Ludwig-Maximilians-Universität, Munich

Empirical Project: Machine Learning in Econometrics

Estimating the Causal Effects of a Family Policy Reform in Germany

Author:

Lolakshi Rajlakshmi

lolakshi.rajlakshmi@campus.lmu.de

Matriculation Number: 12339702

August 4, 2024

Contents

1 Introduction

As (Blome, 2016) notes, one of the fundamental challenges facing modern welfare economies is the issue of work-family policies. An increasing proportion of women are working, but without adequate support for a dual-earner family system. Against this backdrop, German family policy has undergone path-breaking reforms in the 2000s, moving away from the traditional male breadwinner model (Windwehr and Fischer, 2021). It is in this context that the Bavarian state government implemented a significant policy reform in 2017. The reform aims to provide more support to families by increasing financial benefits for parents, allowing them to receive benefits even if they work part-time, and allowing for a flexible division of benefits between parents.

As with any policy change, it is essential to evaluate the effectiveness of this reform for several reasons. First, it entails significant financial expense on the part of the government. Second, it is important to ensure that the policy modifies incentives in a way that leads to desired outcomes. Third, evaluating the reform helps judge the feasibility of extending it to other areas (e.g., neighboring states whose governments may wish to enact similar reforms).

By affecting some parts of the population (i.e. the Bavarian state) but not others (all other German states), this policy reform provides a natural experiment, creating a natural setting for a difference-in-differences (DiD) analysis to identify the causal effect of this policy on important variables such as income, employment, full-time employment, and job satisfaction. Leveraging this natural experiment and a rich panel dataset containing observations on 4,800 individuals living in different German states across the years 2005, 2010, 2015, and 2020, we use debiased or double machine learning (DML) methods to uncover the causal impact of this reform on outcomes of interest.

Using a DiD analysis with DML methods, or DMLDiD analysis, hinges on several assumptions. Two crucial assumptions among these are that there are parallel trends in outcomes over time and that the estimator we use does not violate Neyman Orthogonality. Together these two assumptions allow us to argue that the estimates we are causal and high-quality, ideas that we will discuss in greater detail ahead. This analysis shows that while the reform had no statistically significant impact on income or employment, it increased job satisfaction and full-time employment. Although the result for full-time meployment is counter-intuitive at first sight, heterogeneous responses to the treatment help explain this.

The rest of this report proceeds as follows. We first take a closer look at the theory behind DID analyses and DML methods, deriving an estimator that we argue is of high quality. Here, we also discuss the 4 main assumptions we make in the context of this application and whether the data supports these assumptions. In the section on empirical analysis and results, we discuss the results from applying our proposed estimator estimate the ATT of the reform on income, employed, full-time, job satisfaction, while flexibly controlling for available individual characteristics. Here, we also examine the heterogeneity in treatment effects. We conclude with a brief summary of our results, important assumptions, possible violations and extensions.

2 Theory

In this section, we will dive into the theory underlying difference-in-differences(DiD) analysis and double machine learning (DML) methods, focusing on important assumptions and the suitability of these methods for our analysis.

2.1 Difference-in-Differences Model

Given the quasi-experimental setup of this reform, with repeated observations on the same units across time, the DiD analysis emerges as a natural estimation technique among many other contenders. Canonical DiD proceeds by comparing changes in the average pre - and post-treatment outcomes across groups who are affected by the policy reform (hereon, treatment group) and those who are not (control group), before and after the policy was enacted. To interpret this difference causally, we rely on the assumption that changes in the treatment group would have evolved along the same path as changes in the control group, in the absence of treatment.

However, it is more common to use a *conditional parallel trends assumption* in applied work, a weaker assumption that requires, in the context of this application, that

$$E\left[Y_0(2020) - Y_0(2015) \mid D = 1, X\right] = E\left[Y_0(2020) - Y_0(2015) \mid D = 0, X\right] \quad (1)$$

holds.

This assumption implies that the expected trends in pre- and post-treatment potential outcomes would have been the same in Bavaria (the treatment group) as in other German states (the control group), had the policy reform (or treatment) not been implemented. Note that this assumption allows for different levels of outcome variables across groups, but requires that the change in these outcomes over time should be the same. For example, Bavaria might have a higher level of income than other states in both time periods. What is important is that in the absence of a policy reform, the change in income in Bavaria and other states would have been the same.

How justifiable is this assumption? There are several cases when this assumption could be violated, ranging from differential macroeconomic trends across Bavaria and other states to the reform causing changes in the composition of the treatment or control group, for example through migration. Additionally, idiosyncratic shocks that change the uptake of the policy might also violate this assumption and bias results. Therefore, we try to evaluate this assumption using our panel data.

We create event study plots for pre- and post-treatment outcomes, which can show whether the treatment and control groups are comparable in trends. We show these plots for income and job satisfaction here (Figure 1 and Figure 2), but the overall results point to this assumption being valid for all outcomes.

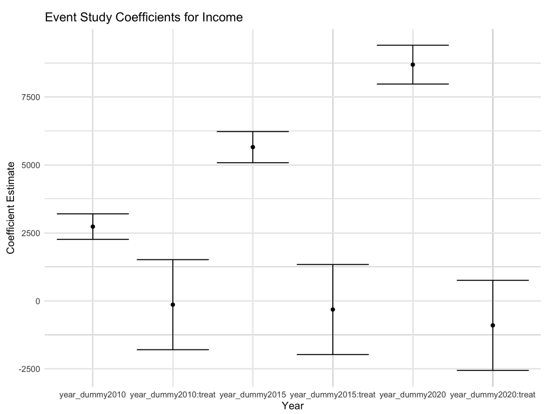

Figure 1: Event Study Plot for Income.

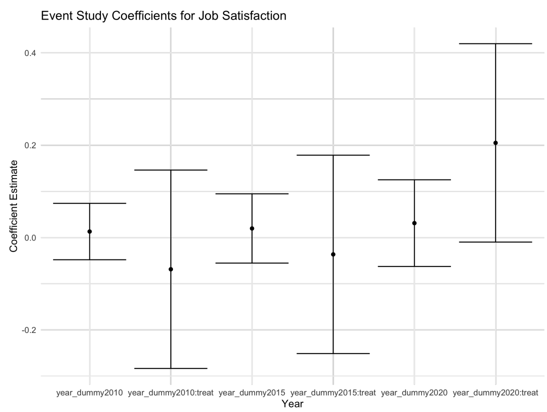

Figure 2: Event Study Plot for Job Satisfaction.

There were no statistically significant differences between the treatment and control group for Income and Job Satisfaction in the pre-treatment period as seen from the interaction term $year\_dummy : treat$.

The second assumption that we make here is common support, i.e.

$$0 < P(D = 1 \mid X) < 1 \tag{2}$$

This requires that every characteristic $X_i = x$ that occurs in the treatment group also occurs in the control group and vice-versa, implying that there exists, for every individual residing in Bavaria, a counterfactual in the control group. We again look at this assumption with the data by inspecting how the propensity score (which is the probability of a unit being assigned to treatment, given a set of observed covariates) is distributed by treatment status. We observe in the analysis that there is an almost complete overlap for the propensity score distributions for the treated and control groups, providing support to the common support assumption.

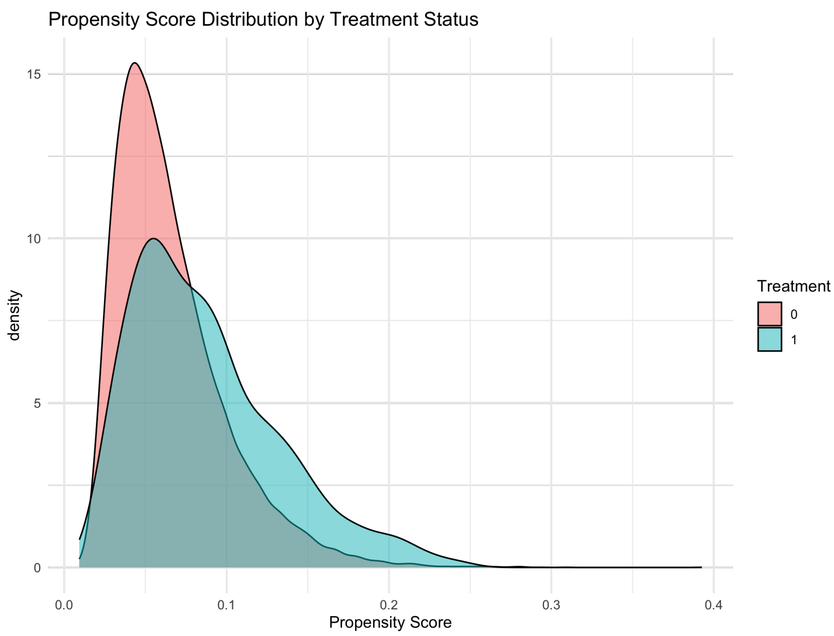

Figure 3: Distribution of propensity scores by treatment status.

The third assumption we make is a simplifying assumption about the conditional means of the treatment status, D, and the change in outcome, ΔY. The linear conditional means assumption is particularly useful because it allows us to use linear regression techniques to estimate the coefficients $gamma_{\Delta}$ and γ_p. These estimates can then be used to adjust for differences in covariates between treated and untreated groups, thereby isolating the treatment effect (ATT). We assume that there exist parameters γ_{Δ}

and γ_p describing the conditional means:

$$E(D \mid X = x) = x'\gamma_p$$

and

$$E[\Delta Y \mid D = 0, X = x] = x'\gamma_\Delta.$$

To summarise, we have the conditional parallel trends assumption, which ensures comparability between the treated and untreated groups by assuming similar trends in outcomes over time. We have the common support assumption, which ensures that every covariate combination has both treated and untreated individuals, allowing for meaningful comparisons. Finally, we have the simplifying linear conditional means. These assumptions ensure that we recover causal estimates of the average treatment effect on the treated (ATT) from a DiD analysis. Now, let's look at the theory and assumptions behind DML methods.

2.2 Double Machine Learning Methods

Double or Debiased Machine Learning (DML) methods are increasingly used in causal inference problems because these predictors possess useful properties that make them desirable. For example, it is common practice to include additional control variables to argue that parallel trends hold among units/individuals that are identical in terms of observed characteristics, which is equivalent to arguing for conditional parallel trends. To obtain estimates of the causal effect in the conditional DiD setting, we must condition on a potentially rich set of observable characteristics. In such cases, traditional DiD methods involving a saturated regression with covariates may not work well due to the high-dimensional nature of the problem (i.e., the number of parameters to be estimated is greater than the number of observations). It is then helpful to employ machine learning methods, such as DML estimation, which help by flexibly estimating parameters used in the measurement of the causal effect.

Another disadvantage of traditional DiD is that it does not capture heterogeneous treatment effects. As (Chang, 2020) notes, the treatment effect estimated from traditional DiD could be weighted with strong negative weights, leading to misleading outcomes. DMLDiD methods overcome this problem and are therefore more suitable in this setting where heterogeneous treatment effects are suspected.

DMLDiD methods work by removing all bias from both the estimators of D and Y, relying on the Frisch - Waugh - Lowell theorem of partialling out. They combine the strengths of machine learning for estimating complex nuisance parameters with traditional econometric techniques for estimating treatment effects. This combination helps in mitigating the challenges posed by high dimensionality, such as overfitting, model misspecification, and inefficiency, thereby providing more accurate and reliable estimates of treatment effects.

The key assumption for estimators using DML estimation is that they have have a moment condition that is Neyman orthogonal. Neyman orthogonality is important for providing high-quality estimation and inference, especially in high-dimensional settings. In high-dimensional settings, we use regularization procedures to estimate the nuisance parameters as solutions to suitable prediction problems. The use of regularization generally results in bias, and we may view using regularized estimates of nuisance parameters as plugging in estimates of these parameters that are close to, but not exactly equal to, the true values of the nuisance parameters. Neyman orthogonality, which guarantees that the target parameter is locally insensitive to perturbations of the nuisance parameters around their true values, then ensures that this bias does not transmit to the estimation of the target parameter, at least to the first order. Therefore, in order to estimate the ATT, we need to derive a Neyman orthogonal estimator.

We are presented with two moment conditions:

$$E\left[ATT - \Delta Y \frac{(D - X'\gamma_p)}{E(D)(1 - X'\gamma_p)}\right] = 0$$

$$E\left[ATT - (\Delta Y - X'\gamma_\Delta) \frac{(D - X'\gamma_p)}{E(D)(1 - X'\gamma_p)}\right] = 0$$

To check for Neyman orthogonality, we need to take gradients with respect to all nuisance parameters and check whether these are zero at the true values of the parameters.

Consider the gradient of the first moment condition

$$\nabla g_0 \mathbb{E}\left[\text{ATT} - \Delta Y \frac{(D - X'g_p)}{E(D)(1 - X'g_p)}\right] = -E\left[\Delta Y \frac{(1 - X'g_p)(-X) - (D - X'g_p)(-X)}{E(D)(1 - X'g_p)^2}\right]$$
$$= E\left[\frac{(\Delta Y/X)(1 - E(D/X))}{E(D/X)(1 - X'g_p)^2}\right]$$

where we have used the law of iterated expectations and substituted in the linear expressions for conditional means. Now, evaluating the derivative at $g_p = \gamma_p$, we obtain

$$E\left(\frac{\Delta Y/X}{E(D/X)(1 - X'\gamma_p)}\right)$$

which is not equal to zero in general as $E(\Delta Y/X)$ is generally not zero.

Let's consider the second moment condition, where we have the nuisance parameters γ_p, γ_Δ. We need to compute gradients with respect to each of these. First:

$$\nabla g_\Delta E\left[\text{ATT} - (\Delta Y - X'g_\Delta)\frac{(D - X'g_p)}{E(D)(1 - X'g_p)}\right] = E\left[\frac{(-X')(D - X'g_p)}{E(D)(1 - X'g_p)}\right]$$
$$= -E\left[\frac{(X')E(D - X'g_p)/X}{E(D)(1 - X'g_p)/X}\right]$$
$$= 0$$

where we again use the Law of Iterated Expectations and substituted the values of g_p. The numerator becomes 0 on substituting $E(D/X) = X'\gamma_p$

Then,

$$\nabla g_p E\left[\text{ATT} - (\Delta Y - X'g_\Delta)\frac{(D - X'g_p)}{E(D)(1 - X'g_p)}\right] = -E\left[\frac{(\Delta Y - X'g_\Delta)(X')(1 - D)}{E(D)(1 - X'g_p)^2}\right]$$
$$= -E\left[\frac{E(\Delta Y - X'g_\Delta)/(D = 0, X) * E((X)(1 - D))/D = 0, X}{E(E(D)(1 - X'g_p)^2)/D = 0, X}\right]$$
$$= 0$$

where the second inequality uses the Law of Iterated Expectations and the third substitutes the true expected value of $E(\Delta Y/D = 0, X) = X'\gamma_\Delta$

Therefore, the second moment condition is Neyman orthogonal and the first is not. We can solve this Neyman orthogonal moment condition for the target parameter to obtain

$$\text{ATT} = E\left[(\Delta Y - X'\gamma_\Delta)\frac{(D - X'\gamma_p)}{E(D)(1 - X'\gamma_p)}\right]$$

We can use this expression to estimate the ATT after estimating the nuisance parameters γ_Δ and γ_p with the following steps:

(i) Construct an estimator $\hat{\gamma}_\Delta$ of γ_Δ: using the observations, run a LASSO/other high-dimensional regression of ΔY on X.

(ii) Construct an estimator $\hat{\gamma}_p$ of γ_p: using all observations, run a LASSO (or another) regression of D_i on X_i.

(iii) Construct an estimator of the ATT using the above Neyman orthogonal expression, substituting in the estimates of the nuisance parameters, and replacing expectations by sample averages:

$$\text{ATT} = \frac{1}{n}\sum_{i=1}^{n}\left[(\Delta Y_i - X_i'\hat{\gamma}_\Delta)\frac{(D_i - X_i'\hat{\gamma}_p)}{E(D_i)(1 - X_i'\hat{\gamma}_p)}\right]$$

It is important to note that we should only use Neyman orthogonal conditions when using DML estimations, because while estimators that do not have this property will also provide an estimate of a causal parameter that will approach the true value in large samples (at a slower than $\sqrt{n}$ rate), the bias of the estimator converges too slowly for standard inference methods to provide reliable inference [1].

We expect this estimator to be of high quality mainly because the orthogonal approach solves two prediction problems – one to predict ΔY and another to predict D – and finds controls that are relevant for both. The resulting residuals are therefore approximately "de-confounded" and provide asymptotically unbiased and normal estimates of the desired causal parameter. This assumption, in addition to the three assumptions related to the DiD setup guarantee that our estimator is of high quality. In the next section, we will look at results from the empirical application of our proposed estimator.

[1] These parameters can however, generally still be used for prediction

3 Empirical Analysis and Results

To compute the DMLDiD estimator of the ATT for the outcomes income, employment, full-time employment and job satisfaction, we use three types of regularized linear regression models - LASSO, Ridge and Elastic Net. We use cross-validation and control flexibly for available individual characteristics, allowing the models to regularize non-predictive covariates.

	Outcome	Alpha	ATT_Estimate	Robust_SE	t_value
1	income	0.00	-201.27	495.13	-0.41
2	employed	0.00	-0.00	0.01	-0.40
3	fulltime	0.00	0.05	0.02	2.45
4	job_satisfaction	0.00	0.23	0.06	3.76
5	income	0.50	-212.16	500.65	-0.42
6	employed	0.50	-0.01	0.01	-0.50
7	fulltime	0.50	0.05	0.02	2.44
8	job_satisfaction	0.50	0.22	0.06	3.66
9	income	1.00	-213.65	501.50	-0.43
10	employed	1.00	-0.01	0.01	-0.51
11	fulltime	1.00	0.05	0.02	2.45
12	job_satisfaction	1.00	0.22	0.06	3.66

Table 1: Results

Across all values of alpha, the ATT estimates for income remain negative, ranging from approximately -201.27 to -213.65, but are not statistically significant with t-values around -0.4, indicating no significant impact on income. For employment status, the ATT estimates are close to zero and also lack statistical significance (t-values between -0.40 and -0.51). Full-time employment and job satisfaction show positive ATT estimates across all alpha values, with full-time employment consistently having a small effect (0.05) and job satisfaction showing a more substantial positive effect (0.22). Both full-time employment and job satisfaction exhibit statistical significance (t-values around 2.44 to 3.76), suggesting a meaningful impact of the treatment on these outcomes. Overall, while the treatment effect on income and employment status is not significant, there is consistent evidence of a positive and significant treatment effect on full-time employment and job satisfaction.

Using all three linear regularization models—LASSO (alpha = 1), Ridge (alpha = 0), and Elastic Net (0 < alpha < 1)—in a Double Machine Learning (DML) estimation is beneficial as each model leverages unique strengths: LASSO performs variable selection by shrinking coefficients to zero, enhancing interpretability and handling high-dimensional data; Ridge shrinks coefficients towards zero without eliminating any, stabilizing estimates when predictors are highly correlated; Elastic Net combines both benefits, balancing variable selection and shrinkage, making it robust against various data structures and multicollinearity.

We use cross-validation because it improves Double Machine Learning (DML) estimates by ensuring that the machine learning models generalize well to unseen data. It helps in selecting optimal hyperparameters, balancing the bias-variance tradeoff, and providing robust performance assessments by averaging results over multiple folds. This leads to more accurate and stable nuisance parameter estimation, which is crucial for unbiased treatment effect estimation.

3.1 Heterogeneity in the treatment effect

We also look at whether the treatment effects are heterogeneous across different types of individuals and find strong support for this. For example, coefficients from the income heterogeneity analysis indicate the impact of various factors on income. Key variables like education (0.044), job satisfaction (0.098), homeownership (0.131), employment status (0.093), and being full-time employed (0.142) have significant positive effects on income. Demographic factors like age (0.0012), gender (0.021), marital status (0.062), and number of children (-0.036) also influence income but to varying degrees. Notably, interactions between these variables, such as age and education (0.0015), or being married and female (-0.014), also play a role in determining income levels. The analysis reveals a complex interplay of personal, professional, and demographic factors affecting income.

This interplay is not just limited to effects on income. Looking at job satisfaction, which had statistically significant estimates of ATT, we see that being employed has a positive impact (Coefficient: 0.13175), while being female and having children negatively affect satisfaction (Coefficients: -0.01662 and -0.00130, respectively). Homeownership also shows a negative association (Coefficient: -0.00840). Additionally, factors such as education and being married positively influence job satisfaction (Coefficients: 0.00090 and 0.00412, respectively). Notably, commuting time has a positive effect (Coefficient:

0.00077), while income exhibits a slight negative impact (Coefficient: -0.00000015).

4 Conclusion

In this report, we tried to identify and estimate the causal effect of a family-policy reform on income, employment, full-time employment and job satisfaction. We used DML-DiD estimators for this, relying on assumptions of conditional parallel trends, common support, linear conditional means and Neyman orthogonality to estimate the ATT. We verified and justified these assumptions theoretically and empirically, providing support for our estimator. We used a variety of tuning parameters in our estimation and can conclude that the ATT is insensitive to this choice, providing further support for our estimates. In our empirical analysis, we controlled flexibly for individual characteristics and used cross-validation to improve our estimates. Overall, our estimators and estimates are both of high quality.

The main results we draw from this analysis is that the policy reform did not affect income or employment, but had positive effects on full-time employment and job satisfaction. It is important to note here that all treatment effects are heterogeneous across individual types, as shown by the heterogeneity analysis. It would be interesting to explore this heterogeneity in further detail. Another possible extension would include looking at other prediction models from machine learning.

References

Blome, Agnes. 2016. *The politics of work-family policy reforms in Germany and Italy*. Routledge.

Chang, Neng-Chieh. 2020. "Double/debiased machine learning for difference-in-differences models." *The Econometrics Journal* 23(2):177–191.

Chernozhukov, Victor, Christian Hansen, Nathan Kallus, Martin Spindler and Vasilis Syrgkanis. 2024. "Applied causal inference powered by ML and AI." *arXiv preprint arXiv:2403.02467* .

Windwehr, Jana and Torben Fischer. 2021. "The limits of change: German family policy and the dynamics of policy transfer 2009–2017." *German politics* 30(2):189–207.

Declaration

I confirm that this report is based on my own work. In preparing this report, I have not received any help from another human nor have I discussed any aspects of the empirical project with others.

Lolakshi

Signature

Munich, August 4 2024

Place, Date