs and derivations regarding Abadie's (2005) SDiD estimator, please refer to the paper cited above.

### 3.3. Parallel Trends in the German Socio-Economic Panel

Figure 1 is a slightly altered replication of Figure C9 in the author's supporting information document, and it shows average immigration attitudes from 1999 to 2003 (Pardos-Prado & Xena, 2019). In Figure 1, I use the unmodified immigration concerns variable while Pardos-Prado and Xena (2019) plot average immigration attitudes as a binary variable in Figure C9. Without conditioning on any covariates, it seems that the parallel trends assumption somewhat holds for the German SOEP data. Nevertheless, the authors still use the more robust SDiD estimator, and in Section 8, I will use an even more robust machine learning (ML) based estimator discussed in the next section. Moreover, in Figure 1, the individuals who were unemployed have higher immigration concerns compared to those that were employed.

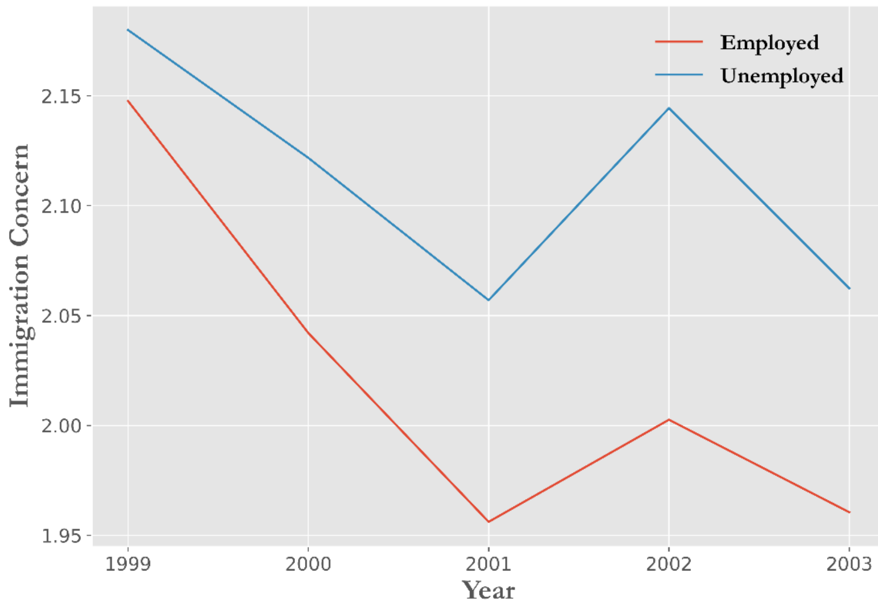


**Figure 1.** Parallel Trends Assumption, Modified from Figure C9 in the Supporting Information (Pardos-Prado & Xena, 2019)

# 4. Double Machine Learning Estimators

## 4.1. DMLDiD Estimator

One concern of implementing Abadie's (2005) SDiD estimator is that classical nonparametric estimators used to estimate the propensity scores are subject to the curse of dimensionality when the number of right-hand side variables start to grow. Therefore, in such a case, ML algorithms which utilize regularization techniques may be a better alternative (Chernozhukov et al., 2018). Despite the robustness of ML estimators, using them to estimate the propensity score in Abadie's (2005) SDiD estimator will likely result in biased estimates due to regularization bias and overfitting (Chernozhukov et al., 2018). To resolve this issue, Chang (2020) builds upon the work of DML by Chernozhukov et al. (2018) and modifies Abadie's (2005) SDiD estimator to accommodate appropriate usage of ML algorithms which he calls the DMLDiD estimator. Chang (2020) presents Abadie's (2005) SDiD estimator as such:

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i(1) - Y_i(0)}{\hat{p}}\frac{D_i - \hat{g}(X_i)}{1 - \hat{g}(X_i)}.$$

Chang (2020) designates $\hat{p}$ and $\hat{g}(X_i)$ as the sample estimates of $p_0 \equiv P(D = 1)$ and $g_0(X) \equiv P(D = 1|X)$ for each $ith$ individual, respectively. Chang modifies the SDiD estimator above in the following way:

$$\tilde{\theta}_k = \frac{1}{n}\sum_{i \in I_k}\frac{D_i - \hat{g}_k(X_i)}{\hat{p}_k\big(1 - \hat{g}_k(X_i)\big)} \times \Big(Y_i(1) - Y_i(0) - \hat{\ell}_{1k}(X_i)\Big).$$

Chang designates $I_k$ as the $kth$ fold of a total of $K$ random folds of the entire data set and $n$ as the number of observations in the $kth$ fold. In each of the respective subsamples where one of the $K$ folds are held out, $\hat{p}_k$, $\hat{g}_k(X_i)$, and $\hat{\ell}_{1k}(X_i)$ are the sample estimates of $p_0 \equiv P(D = 1)$, $g_0(X) \equiv P(D = 1|X)$, and $\ell_{10} \equiv E[Y(1) - Y(0)|X, D = 0]$ for each $ith$ individual, respectively.

## 4.2. Neyman Orthogonality Condition

Chang (2020) defines the score function of the DMLDiD estimator mentioned in the previous subsection as the following:

$$\begin{aligned}\psi_1(W, \theta_0, p_0, \eta_{10}) = {} & \frac{Y(1) - Y(0)}{P(D = 1)}\frac{D - P(D = 1|X)}{1 - P(D = 1|X)} - \theta_0 \\ & - \frac{D - P(D = 1|X)}{P(D = 1)\big(1 - P(D = 1|X)\big)}\mathrm{E}[Y(1) - Y(0)|X, D = 0].\end{aligned}$$

In the above score function, $W = (Y(1), Y(0), D, X)$, $\theta_0 \equiv E[Y^1(1) - Y^0(1)|D = 1]$, and $\eta_{10} = (g_0, \ell_{10})$. Chang proves that the above score function meets the Neyman orthogonality condition as expressed below:

$$\partial_{\eta_1}\ \mathrm{E_p}[\psi_1(W, \theta_0, p_0, \eta_{10})] = 0$$

The expression above is found in the appendix of Chang's (2020) paper, and it proves that the Gateaux derivative of the score function, $\psi_1(W, \theta_0, p_0, \eta_{10})$, with respect to $\eta_{10}$ is zero. Therefore, Chang demonstrates the unbiasedness of the DMLDiD estimator through the Neyman orthogonality condition. Moreover, he proves that it is possible for the DMLDiD estimator to be both consistent and asymptotically normal. For more detailed proofs and derivations regarding DML estimators, please refer to the papers by Chernozhukov et al. (2018) and Chang (2020) cited above.

## 4.3. Application to the German Socio-Economic Panel

In Section 8, I will apply the DMLDiD estimator to the German SOEP data and compare the results by Pardos-Prado and Xena (2019). Before doing so, I will first discuss some adjustments to variable definitions. When estimating the parametric DiD model, the authors were able to use multiple time periods provided in the panel data using the generalized DiD model (Pardos-Prado & Xena, 2019). Since the DMLDiD estimator is based on two time periods (Chang, 2020), I redefine the treatment/policy variable, $D_{it}$, to be equal to $1$ if the individual experienced unemployment at least once throughout the years 2003 to 2005 and $0$ otherwise. In addition, I set 2002 as the before-treatment period and 2011 as the after-treatment period. This means that $Y_i(1)$ and $Y_i(0)$ in the DMLDiD estimator represent immigration attitudes for each $ith$ individual in 2002 and 2011, respectively. Therefore, this setup is in alignment with the traditional two-time period DiD setting the DMLDiD estimator is based on (Chang, 2020). I will use the same covariates that Pardos-Prado and Xena (2019) use for the SDiD estimator for the DMLDiD estimator. I will also consider polynomial transformations of these covariates and additional covariates in the German SOEP data that Pardos-Prado and Xena (2019) did not use for the SDiD estimation. Given that the vector of covariates, $X_i$, is highly dimensional ($p = 8$), I can justify the use of the DMLDiD estimator as classical nonparametric estimators would suffer from the curse of dimensionality.

The double ML aspect of the DMLDiD estimator requires using ML algorithms to estimate $\hat{g}_k(X_i)$ and $\hat{\ell}_{1k}(X_i)$ (Chang, 2020). Since $\hat{g}_k(X_i)$ is an estimator for $\mathrm{P}(D = 1|X)$, I will apply supervised classification algorithms to this component. By the same logic, since $\hat{\ell}_{1k}(X_i)$ is an estimator for $E[Y(1) - Y(0)|X, D = 0]$, I will apply supervised regression algorithms to this component.

## 5. Covariates and Feature Importance

### 5.1. Skill Specificity and Labor Demand

To estimate the propensity scores, Pardos-Prado and Xena (2019) use the vector of covariates, $X_i$, described in Subsection 3.2. Income levels, gender, manual work, age, and being temporarily employed are self-explanatory covariates commonly used in research in the social sciences. Iversen and Soskice (2001) define skill specificity using the International Standard Classification of Occupations (ISCO-88) created by the International Labour Organization (2004). According to the International Labour Organization (2004), the ISCO-88 categorizes occupations in varying amounts of granularity. Iversen and Soskice (2001) calculate skill specificity "by comparing the share of unit groups in any higher level class to the share of the workforce in that class" (p. 881). Pardos-Prado and Xena (2019) also calculate skill specificity this way by using the unit groups and major groups in the ISCO-88 as defined by the International Labour Organization (2004). In their paper, Pardos-Prado and Xena (2019) refer to this variable as the "compartmentalization of occupational tasks within major occupational groups," but I will call it task compartmentalization for simplicity (p. 290). Rehm (2009) initially used occupational unemployment rates to examine whether unfavorable job outlooks can garner more approval for income redistribution. Pardos-Prado and Xena (2019) use occupational unemployment rates to represent labor demand. Therefore, it may be reasonable to assume that individuals who are in occupations that use very specific skills and/or have frequent layoffs are more prone to self-selection into the treatment group.

### 5.2. Feature Importance Using Machine Learning

To assess Pardos-Prado and Xena's (2019) choice of covariates for estimating the propensity scores, I will utilize commonly used ML algorithms for feature selection/importance which includes Random Forest and the least absolute shrinkage and selection operator (LASSO). Random Forest

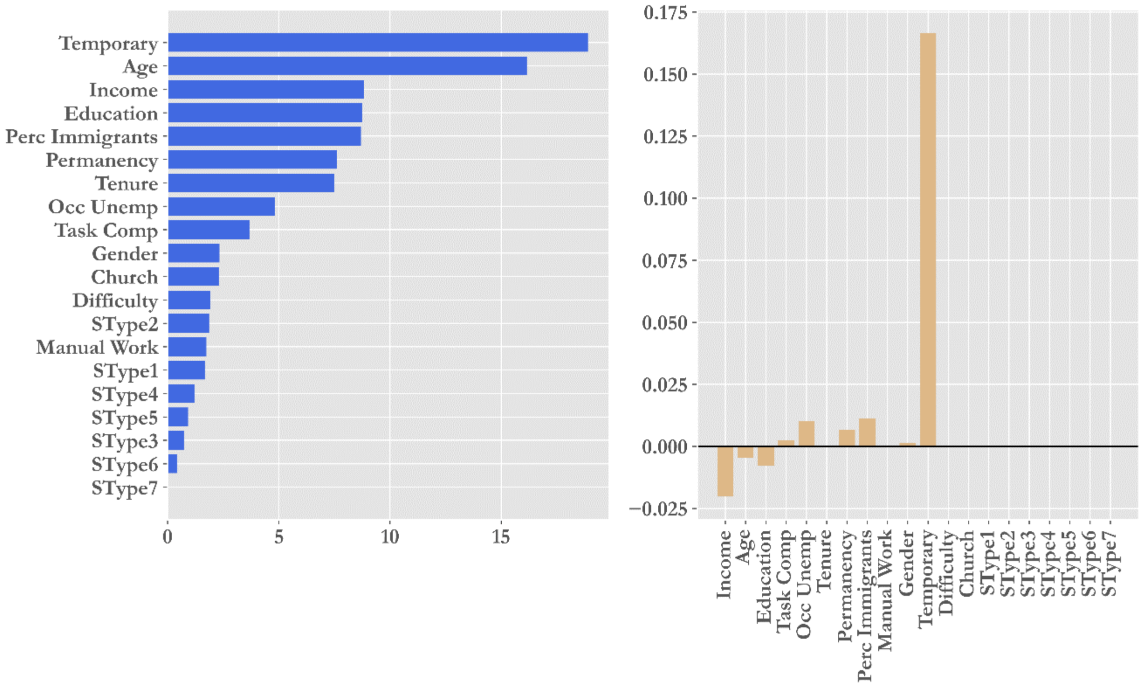


**Figure 2.** Feature Importance Using Random Forest (left) and LASSO (right)

can give information regarding feature importance based on which features the algorithm chooses to make splits in the data. Random Forest chooses features that will minimize the loss function. LASSO can perform feature selection by adding the $\ell_1$ penalty parameter in the loss function which biases coefficient estimates to zero. I fit Random Forest and LASSO using the covariates as the features and the treatment dummy variable as the dependent variable. I tune the hyperparameters using 3-fold cross validation.

There were some covariates available in the data set which Pardos-Prado and Xena (2019) did not consider for the SDiD estimation, so I will include them as well. These covariates are occupational tenure, "occupational permanency rates," "occupational migrant share," "perceived difficulty of finding a new job," and "church attendance" (Pardos-Prado & Xena, 2019, pp. 293, 295, 297, 300). I also consider individuals' "type of schooling" which is in the codebook provided by the authors (Pardos-Prado & Xena, 2019, p. 5). Occupational tenure is another representation of skill specificity which Kambourov and Manovskii (2009), Sullivan (2010), and Zangelidis (2008) use to study wage distributions. Pardos-Pardo and Xena (2019) also use occupational tenure to represent

**Table 1.** AbadieML Estimates ($p = 8$)

| | DiD | SDiD | Logit-$\ell_1$ | FNN | RF | GB | AdaB | XGB |
|---|---|---|---|---|---|---|---|---|
| ATT | $-0.019$ | $-0.384$ | $-0.013$ | $-0.013$ | $-0.023$ | $-0.012$ | $-0.005$ | $-0.012$ |
| SE | $(0.010)$ | $(0.042)$ | $(0.054)$ | $(0.052)$ | $(0.054)$ | $(0.054)$ | $(0.058)$ | $(0.053)$ |

skill specificity, and they define it as the amount of time that an individual has been employed in their most recent occupation. Occupational permanency rates are another representation of labor demand, and they are the frequencies at which individuals shift to different occupations (Shaw, 1987, as cited in Pardos-Prado & Xena, 2019). Those that have prolonged occupational tenure and/or shifted to different occupations fewer times may also be more predisposed to self-select into the treatment group. Occupational migrant share is defined as the percentage of immigrants within the occupation the individual is in (Pardos-Prado & Xena, 2019). The rest of the covariates are self-explanatory.

In Figure 2, both Random Forest and LASSO choose having a temporary work contract as the most important feature. There is also some overlap between the other covariates that Pardos-Prado and Xena (2019) chose (income, manual work, gender, education, age, task compartmentalization, and occupational unemployment) and what Random Forest/LASSO chooses. Some notable features that the authors did not choose include occupational migrant share, occupational permanency rates, and occupational tenure.

## 6. Machine Learning Algorithms

To obtain the DMLDiD estimates, I utilize the following popular ML algorithms: $\ell_1$-penalized Logistic Regression, LASSO, $\ell_1$-penalized feed-forward neural networks (FNN), Random Forest (RF), Gradient Boosting (GB), AdaBoost (AdaB), and XGBoost (XGB). These ML methods can be helpful when using highly dimensional data because they utilize some form of regularization or dimensionality reduction techniques.

**Table 2.** DMLDiD Estimates ($p = 8$)

| | DiD | SDiD | LASSO | FNN | RF | GB | AdaB | XGB |
|---|---|---|---|---|---|---|---|---|
| ATT | $-0.019$ | $-0.384$ | $-0.019$ | $-0.016$ | $-0.013$ | $-0.019$ | $-0.014$ | $-0.013$ |
| SE | $(0.010)$ | $(0.042)$ | $(0.046)$ | $(0.044)$ | $(0.045)$ | $(0.045)$ | $(0.045)$ | $(0.045)$ |

ML algorithms that use regularization techniques are very powerful because they can decide which covariates/features are the most relevant (Chernozhukov et al., 2018). Both the $\ell_1$-penalized Logistic Regression and LASSO use the $\ell_1$ penalty to send coefficient estimates to zero. Feed-forward neural networks can also be designed to apply $\ell_1$ penalties to the weights in each layer. Moreover, neural networks with decreasing numbers of neurons in each subsequent layer is a form of dimensionality reduction. Neural networks can also be very useful for fitting non-linear relationships within the data. Random Forests are a good choice for high dimensional data since the algorithm decides which features/covariates to use for splitting the data. Random Forests can also be regularized by tuning the maximum tree depth parameter. Boosting methods also utilize regularization through the learning rate. XGBoost, which is a form of boosting, allows for the use of both the $\ell_1$ and $\ell_2$ penalty parameters. Since these algorithms will likely result in regularization bias, using the DMLDiD estimator to estimate the ATT will take care of this issue.

## 7. Semiparametric Difference-in-Differences Results

In their respective papers, Chang (2020) and Chernozhukov et al. (2018) show that using ML algorithms in traditional econometric estimation procedures without taking any intermediary steps can lead to regularization bias. For completeness, before applying the DMLDiD estimator, I will use ML algorithms to obtain SDiD estimates to compare them to the DMLDiD estimates. To estimate the ATT using the SDiD estimator, I will estimate $g_0(X) \equiv P(D = 1|X)$ using supervised classification algorithms. Table 1 summarizes the SDiD results calculated by using ML algorithms. In addition, I will borrow notation from the replication file in Chang's (2020) paper and refer to the results in Table 1 as the AbadieML results.

**Table 3.** DMLDiD Estimates with 4th Order Polynomial Covariates ($p = 494$)

| | LASSO | FNN | RF | GB | AdaB | XGB |
|---|---|---|---|---|---|---|
| ATT | $-0.025$ | $-0.020$ | $-0.026$ | $-0.021$ | $-0.012$ | $-0.024$ |
| SE | $(0.045)$ | $(0.044)$ | $(0.045)$ | $(0.045)$ | $(0.046)$ | $(0.045)$ |

When using ML algorithms to estimate the propensity scores for the SDiD estimator in Table 1, I obtain much smaller estimates for the ATT compared to the authors' SDiD estimate, and the standard errors are slightly larger. These results are computed for comparison against the DMLDiD estimates and should not be used for inference due to potential regularization bias since the SDiD estimator does not meet the Neyman orthogonality condition (Chang, 2020; Chernozhukov et al., 2018). Additionally, the AbadieML point estimates are all statistically insignificant at the 5% level.

# 8. Double Machine Learning Results

## 8.1. DMLDiD Estimation Results

For the DMLDiD estimator labeled as LASSO, I apply the $\ell_1$-penalized logistic regression to $\hat{g}_k(X_i)$ and LASSO to $\hat{\ell}_{1k}(X_i)$. All other estimates were computed using the classification and regression counterparts of the specified algorithm accordingly. The parametric DiD and SDiD estimates were calculated by Pardos-Prado and Xena (2019) in their original paper. All hyperparameters were tuned using 2-fold cross validation. I use 2-fold cross-fitting as discussed in Chang (2020) and Chernozhukov et al. (2018) when using the DMLDiD estimator to estimate the ATT.

The results in Table 2 can be interpreted as the average change in immigration attitudes for those treated by the Hartz reforms. As shown in Table 2, the DMLDiD estimates are much smaller in absolute value (approximately 30 times smaller at the most) compared to the SDiD estimate and are much closer to the linear DiD estimate, especially the estimates using LASSO and Gradient Boosting. I also notice that the three boosting methods show different results indicating that using adapted weights in AdaBoost and both $\ell_1$ and $\ell_2$ regularization in XGBoost most likely provided additional

**Table 4.** DMLDiD Estimates with Additional Covariates ($p = 19$)

| | LASSO | FNN | RF | GB | AdaB | XGB |
|---|---|---|---|---|---|---|
| ATT | $-0.032$ | $-0.018$ | $-0.043$ | $-0.044$ | $-0.030$ | $-0.041$ |
| SE | (0.044) | (0.044) | (0.041) | (0.041) | (0.042) | (0.041) |

learning from the data. Moreover, it appears that using both the parametric and semiparametric DiD estimators are resulting in statistically significant estimates while the more robust DMLDiD estimates fail to reach statistical significance at the 5% level. Pardos-Prado and Xena's (2019) DiD and SDiD estimates imply that the Hartz reforms lowered existing anti-immigration sentiment in Germany, and this evidence supports their theory on immigration attitudes. However, I cannot make any conclusions based on the DMLDiD results. Due to the statistical insignificance of the DMLDiD estimates, further investigation on the effect of skill specificity and labor demand on immigration attitudes may be required. When comparing the DMLDiD results in Table 2 to the AbadieML results in Table 1, the DMLDiD point estimates appear to be more stable while the AbadieML point estimates vary more. This observation could be due to regularization bias from using ML algorithms in the SDiD estimation procedure (Chang, 2020; Chernozhukov et al., 2018). The standard errors of the DMLDiD estimates are also smaller than the standard errors of the AbadieML estimates.

## 8.2. Polynomial Transformed Covariates

To explore the different possible relationships between the dependent variable and the covariates, I also compute the DMLDiD estimates using 4th order polynomial transformed covariates. Applying this transformation changes the dimension of the vector of covariates, $X_i$, from $p = 8$ to $p = 494$. The estimates are summarized in Table 3. Except for the DMLDiD estimate using AdaBoost, which resulted in a slightly smaller ATT, all other estimates resulted in larger ATTs (in absolute value) compared to the estimates in Table 2. Other than the estimate using AdaBoost, the magnitudes of the DMLDiD estimates in Table 3 are relatively close to each other. These results indicate that there may be more dynamic relationships between the covariates and immigration attitudes. However, because all of the point estimates in Table 3 are still statistically insignificant at

**Table 5.** DMLDiD Estimates with 2nd Order Polynomial Covariates ($p = 209$)

| | LASSO | FNN | RF | GB | AdaB | XGB |
|---|---|---|---|---|---|---|
| ATT | $-0.039$ | $-0.045$ | $-0.044$ | $-0.041$ | $-0.027$ | $-0.037$ |
| SE | $(0.045)$ | $(0.042)$ | $(0.041)$ | $(0.042)$ | $(0.042)$ | $(0.042)$ |

the 5% level, I cannot make any concrete conclusions. Additionally, the standard errors in Table 3 are similar to the standard errors in Table 2.

## 8.3. Additional Covariates

I also compute DMLDiD estimates using the original eight covariates (Pardos-Prado & Xena, 2019) used in Table 2 and all of the additional covariates mentioned in Subsection 5.2 (occupational tenure, occupational permanency rates, occupational migrant share, perceived difficulty of finding a new job, church attendance, and schooling type). These results are summarized in Table 4. Adding these additional covariates changes the dimension of the vector of covariates from $p = 8$ to $p = 19$. Instead of using just the covariates the authors use, I will let the ML algorithms and cross-validation/hyperparameter tuning decide which covariates are the most relevant. Compared to the DMLDiD results in Table 2, the results in Table 4 have point estimates that are all larger in absolute value and standard errors that are either similar or smaller. When comparing the results from Table 4 to Table 3, the DMLDiD estimate using neural networks is slightly smaller in absolute value while all other estimates are larger in absolute value. The estimates in Table 4 may indicate that the additional covariates are relevant for estimating the ATT, but they continue to fail to reach statistical significance at the 5% level. Therefore, just as with the previous DMLDiD estimates in Tables 2 and 3, I cannot make any conclusions regarding the differences or similarities in values between the DMLDiD estimates in Table 4 and Pardos-Prado and Xena's (2019) results in their original paper.

## 8.4. Additional Polynomial Transformed Covariates

Continuing from the previous subsection, I also estimate the ATT using the polynomial transformation of the additional covariates used above. Since there are now 19 covariates, I only

Table 6. Hyperparameter Choices

| LASSO | FNN | RF | GB | AdaB | XGB |
|---|---|---|---|---|---|
| $\ell_1$ Penalty | $\ell_1$ Penalty | Max Depth | Max Depth | Max Depth | $\ell_1$ Penalty |
| | # of Layers | Max Features | # of Rounds | # of Rounds | $\ell_2$ Penalty |
| | Optimizers | | Max Features | Learning Rate | Max Depth |
| | # of Neurons | | Learning Rate | | # of Rounds |
| | Learning Rate | | | | Learning Rate |

consider the $2nd$ order polynomial transformation due to computational constraints. This changes the dimension of the vector of covariates from $p = 19$ to $p = 209$. If I were to take the 3rd or 4th order polynomial transformations, the dimensions would increase to $p = 1539$ and $p = 8854$, respectively, and hyperparameter tuning would be infeasible for some ML algorithms. The results are summarized in Table 5. Most of the point estimates from Table 5 are similar to the results in Table 4. However, the estimate using neural networks in Table 5 is relatively larger in absolute value (approximately 2.5 times larger) compared to the estimate using neural networks in Table 4. Moreover, the standard errors in Table 5 are similar to the standard errors in Table 4. Again, all of the point estimates in Table 5 are statistically insignificant at the 5% level, so I will not make any conclusions regarding these results.

## 8.5. Hyperparameter Tuning

To estimate LASSO, Logistic Regression, Random Forest, Gradient Boosting, and AdaBoost, I use the sci-kit learn module (Blondel et al., 2011). For XGBoost and Neural Networks, I use the Distributed (Deep) Machine Learning Community's XGBoost and TensorFlow modules, respectively (Abadi et al., 2015; Chen & Guestrin, 2016). For the estimator labeled as LASSO, I use cross-validation and the rule proposed by Belloni, Chen, Chernozhukov, and Hansen (2012) to choose the $\ell_1$ penalties for the $\ell_1$-penalized logistic regression and LASSO, respectively. For the estimates using neural networks, I consider some of the optimization techniques and activation functions available in the TensorFlow module (Abadi et al., 2015). I also apply the $\ell_1$ penalty parameter to the weights in each layer. For completeness, Table 6 outlines which hyperparameters were considered for tuning for each ML algorithm.

It is important to note that using different hyperparameters and ranges of values for the hyperparameter grids can potentially result in different DMLDiD estimates since only finite sets of hyperparameters and the range of their respective values can be chosen. This is especially true for neural networks which have a plethora of combinations of hyperparameters that could have also been considered including dropout layers, different optimization techniques, and different activation functions (Abadi et al., 2015). Moreover, using different programming languages and/or ML libraries could also result in different DMLDiD estimates due to differences in the optimization algorithms and how the objective functions are defined.

## 9. Conclusion

In this paper, I revisit an important research question regarding immigration attitudes that Pardos-Prado and Xena (2019) answer. The authors propose a new theory for how individuals form immigration attitudes (Pardos-Prado & Xena, 2019). To assess this theory, Pardos-Prado and Xena (2019) use traditional econometric methods which may lead to inaccurate conclusions when working with high dimensional data as explained by Chang (2020) and Chernozhukov et al. (2018). To check the robustness of Pardos-Prado and Xena's (2019) results, I re-estimate the ATT using the DMLDiD estimator (Chang, 2020) based on DML by Chernozhukov et al. (2018). I use the same covariates that Pardos-Prado and Xena (2019) use but also explore the polynomial transformed covariates as well as the additional covariates available in the data set. Some of the DMLDiD estimates I obtain are similar to the linear DiD estimates, and all of the DMLDiD estimates are quite different from the SDiD estimates (Pardos-Prado & Xena, 2019). Regardless of the magnitudes of the DMLDiD estimates, I cannot make any conclusions on their similarities or differences when compared to the authors' results because they are all statistically insignificant at the 5% level. Since using a more robust ML based estimator resulted in statistically insignificant estimates while using traditional econometric methods resulted in statistically significant estimates, further investigation may be needed on the causal effect of skill specificity and labor demand on immigration attitudes.

## 10. References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., . . . Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous distributed systems* [White paper]. Google Research. https://www.tensorflow.org/about/bib

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, *72*(1), 1-19. https://doi.org/10.1111/0034-6527.00321

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. https://doi.org/10.3982/ECTA9626

Blondel, M., Brucher, M., Cournapeau, D., Dubourg, V., Duchesnay, E., Gramfort, A., Grisel, O., Michel, V., Passos, A., Pedregosa, F., Perrot, M., Prettenhofer, P., Thirion, B., Vanderplas, J., Varoquaux, G., & Weiss, R. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Chang, N. C. (2020). Double/debiased machine learning for difference-in-differences models. *The Econometrics Journal*, *23*(2), 177-191. https://doi.org/10.1093/ectj/utaa001

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

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. https://doi.org/10.1111/ectj.12097

German Institute for Economic Research (2022). *German socio-economic panel* [Data set]. https://www.diw.de/en/diw_01.c.678568.en/research_data_center_soep.html

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, *80*(2), 313-318. https://www.jstor.org/stable/2006591

Houngbedji, K. (2016). Abadie's semiparametric difference-in-differences estimator. *The Stata Journal*, *16*(2), 482-490. https://doi.org/10.1177/1536867X1601600213

International Labour Organization. (2004, September 15). *Design and structure*. https://www.ilo.org/public/english/bureau/stat/isco/isco88/publ3.htm

Iversen, T., & Soskice, D. (2001). An asset theory of social policy preferences. *American Political Science Review*, *95*(4), 875-893. https://doi.org/10.1017/S0003055400400079

Kambourov, G., & Manovskii, I. (2009). Occupational mobility and wage inequality. *The Review of Economic Studies*, *76*(2), 731-759. https://doi.org/10.1111/j.1467-937X.2009.00535.x

Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*, *13*(2), 151-161. https://doi.org/10.1080/07350015.1995.10524589

Pardos-Prado, S., & Xena, C. (2019). Skill specificity and attitudes toward immigration. *American Journal of Political Science*, *63*(2), 286-304. https://doi.org/10.1111/ajps.12406

Rehm, P. (2009). Risks and redistribution: An individual-level analysis. *Comparative Political Studies*, *42*(7), 855-881. https://doi.org/10.1177/0010414008330595

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688-701. https://doi.org/10.1037/h0037350

Sullivan, P. (2010). Empirical evidence on occupation and industry specific human capital. *Labour Economics*, *17*(3), 567-580. https://doi.org/10.1016/j.labeco.2009.11.003

Zangelidis, A. (2008). Occupational and industry specificity of human capital in the British labour market. *Scottish Journal of Political Economy*, *55*(4), 420-443. https://doi.org/10.1111/j.1467-9485.2008.00460.x