

NBER WORKING PAPER SERIES

SEEING BEYOND THE TREES:  
USING MACHINE LEARNING TO ESTIMATE THE IMPACT OF  
MINIMUM WAGES ON LABOR MARKET OUTCOMES

Doruk Cengiz  
Arindrajit Dube  
Attila S. Lindner  
David Zentler-Munro

Working Paper 28399  
<http://www.nber.org/papers/w28399>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
January 2021

We thank Erin Conlon, Ezgi Cengiz, Ina Ganguli, Laura Giuliano, Carl Nadler, Hasan Tekguc, Michael Reich, Jesse Rothstein and participants at LERA 70th Annual Meeting, IRLE Research Presentation seminar, 44th Eastern Economic Association Conference and 2018 The New School-UMass Economics Graduate Student Workshop, and the Authors Conference in honor of Alan Krueger for very helpful comments. We are also grateful to Jon Piqueras for outstanding research assistance. A previous version of the draft was circulated under the title "Seeing Beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Affected Individuals." Lindner acknowledges financial support from the Economic and Social Research Council (new investigator grant, ES/ T008474/1) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement Number 949995). Dube acknowledges financial support from Russell Sage Foundation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Doruk Cengiz, Arindrajit Dube, Attila S. Lindner, and David Zentler-Munro. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Seeing Beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes

Doruk Cengiz, Arindrajit Dube, Attila S. Lindner, and David Zentler-Munro

NBER Working Paper No. 28399

January 2021

JEL No. J08,J2,J3,J38,J8,J88

**ABSTRACT**

We assess the effect of the minimum wage on labor market outcomes such as employment, unemployment, and labor force participation for most workers affected by the policy. We apply modern machine learning tools to construct demographically-based treatment groups capturing around 75% of all minimum wage workers—a major improvement over the literature which has focused on fairly narrow subgroups where the policy has a large bite (e.g., teens). By exploiting 172 prominent minimum wages between 1979 and 2019 we find that there is a very clear increase in average wages of workers in these groups following a minimum wage increase, while there is little evidence of employment loss. Furthermore, we find no indication that minimum wage has a negative effect on the unemployment rate, on the labor force participation, or on the labor market transitions. Furthermore, we detect no employment or participation responses even for sub-groups that are likely to have a high extensive margin labor supply elasticity—such as teens, older workers, or single mothers. Overall, these findings provide little evidence for changing search effort in response to a minimum wage increase.

Doruk Cengiz  
OM Partners  
2727 Paces Ferry Road  
Atlanta, GA 30339  
dcdorukcengiz@gmail.com

Arindrajit Dube  
Department of Economics  
University of Massachusetts  
Crotty Hall  
412 N. Pleasant Street  
Amherst, MA 01002  
and NBER  
adube@econs.umass.edu

Attila S. Lindner  
Department of Economics  
University College London  
30 Gordon Street  
London  
WC1H 0AX  
United Kingdom  
and CERS-HAS, IZA and IFS  
a.lindner@ucl.ac.uk

David Zentler-Munro  
Department of Economics  
University College London  
30 Gordon Street  
London WC1H 0AX  
United Kingdom  
and CReAM  
david.zentler-munro.13@ucl.ac.uk

# 1 Introduction

A long-standing question in economics centers around understanding how minimum wages affect low-wage labor markets. A key challenge to convincingly answer this question comes from being able to successfully identify the workers who are actually affected by the policy. This has led many researchers to focus on specific industries or demographic groups such as teens ([Allegretto et al., 2017](#); [Card, 1992](#); [Neumark and Wascher, 1992](#); [Neumark et al., 2014](#); [Totty, 2017](#)), younger workers with lower educational credentials ([Clemens and Wither, 2016](#); [Clemens and Strain, 2017](#); [Manning, 2016](#); [Sabia et al., 2012](#)), and individuals with no high school education ([Addison and Blackburn, 1999](#); [Addison et al., 2011](#)). However, these groups constitute relatively small shares of all minimum wage workers. As a result, there is a tension between what is often analyzed (e.g., minimum wage effects on teens) and what is argued (effects of the policy on affected workers largely composed of adults) ([Belman et al., 2015](#); [Manning, 2016](#)).<sup>1</sup>

In this paper, we use machine learning tools to predict which individuals were likely affected by minimum wage increases, and then estimate the impact of minimum wages on the individuals predicted to be exposed to the minimum wage. This approach extends the [Card and Krueger \(1995\)](#) prediction approach that was proposed in “Myth and Measurement” (see p. 135), but has been undeservedly neglected in the literature ever since.<sup>2</sup>

We construct various groups based on the predicted probabilities to assess the impact of the policy on workers who are highly likely to be exposed to the minimum wage (whom we refer to as the “high-probability” group) and also a wider group where we can retrieve 75 percent of likely minimum wage workers (whom we refer to as the “high-recall” group). We then study the impact of the policy on various labor market outcomes such as employment, unemployment and labor force participation of workers with different exposure to the minimum wage. The impact of the

---

<sup>1</sup>The discrepancy is particularly significant when the outcome of interest is teen employment rate, which is the subject of research in the U.S. [Belman and Wolfson \(2014\)](#) scrutinize 30 studies, which examined the employment effects of the minimum wage on some demographic groups between 2001 and early 2013 and find that 17 of them had teen employment as the dependent variable. [Neumark \(2017\)](#) shows that 12 out of 13 studies that examined minimum wage effects on unskilled employment between 2010 and 2016 focused on teens (see Table ??). At the same time, teens are less likely to be in the target group than non-teen adult minimum wage workers ([Lundstrom, 2016](#)). Compared to affected non-teens, only a relatively small share of teens live in poor households. According to the 2016 American Community Survey, 18.4% of teens are poor.

<sup>2</sup>We are only aware of one previous publication that applied this method ([Cengiz et al., 2019](#)), whose co-authors list includes three of the authors of this paper. That paper primarily utilized this approach as a side note to show the differences between the Card and Krueger’s prediction approach and the bunching method developed there.

minimum wage on these latter outcomes has been extensively studied in the theoretical literature (e.g. [Flinn \(2011\)](#)), despite the scant empirical evidence on them.

Our proposed approach has several advantages. First, we can assess the effect of the minimum wage on a large fraction of low-wage workers, and not just on some specific sub-groups with high exposure such as teens or youth (see e.g. [Laws \(2018\)](#)). In that sense, our approach is close in spirit to [Cengiz et al. \(2019\)](#), who assess the overall employment impact of the policy. However, here we assess the effect of the minimum wage on important labor market outcomes that have not been studied, and cannot be readily studied using the frequency distribution based approach of [Cengiz et al. \(2019\)](#). Additionally, we can also directly study the impact of the policy on the affected non-teen (20-64), and prime age (25-55) individuals who are more likely to live in poor households than teens, and tend to be the intended beneficiaries.

Third, we can also study the impact of the policy on the groups of workers with low predicted probability—i.e., who should not be affected by the policy. Specifications that show an unrealistically large impact on workers not operating on low-wage labor markets should be cautiously interpreted as they raise questions about the credibility of the particular research design. Using the impact on the “low-probability” group for a falsification test is equivalent to studying the upper tail employment changes – a fruitful approach that successfully resolved some of the discrepancies in the minimum wage literature (see [Cengiz et al. \(2019\)](#)).

Finally, our approach allows us to study the impact of the policy on various labor market outcomes that would not be possible with the distribution based approach developed in [Cengiz et al. \(2019\)](#). This is a major advantage as we can provide the first comprehensive picture on how low-wage labor markets evolve in response to the minimum wage. The impact of the policy on unemployment and participation rate is often discussed in the policy discourse, but compelling evidence is limited and mainly focuses on some specific subgroups. Furthermore, as we will demonstrate later, understanding the impact of the policy on outcomes besides employment is relevant since it has welfare implications in various non-competitive models (e.g. [Flinn \(2011\)](#)),

We implement the proposed approach by using machine learning (ML) methods and demographic information to predict which individuals are likely to be minimum wage workers in the Current Population Survey (CPS) data between 1979 and 2019. We use individual’s demographic and educational characteristics to predict their probability of having an hourly wage less than 125%

of the statutory minimum wage.<sup>34</sup> We apply three main tree-based learning tools in the training data: decision trees, random forests and gradient boosting tree and we also explore the elastic net regularization of the logistic regression. A key advantage of the ML tools over the [Card and Krueger \(1995\)](#) approach is that they do not require the researcher to pre-specify the functional form of the prediction model, which is instead determined in a data-driven way. Then we compare the performance of various prediction models in the test data.

The best performing prediction comes from the gradient boosting tree model. At the same time, it is worth pointing out that the original linear prediction model proposed by [Card and Krueger \(1995\)](#) (with a judiciously chosen set of interactions) also performs relatively well, although a little worse than the state-of-the-art machine learning tools. When compared to a commonly used demographic groups in the literature (such as teens, or high school or less under the age of 30), the boosting approach can form groups with a similar number of (correctly classified) minimum wage workers while substantially reducing the number of (mis-classified) non-minimum wage workers. The gains in precision (i.e. the correctly classified share) for a given level of recall (i.e. share of minimum wage workers included in the group) is sizable when we limit attention to non-teen workers—a group that is of particular interest to policymakers.

Armed with the prediction model, we implement an event study analysis that exploits 172 prominent state-level minimum wage increases between 1979 and 2019. We assess the impact of the policy on various groups formed based on the predicted exposure probability. The “high-probability” group comprises 10% of the population with the highest likelihood of being affected by the policy. We also study the impact of the policy on the high-recall group that captures 75% of all minimum wage workers.

For both groups, we find a considerable increase in wages after the policy change though the wage increase is somewhat lower for the high-recall group. At the same time, we detect a small, positive, and statistically insignificant effect on employment for both groups. The implied

---

<sup>3</sup>As we describe below, we assess the effect of the minimum wage in an event study framework by focusing on prominent state-level minimum wage changes. In the prediction exercise we only use data preceding these minimum wage hikes.

<sup>4</sup>The full set of predictors and how they are coded are reported in Appendix A. In the prediction model, we do not use variables related to past employment status or occupation/industry in the CPS-ORG. They are sometimes missing even if the individual is currently in the labor force and looking for job. We prefer to keep these observations in the sample, as they potentially carry information about labor supply effects of the policy. In our preferred prediction model, we do not use state of residence or year information either. This choice is primarily to be able to build samples that are comparable and consistent across time and space.

employment elasticity with respect to own wage—the labor demand elasticity in the standard competitive model of the labor market—is 0.28 (s.e. 0.34) for the high-probability group and 0.18 (s.e. 0.22) for the high-recall group. Both of these estimates can rule out anything more than modest negative disemployment effects at the conventional significance levels.

At the same time, there is no evidence of substantial changes in the unemployment or participation rates in response to the policy. We are also not able to detect any economically meaningful (or statistically significant) changes in labor market transitions between employment, unemployment and non-participation. This lack of response on the labor-force participation margin provides new evidence that minimum wages have a limited impact on search effort when we focus on most individuals affected by the policy.

Our results are robust to controlling for time-varying heterogeneity in a wide variety of ways. Moreover, the increase in wages lines up well with the timing of the minimum wage increases, and the effects only emerge in the group of individuals likely to be exposed to the minimum wage. We find no significant differences in labor market outcomes for the low-probability group suggesting that no unusual changes took place around the minimum wage hikes. All these findings underline the credibility of our research design.

Furthermore, we also study whether the responses to the policy vary across demographic groups. Most importantly, we study whether differential responses can be detected on employment, unemployment, and participation margins for workers who are thought to have larger extensive margin labor supply elasticities—such as teens, older workers, and single mothers. In addition, we also assess the impact of the policy by the likelihood of moving in or out from the labor force. We use demographic information and apply machine learning tools again to classify workers as being more likely to move in and out from the labor force. Even when we focus on the group of workers with the highest predicted transition probabilities, we find no evidence of substantial change in the employment or participation rates.

This paper contributes to several strands of the minimum wage literature. Although there is a thick literature on the minimum wage, there are only a handful of studies that examine large segments of affected workers in the U.S.<sup>5</sup> Notable exceptions are [Cengiz et al. \(2019\)](#) who use the frequency distribution of wages to focus on the number of low-wage jobs, and [Meer and West \(2015\)](#)

---

<sup>5</sup>See [Belman and Wolfson \(2014\)](#) for a thorough literature review on the subject.

who simply consider total state-level employment to study the overall impact of the policy. Our estimates on the high-recall group complement the existing evidence by providing an alternative way of assessing the impact of the minimum wage on overall employment.

Our paper also fills an important gap in the literature by going beyond studying the wage and employment effects of the policy. Only a few papers have studied directly the effect of the minimum wage on unemployment or participation rate, which are likely to reflect changes in job search behavior. [Laws \(2018\)](#) and [Godoy et al. \(2020\)](#) study changes in labor force participation of specific subgroups such as youth and parents. [Borgschulte and Cho \(2019\)](#) studies the impact on workers close to retirement. Furthermore, [Adams et al. \(2018\)](#) study the impact on aggregate job search. Similarly to us, they find no indication for significant changes in job search. Nevertheless, it is unclear whether they have enough statistical power to detect significant changes by studying the impact on all U.S. workers, including those with high wages. Since in the U.S. context, a relatively small fraction of the workforce is affected by the policy, positive wage effects cannot be detected without focusing on workers around the minimum wage ([Cengiz et al., 2019](#)). As a result, it may not be surprising that [Adams et al. \(2018\)](#) were not able to detect significant changes in aggregate job search.

Methodologically, our use of a demographics-based predictive model for minimum wage workers is inspired by [Card and Krueger \(1995\)](#) who examine the 1988 California minimum wage increase and use a linear prediction model to sort individuals living in the state in 1987 according to their likelihood of having hourly wages between the old and new minimum wage (\$3.35 and \$4.25). The [Card and Krueger \(1995\)](#) model is based on subjective judgments about predictors and the functional form and includes complicated multi-way interactions. Even though this subjective assessment turns out to have been implemented incredibly well in [Card and Krueger \(1995\)](#), the key advantage of the machine learning-based approach proposed here is that we do not need to rely on such judgments. Instead, the ML tools determine the prediction model in a data-driven way, and can provide a much better guarantee against overfitting and specification hunting.

The remainder of the paper is organized as follows. In the next section, we briefly discuss the benefits of examining outcomes besides the employment change. Section [3](#) describes the data sets employed. Section [4](#) explains how we apply various learning tools to predict exposure to the minimum wage. Section [5](#) examines the empirical implementation and the key results. Section [6](#)

concludes.

## **2 Participation, Unemployment and the Welfare Impacts of the Minimum Wage**

The approach developed in this paper allows us to analyze the impact of the minimum wage on a wide range of outcomes that go beyond the traditional employment and wage impacts emphasized in the literature. Studying the impact on unemployment and participation rate is particularly interesting since a large class of theoretical models have direct predictions on these outcomes. For instance, efficiency wage models (Drazen, 1986), models with information asymmetry (Lang, 1987), or search models (Swinnerton, 1996; van den Berg, 2003) argue that minimum wages will raise unemployment rates.

The participation margin is also interesting given the growing number of empirical studies finding close to zero effect on employment. [Flinn \(2006\)](#) and [Ahn et al. \(2011\)](#) discuss in detail what mechanisms are needed to find positive employment effects in a search model. They highlight endogenous participation rate or endogenous search effort as being essential. [Flinn \(2006\)](#) also points out that to find simultaneously an increase in employment rate and a decrease in the unemployment rate, one needs to introduce both elements: an increase in the participation rate and search effort. Therefore, by providing direct evidence on employment, unemployment and participation rate simultaneously, we can assess the empirical validity of this search-based explanations.

In addition to that, the policy impact on labor market participation is potentially informative with respect to welfare for some groups of minimum wage workers. While the employment and the unemployment responses often reflect both the supply and demand side of the market, whether someone is participating on the labor market solely depends on the worker's decision. As a result, if someone chooses to participate, then that individual directly reveals that the pay-off from searching and finding a job in expectation makes him/her better-off. Assuming that minimum wages do not have an impact on non-participation payoffs, the individuals who decide to participate are going to be made better-off by the minimum wage. Conversely, the workers who leave the labor force in response to the minimum wage will be made worse-off.

Our emphasis on studying participation differs somewhat from the literature that emphasizes the presence of wage spillovers as an indicator of welfare impacts (e.g. [Flinn \(2002\)](#)). A key requirement for wage spillovers to fully reflect welfare changes is that wages are the only welfare relevant dimension of a job, i.e., there are no non-wage amenities. If minimum wage policies alter job attributes, the positive (negative) wage spillovers are neither a necessary nor a sufficient condition for a welfare gain (loss). Empirically, a growing body of evidence suggest that minimum wage has an impact on job amenities. For instance, [Dustmann et al. \(2020\)](#) find that the introduction of the German minimum wage led to an increase in commuting times. [Clemens et al. \(2018\)](#) find that minimum wages alter the provision of the employer provided health insurance. [Dube et al. \(2016\)](#) find that separations fall following minimum wage increases, which would be consistent with improving job quality.

The advantage of considering participation decisions is that those are driven by workers' overall assessment of the quality of jobs available (and the probability of getting those jobs). As a result, examining the impact on participation can improve upon a piecemeal approach of looking at wage spillovers or specific measures of amenities. Nevertheless, the change in participation is a discrete decision and so the welfare impact of the minimum wage is only revealed for workers who are on the margin. As a result, we also study the intensive margin by considering the impact of the minimum wage on unemployment rates and transition probabilities, while noting that this impact reflects an equilibrium between supply and demand changes.

### **3 Data**

The primary data set we employ throughout the analysis is the Current Population Survey (CPS). We use the 1979-2019 CPS Outgoing Rotation Group (CPS-ORG) sample for the hourly wage and weekly earnings variables. This is a subset of the Basic Monthly CPS, a monthly survey of approximately 60,000 households in the U.S. The CPS-ORG includes only the fourth and eighth sample months, when usual hourly wages, weekly earnings and weekly hours worked are asked. These variables are of primary importance for the prediction as well as for the estimation and thus we rely on their accuracy. For this reason, we exclude observations with imputed hourly wages, imputed weekly earnings or imputed hours worked. For hourly workers, we use the reported

hourly wage, and for other workers we define the hourly wage to be their usual weekly earnings divided by usual weekly hours. We also use a range of demographic variables in the data set when predicting individual's likelihood of having a wage close to the minimum. These variables indicate individual's age, race, Hispanic status, gender, education, veteran status, marital status, and rural status of the residency (see Appendix B for the exact definitions).

We use the 1979-2019 CPS Basic files (CPS-Basic) for the employment, unemployment, and labor force participation (LFP) variables as well as for a number of secondary variables describing the nature of employment (part-time, over-time and self-employment). Unlike the CPS-ORG, CPS-Basic contains observations for every month that a respondent is surveyed. Using CPS-Basic to estimate employment, unemployment and LFP effects of the minimum wage therefore results in greater precision than using CPS-ORG to estimate these effects. It also allows us to estimate the impacts of the minimum wage on transitions between employment, unemployment and inactivity.

We obtain the minimum wage data from [Vaghul and Zipperer \(2016\)](#), which has been extended until 2019 by the authors.

## 4 Predicting to Be a Minimum Wage Worker

We examine the likelihood of an individual being exposed to the minimum wage. This is a classic prediction problem. We build a prediction model to explain the relationship between being a minimum wage worker (defined as having an hourly wage less than 125% of the statutory minimum wage) and the demographic and educational variables.<sup>6</sup> Then, we use the model to predict the likelihood of an individual to be a minimum wage worker. As the model relies on demographic and educational variables, we can ascertain likelihood of an individual being affected by the policy even if the individual is currently non-employed or has no wage. As a result, we can examine the effects of the policy not only on incumbents but also on those that are currently non-employed.

To build the prediction model and assess its power, we divide the data into three mutually exclusive samples: training, test and leave-out. To construct the training and test samples, we

---

<sup>6</sup>Our results are not sensitive to the definition of minimum wage workers based on alternative cut-off values. Setting the thresholds to 3% above the minimum wage or 200% above the minimum wage produces virtually the same ordering of observations according to predicted probabilities, suggesting that the specific definition we use has essentially no bearing on the conclusions.

focus on the quarters where there is a minimum wage event in the next 12 quarters and where there is no minimum wage event in the last 20 quarters (to fully allow wages to adjust to the new environment). Then, we randomly pick 150,000 out of 469,174 observations that satisfy the conditions. The test sample is composed of every wage worker that is not in the training sample. All the remaining observations are in the leave-out sample. The leave-out sample is not used in building the prediction model or assessing it.

We apply various learning tools such as random forests, tree boosting, basic logistic, elastic net, and the linear probability model along the lines of [Card and Krueger \(1995\)](#). We fit each model using the training data and then compare the performances of the models in the test data by plotting the precision-recall curves. We describe next the key idea behind each prediction algorithm. For further details, we refer the reader to Appendix C and [Friedman et al. \(2009\)](#).

## 4.1 Prediction Algorithms

**Decision Trees:** A single decision tree lies at the basis of many learning techniques, including random forests and gradient tree boosting. A decision tree recursively divides the feature (predictor) space into two in a way that reduces the pre-specified loss function the most.<sup>7</sup> More concretely, in the beginning, the tree tries every possible split to divide the entire sample space into two, and picks the one that diminishes the loss function the most. Subsequently, each subsample is treated as the new sample, and the first step is repeated. Once the splitting is over, it predicts the class of every observation according to the majority vote in the subspace (terminal node) to which the observation belongs.

This procedure requires a decision on when to stop the splitting. In principle, the splitting could continue until there is only one data point at each terminal node. Such a tree fits the training sample perfectly, but would suffer from overfitting. To overcome the problem, it is common to use cross-validation to determine the complexity of the tree. For a more accurate prediction, we collapse some internal nodes (“prune the tree”), and decrease the prediction variance at the expense of bias.

Decision trees are not among the most successful learners, yet they are relatively easy to

---

<sup>7</sup>The loss function is the deviance, defined as  $-2 \sum_m \sum_k n_{mk} * \log(\hat{p}_{mk})$ ; where  $n_{mk}$  indicates the number of observations at terminal node  $m$  that belongs to class  $k$  and  $\hat{p}_{mk}$  is the share of observations at terminal node  $m$  that belongs to the class  $k$ .

interpret. In figure 1, we plot a pruned decision tree produced to predict whether a wage worker has an hourly wage of less than 125% of the statutory minimum wage using the demographic and educational characteristics. The tree predicts that the only group in the training sample with hourly wages less than the threshold is the one with those younger than 19 years old. The majority vote in all the other terminal nodes is “FALSE”, indicating that non-teen observations are expected to work for hourly wages higher than the threshold.

Interestingly, the recommendation based on a simple decision tree is to proxy minimum wage workers with something that is very close to the commonly used teen sample (which includes individuals younger than 18 years of age). Nevertheless, as we show below, there are ways to obtain much better predictions by combining multiple decision trees. The two most common ways to do so are the random forest by Breiman (2001) and the gradient boosted trees by Friedman (2001).

**Random Forest:** The random forest is a tree-based ensemble learning technique. It provides a way to overcome the bias-variance trade-off of a single tree. It constructs a multitude of fully grown decision trees formed using different training bootstrap samples. Each tree produces unbiased predictions that have large variances. We calculate the average of the predictions, thereby reducing the variance. To further reduce the variance, we decrease the correlation among trees by employing a randomly selected portion of the predictors at each split. Although this results in the loss of the interpretability of individual trees, it has no impact on the bias since individual trees are still fully grown.<sup>8</sup> Our 5-fold cross-validation finds that the “optimum” random forest is achieved with 2,000 trees and only two predictors tried at each split.

**Boosting:** The boosting approaches the problem of how to combine multiple trees from a different angle. Instead of producing many fully-grown trees and averaging them, the trees in this model are grown sequentially where subsequent trees attempts to fix the errors of the preceding ones. As a result, while the first tree in the boosting is interpretable, the subsequent trees are not independently meaningful. Intuitively, with the boosted trees one starts with the lowest-hanging fruit and, say, classify teens as minimum wage workers, and others as not minimum wage workers. Then, the second or subsequent tree builds a model that focuses more on correctly classifying non-teen minimum wage workers and teen non-minimum wage workers, namely the observations

---

<sup>8</sup>Note that if trees are perfectly correlated, the reduction of the variance would be nil. If they are independent, the variance of the final model would be  $\frac{\sigma^2}{B}$ , where  $\sigma^2$  is the prediction variance of a single tree and  $B$  indicates the number of trees.

misclassified by the first tree. The change of focus is usually achieved by altering the outcome variable (e.g. using the residual as the outcome variable) or slightly changing the loss function (weighting the misclassified observations more heavily). After building the subsequent tree, we combine the predictions of all trees through a weighted majority vote. Based on our 5-fold cross-validation, the “optimum” boosted trees model is obtained with the following parameters: number of trees = 4,000; shrinkage factor = 0.005; depth of tree = 6; minimum observations in a node = 10.

**Elastic net:** We use the elastic net regularization developed by [Zou and Hastie \(2005\)](#). The underlying model is very similar to the logistic regression, except that the elastic net model penalizes model complexity. The penalty term is a linear combination of the lasso and ridge methods. It intuitively aims to compensate for the limitations of one of the latter methods with the advantages of the other and thus tends to produce more accurate predictions than both of them. As opposed to the previous models, the elastic net regression requires pre-specification of the exact functional form for the predictors in the prediction equation. Therefore, we purposefully build a fairly complex model, where we include all the predictors, their four-way interactions, all the interactions with the quadratic age variable, and the cubic and quartic terms of the age variable. We rely on the regularization to simplify the model and prevent overfitting.

**Card and Krueger’s linear probability model:** This is a trial and error method employed by [Card and Krueger \(1995\)](#). We follow the functional form proposed on page 135. They apply a linear probability model and use the following right hand side variables: a set of three-way interaction variables between teen, non-white, and gender indicators; three-way interaction variables between young adult (age 20-25), non-white, and gender indicators; three-way interactions of age, categorical education, and gender variables; quadratic and cubic terms of the age variable; indicator variables for Hispanic, and non-white individuals. We can think of the Card and Krueger’s model as a sort of elastic net approach to predictor selection, but where the regularization is based on subjective judgment instead of a formal learning algorithm.<sup>9</sup>

---

<sup>9</sup>We also tried to implement neural networks and support vector machines. While the model constructed using the neural networks performs slightly worse than the boosting, the models using the support vector machines fail to provide a well-performing prediction model.

## 4.2 Precision-Recall Curves and Predicted Probabilities

To compare the models with each other, we employ two concepts from the machine learning literature: precision and recall. Precision refers to the number of minimum wage workers in the predicted group as a proportion of the sample size. On the other hand, recall refers to the number of minimum wage workers in the predicted group as a proportion of the number of minimum wage workers in the entire population. For instance, if a predicted group has only one observation and the observation is a minimum wage worker, then the precision is 1; however, here the recall is very small as the sample will cover only a minuscule fraction of the minimum wage workers in the population. On the other hand, if the predicted group contains every observation in the population, then the recall rate is 1 as the sample, by construction, includes all the minimum wage workers in the population. However, the precision is going to be small since the predicted group also includes all the non-minimum wage workers in the population. The ideal is to construct a predicted group that includes all the minimum wage workers and none of the non-minimum wage workers, so both the precision and the recall are 1. Generally, the higher the precision for a given recall rate, the better the performance of the model is.<sup>10</sup>

Figure 2 Panel (a) shows the precision-recall curves corresponding to the above described prediction algorithms. We also estimate and report the performance of a basic logistic model with age and the categorical education variables as predictors for comparison. To plot the curve we calculate the predicted probabilities for each individual in the test sample. Then, we define the predicted group for alternative probability thresholds where all workers in the group have a predicted probability greater than the threshold. We calculate the precision and the recall of each one of these groups and obtain the curve. In other words, each point on the curves corresponds to a separate predicted group. When we raise the threshold, we expect the precision to increase, but at the cost of a reduced recall rate. How strong this trade-off is between precision and recall rate for various prediction models is shown on the figure.

The figure shows that the boosted tree (black solid line) outperforms other prediction models

---

<sup>10</sup>Another approach commonly used to compare models is to plot the receiver operating characteristic (ROC) curve. The ROC curve plots the recall against false positive rate, the latter defined as the number of non-minimum wage observations as a proportion of the number of non-minimum wage workers in the population. The difference in interpretation between the precision-recall and the ROC curves is that while the former answers the question of “For the given data set, how well the model performs?”, the latter answers “How well the model performs in general?” In our case, we reach the same conclusion whether we use the ROC curve or the precision-recall curve.

since it provides the highest precision at almost all recall levels. For comparison, in Panel (b) we report the other prediction models relative to the boosted tree model. The boosted tree model (and also the other prediction algorithms) improves precision considerably relative to the basic logistic model. At the same time, the differences between the other prediction models and the boosted tree are relatively small especially at higher recall rate levels. The random forest model achieves almost the same result as the boosted tree. It is also notable that the Card and Krueger’s subjective judgment approach does almost as well as the elastic regularization of the logistic model and the performance of their model is not far behind of the best performing prediction model.

In Figure 3 we compare the performance of the best prediction model, the boosted tree, relative to the strategy of choosing specific subgroups to proxy minimum wage workers.<sup>11</sup> First, in terms of precision, the teen sample performs better than all the other commonly used samples that we compare (workers younger than 30 with no high school degree (LTHS, Age<30), workers younger than 30 with high school or less education (HSL, Age<30), and workers with no high school degree (LTHS)). In fact, when compare to “LTHS, Age<30”, both recall and precision values of the teen sample are higher. This indicates that the teen sample includes more workers that truly have hourly wages lower than 125% of the minimum, and it captures them more accurately than the former samples. Second, the commonly used samples that include non-teen workers tend to achieve a higher recall than the teen sample. However, the rise in the recall is expensive in terms of the lost precision. For instance, including all high school or less workers younger than 30 increases the recall value by 0.132 compared to the teen sample, yet the precision decreases by 0.24. It implies that many non-teen observations in the “HSL, Age<30” sample are actually non-minimum wage workers.

Overall, it is clear that points on the curve are closer to the top-right corner than the points corresponding to the commonly used samples. Furthermore, the difference in the precision values between the commonly used samples and the samples recommended by the tools increases as the recall increases. For instance, the precision value that the learning tools achieve for the teen sample’s recall value is only slightly greater than that of the teen sample (the vertical distance

---

<sup>11</sup>Of course, it is possible that someone is directly interested in the impact of the policy on the labor market outcomes of certain demographic groups or industries. Nevertheless, in most cases researchers pick specific subgroups (e.g. teens) or sectors (e.g. restaurants) not because they are the main subjects of interest, but because these are subgroups where the fraction of minimum wage workers is high. Furthermore, the prediction approach can be also applied if someone is specifically interested in the impact of the minimum wage on some subgroups (see Table 5).

between the dark triangle and the curve). For the recall value of the “HSL, Age<30” sample, however, the learning tools achieve a substantially larger precision value than that of the former sample. Therefore, the figure highlights that the learning tools improve the precision-recall trade-off considerably, especially if the aim is to include non-teen observations.

### **4.3 Who are the Minimum Wage Workers?**

Before we study the impact of the minimum wage on labor market outcomes, we examine who are the most likely minimum wage workers according to our (best) performing prediction model. To do this, we examine the main characteristics of individuals in various predicted probability deciles. Table ?? shows the share of workers with various demographic characteristics in each column, while predicted probabilities deciles are in descending order (starting with the highest probability group).

Examining the table, we observe that age is a highly important predictor. Teens constitute 70% of the workers in the decile most likely to be exposed to the minimum wage. This suggests that an analysis using only the highest decile would be very similar to an analysis based just on the teen sample. However, the teen share drops to 5.6% in decile 9 and is virtually zero thereafter, and more than 35% of the observations in deciles 9 and 8 are adults of age 20-30. This suggests that the importance of the age variable is not limited to determining whether an observation belongs to the highest decile. The model also employs educational attainment heavily in determining the predicted probabilities. Observations with less than high school education are mostly in the highest deciles and high school graduates are concentrated on the top half of the exposed individuals. There is almost no observation without some college education in the lowest deciles. Another finding worth mentioning is that the share of women workers is high in the top deciles (e.g. 66.0% in the 9th decile), and is lower in the bottom deciles (36.9% in the least likely decile), indicating that individual’s gender also play an important role. At the same time, the race variable appears to play a smaller role in determining if an individual is in the top or the bottom deciles. Lastly, Hispanic individuals’ share in the least likely decile is virtually zero, whereas they make up around 20% in the top two deciles.

An alternative way to examine who are the minimum wage workers is to consider the relative

importance of each predictor. In Figure 4 we plot the “relative influences” of the variables calculated following Friedman (2001).<sup>12</sup> The figure largely confirms our previous observations. It finds “age” as the most important predictor in the sample with a very large margin. The variable for educational credentials comes after age.<sup>13</sup> Gender variables are also relatively important in the prediction. The indicator variables for Hispanic, rural, race, and veteran status appear to have less influence on the prediction.

#### 4.4 High-Probability and High-Recall Groups

While we study responses to the minimum wage throughout the whole predicted probability spectrum, we will mainly focus on two subgroups in the baseline specifications. We follow the definition of Card and Krueger (1995) and define the high-probability group as the individuals with the 10% highest predicted probability. When we use the boosted tree prediction model, the threshold probability that we need to apply to get this group is 39% – all individuals with predicted probability above that value are in the high-probability group. At this threshold the precision rate is almost 60%, which means that almost 60% of the individuals in the high-probability group are indeed minimum wage workers, while the recall rate is 28%, which means that the group covers around 28% of all minimum wage workers.

Since the high-probability group covers less than 30% of all minimum wage workers, we also study the impact of the minimum wage on a more broadly defined group. In the high-recall group we set the threshold probability such that 75% of all minimum wage workers are captured. To do so we need to set the threshold at 11% and at that level we achieve a 33% precision rate. This group covers little less than 50% of all workers in the data. To study the impact of the policy on workers unlikely to be affected by the policy we also define a group for whom the predicted probability is

---

<sup>12</sup>The figure shows the reduction in the loss function caused by each variable used in the non-terminal nodes. We normalize the relative influences so that they sum up to 100. The average importance of each variable is;

$$\mathcal{I}_l^2 = \frac{1}{M} \sum_{m=1}^M \mathcal{I}_l^2(T_m)$$

where  $\mathcal{I}_l(T_m)$  is the reduction in the loss function due to the use of variable  $l$  in the non-terminal nodes of tree  $m$ . However, we wish to caution against interpreting them directly. First, the importance is in terms of prediction, not explanation. Second, there are cases where one variable needs to be interacted with another one for high predictive power. In those cases, only one of them is deemed to have a strong influence, whereas both are essential.

<sup>13</sup>In fact, dropping teen observations from the sample decreases the relative importance of the age variable substantially. It renders age to be the close second most important variable in the prediction. The educational credentials variable of the observation becomes the most important predictor.

less than 11%. Throughout the text we will refer to that group as the low-probability group.

## 5 Impact of the Minimum Wage on Labor Market Outcomes

**Identification Strategy.** We estimate the effect of the minimum wage by implementing an event study strategy similar to the baseline specification in [Cengiz et al. \(2019\)](#). We will focus on prominent state-level minimum wage changes occurred between 1979-2019. Our basic regression specification is the following:

$$Y_{st}^g = \sum_{\tau=-3}^4 \beta_{\tau} treat_{st}^{\tau} + \Omega_{st} + \mu_s + \rho_t + u_{st}, \quad (1)$$

where  $Y_{st}^g$  is the the labor market outcome (e.g. employment rate, unemployment rate, participation rate) in state  $s$  and at quarter  $t$  for group  $g$ . As we discussed above we study the impact of the minimum wage on various groups defined by the prediction model such as the high-probability group and the high-recall group.  $treat_{st}^{\tau}$  is a binary variable that takes the value of 1 if the minimum wage increased in  $\tau$  years from date  $t$  in state  $s$ . This definition implies that  $\tau = 0$  represents the first year following the minimum wage increase (i.e. the quarter of treatment and the subsequent three quarters), and  $\tau = -1$  is the year (four quarters) prior to treatment. Our benchmark specification controls for state and period fixed effects,  $\mu_s$  and  $\rho_t$ , and we also include controls for small or federal increases,  $\Omega_{st}$ .<sup>14</sup> We cluster our standard errors by state, which is the level at which policy is assigned.

**Main Results.** Table 2 shows the estimated effects on the high-probability, the high-recall and the low-probability groups. We report five year averaged post-treatment estimates for the key labor market outcomes (wages, employment, unemployment and labor force participation), formally  $\frac{1}{5} \sum_{\tau=0}^4 \beta_{\tau}$ . Columns 1 and 3 highlight that the minimum wage has a significant positive impact on wages at groups of workers predicted to be exposed to the minimum wage. In the high-probability group wages increased by around 2.3% (s.e. 0.2%), while the high-recall group, which captures 75% of the minimum wage workers, experienced a somewhat lower but still significant wage increase (1.6%, s.e. 0.3). At the same time, Column 5 shows no indication for any significant wage

<sup>14</sup>The variables we use to control for federal and small events are the same as the ones employed in [Cengiz et al. \(2019\)](#). We collapse the windows for small and federal events into three periods: EARLY, PRE, and POST. EARLY is for 3 and 2 years before, and PRE is for 1 year before the event. POST is for 0 to 4 years after the event.

change relative to the untreated states at the low-probability group. This highlights that the wage growth was only shared by workers exposed to the minimum wages, and we find no evidence for substantial wage gain for individuals unlikely to be directly exposed to the minimum wage shock.

Columns 2, 4, and 6 of Table 2 show the estimates when we classify workers based on the Card and Krueger's linear probability model. It is worth highlighting that the wage effects are almost the same for the best performing prediction model and for the Card and Krueger's approach. This just highlights that the Card and Krueger model performs very well.<sup>15</sup>

Table 2 also reports the effect of the minimum wage on employment. Considering the high-probability and high-recall groups, we find a small and statistically insignificant positive effect on employment. The employment elasticity with respect to minimum wage is around 0.07 (s.e. 0.08) and 0.03 (s.e. 0.0), respectively. The 95th confidence intervals around these estimates can rule out -0.1 – the lower bound elasticity suggested by Neumark and Wascher (2008) .

Table 2 also highlights that the employment estimates are somewhat smaller for the high-recall than for the high-probability group, which is in line with the wage effects. This leads to a similar elasticity of employment with respect to own wage, which would be the labor demand elasticity in the standard model, in the two groups. When we calculate the employment elasticity with respect to own wage, we obtain an elasticity of 0.28 (s.e. 0.34) for the high-probability group and 0.18 (s.e. 0.22) for the high-recall group. These are quite precise estimates and can rule out a modest negative impact of the policy on employment.

A key advantage of the probability based approach is that we can study outcomes besides employment and wages. In Table 2 we also report the effect of the minimum wage on unemployment rate and on participation rate. For the high-probability group (Column 1), we find a slight decrease in unemployment and a slight increase in the participation rate. The estimates on high-recall group (Column 3) show a similar picture though there is some evidence for a slight increase in the participation rate. However, these changes in unemployment and participation are not statistically significant .

At the same time, the decrease in unemployment and the slight increase in participation rate

---

<sup>15</sup>Nevertheless, the key advantage of applying machine learning tools is that someone can select the predictors in a data-driven way without knowing much about the context. Even if the functional form chosen by Card and Krueger (1995) performs very well, it is unclear how someone with less knowledge about U.S. labor markets could come up with that functional form.

suggest that search effort is unlikely to fall in response to the policy. Furthermore, as we discussed in Section 2, the lack of a considerable drop in participation highlights that the policy cannot have significant negative welfare effects on workers who are on the margin of participating in the labor market. At the same time, our estimates also highlight that the participation margin, emphasized by Flinn (2006), is unlikely to be a major margin of adjustment. Moreover, our results are hard to reconcile models that predict an increase in unemployment rate in response to the policy (see e.g. Drazen, 1986, Lang, 1987, Swinnerton, 1996; van den Berg, 2003).

**Non-linearity in the extent of exposure.** In Table 2 we see that the high-probability group which captures the 10% most exposed individuals, and the high-recall group that covers a broader group comprising 75% of the minimum wage workers provide surprisingly similar estimates across the subgroups studied here. This suggests that the impact of the policy does not seem to depend on non-linearity in the extent of exposure. Figure 5 explores this non-linearity further. We show the estimated impact on the key labor market outcomes by applying various cut-offs to the predicted probability. The green solid lines show the five year averaged post-treatment estimates for individuals whose predicted probability is above the particular predicted probability threshold. As this threshold decreases, we also naturally decrease the precision as our group will include more non-minimum wage workers in the sample. At the same time, we also capture a larger fraction of minimum wage workers and so we attain a higher recall rate.

Panel (a) shows that the wage effects decline almost monotonically as we lower the minimum predicted probability. This is what we expect given the declining precision rate: as we lower the threshold our sample will cover more and more non-minimum wage workers whose wages are unaffected (directly) by the policy. At the same time, Panel (b) shows that employment responses, which start from a small positive effect, decline as we lower the precision. As a result, when we divide the employment effects by the wage effects, we obtain a stable employment elasticity with respect to wage.

Furthermore, Panel (c) shows that the unemployment rate declines at the high-probability group but that decline shrinks as we increase the recall rate. . At the same time, the change in participation is close to zero and unrelated to the change in thresholds (see Panel (d)). Overall, the graphical evidence on Figure 5 finds no clear indication for non-linearities in response to the policy.

**Timing of the Impact.** Figures 6 and 7 show the impact of the minimum wage on various labor

market outcomes over time for the high-recall and the high-probability groups respectively. In the figure we plot  $\beta_\tau$  expressed relative to the event date  $\tau = -1$ , or the year just prior to treatment. Panel (a) shows the evolution of wages around the minimum wage increase. There is a clear increase in wages just in line with the timing of the minimum wage increase, and is not driven by some pre-existing trends. Over time the wage effects are attenuated, which reflect that the most recent minimum wage changes tend to be larger.<sup>16</sup>

Panel (b) shows the impact of the policy on employment around the minimum wage hike. For both the high-recall and the high-probability group we see a similar pattern: there is no clear evidence for pre-existing trends in employment, although there is a slight dip in employment 2 years before the minimum wage increase when we look at the high recall group (Figure 6). Nevertheless, there is no unusual unemployment change taking place if we look at the longer change between 1 year and 3 years preceding the minimum wage increase. Furthermore, the small drop in employment between 1 and 2 year preceding the minimum wage increase would imply that the economy slightly deteriorates before an average minimum wage hike, and so we would expect a decrease in employment rate after the policy change. In contrast, we see no clear break in employment after the policy change—if anything, there is a small, statistically significant increase.

Panel (c) shows the impact on unemployment. There is no clear evidence of any pre-existing trends, and no break after the minimum wage increase. Panel (d) shows the impact of the policy on participation rate. Overall, the pre-existing trends are reassuring, although we see a similar pattern for employment. At the same time, we find no indication of any break in the participation rate, which is more or less stable in the post policy years.

**Impact by predicted probability quintiles.** Figure A.2 shows the effect of the minimum wage for each predicted probability quintile separately. Panel (a) shows that there is a clear wage effect among the individuals most exposed to the minimum wage. Wages also seems to increase for the second highest quintile of predicted probabilities, though this increase is not statistically significant. Panels (c)-(d) show the effect of the policy on labor market outcomes such as employment, unemployment and participation. None of these outcomes show any substantial

---

<sup>16</sup>Since we do not fully observe the impact of the policy 5 years after the minimum wage hike for the most recent events, those events only impact the estimates in earlier post treatment years. In Appendix Figure A.2 and A.3 we assess the timing of the policy when we focus only on minimum wage changes where we see responses for the entire event window. For these events we find no decline in wage effects over time.

change even at the highest predicted probability quintile.

**Robustness.** In Table 3 we assess the robustness of the main results shown in Table 2 to the inclusion of various versions of time-varying heterogeneity for the high-recall group, while we report the same robustness checks for the high-probability group in Table A.1. In Column (1) we report the estimates for the baseline specification shown in Table 2. Columns (2) and (3) add state specific linear and quadratic time trends, respectively, while Columns (4)-(5) additionally allow the period effects to vary by the 9 Census divisions. Column (5) represents a highly saturated model allowing for state-specific quadratic time trends and division-period effects. The results in all these specifications are very similar to the baseline specification: we find a positive and statistically significant wage effect; a small positive (non-significant) employment effect, which comes from a drop in unemployment rate; and an increase in the participation rate.

So far we have only focused on prominent state-level minimum wage changes. In Column (6) of Table 3, we expand the event definition to include (nontrivial) federal minimum wage increases. This leads us to 406 minimum wage changes. Inclusion of these federal-level events provides a very similar result. In Column (7) we provide estimates without using population weights. The estimates are similar though the employment estimate without population weight is closer to zero and more noisily estimated. The implied employment elasticity with respect to wage is -0.01 (s.e. 0.32), which can only rule out considerable negative responses to the policy at the conventional significance levels.

Finally, in Column (8) we restrict our analysis to prominent state-level minimum wage changes where we observe all the five years post reform. Since our data ends in 2019, we include only events in this analysis that took place before January 2015. Restricting the sample to these events does not affect the key findings: we obtain a bit larger participation increase (0.2 percentage points instead of 0.1), but the estimates become also a bit more noisy.

**Effects on labor market transitions.** Table 4 shows the estimated impact on monthly labor market transition rates between employment, unemployment, and participation. We find no statistically or economically significant effects on the transition probabilities in response to the policy change for the high-probability (column 1) or the high recall group (column 2). This highlights that the minimum wage increase did not lengthen unemployment durations; if anything, the policy accelerated monthly transition from unemployment to employment by 0.007 (s.e. 0.005)

percentage point for the high probability individuals and by 0.004 (s.e. 0.004) percentage point for the high recall group. Therefore, we find no indication that some individuals were pushed permanently into long term unemployment or that the labor markets became more sclerotic in response to the minimum wages.

**Effects by demographic subgroups.** In some cases, policy makers may be interested in the impact of minimum wages on specific demographic groups. Moreover, individuals are more likely to be on the participation margin in some demographic groups than in others. As a result, the impact of the policy on participation decisions might vary considerably across demographic groups. We study the heterogeneity in response to the minimum wage in Table 5. To make sure that changes in labor market outcomes are driven by the minimum wage itself and not something else, in each subgroup we focus on workers who are in the high-recall group.<sup>17</sup> Restricting the sample to workers and who are likely to be minimum wage worker is also necessary for getting first stage wage effects, which would not be possible if we had all workers in the sample (see Table A.1. in the Online Appendix of [Cengiz et al. \(2019\)](#)).

For most subgroup in Table 5 we find a clear and significant impact on wages, which highlights the key advantage of restricting the sample to groups that are likely to be exposed to the minimum wage. In Column (1) we simply report the estimates of our benchmark analysis on the overall high-recall group for comparison. In Column (2) we show the estimates on workers who are black or Hispanic.<sup>18</sup> The point estimates indicate a drop in employment and participation, but the estimates are noisy. This suggests that in the black or Hispanic group, some workers (who were at the participation margin) may have been made worse off by the policy. However, the standard errors are too large to make a definitive conclusion on this question, and we also cannot rule out a small positive employment effect.

In Column (3) and (4), we examine the impact of the policy on all women and on married women, respectively. Since the extensive margin labor supply elasticities are often found to be larger

---

<sup>17</sup>We use the predicted probabilities estimated on all workers. One could estimate a separate prediction model for each subgroup and then use those predicted probabilities. Nevertheless, it is unlikely that there are substantial gains from estimating a separate prediction model for each subgroups. If there were substantial gains from using differently some predictors within a sub-group, then the boosting tree model would incorporate that in the prediction model even if it applied in the overall sample. We also explored for some demographic groups whether estimating a separate prediction model helps, but we find negligible changes in the precision of our estimates.

<sup>18</sup>We tried to estimate the effect of the policy on black and Hispanics separately, but the estimates were too imprecise to be informative.

for women than for men, it might be that case that minimum wages lead to a larger increase in women's participation. Furthermore, married women are also the ones who are typically thought to have larger responsiveness on the extensive margin. However, our estimated effects on the key labor market outcomes for women and married women are very similar to the benchmark specification.

In Column (5) and (6), we study the impact on teens and on older individuals (aged between 60-70), respectively. Both the young and the old have lower participation rate than prime age individuals, and they also have more elastic labor supply [Blundell et al. (2011)]. For teens, we find larger wage effects and slightly larger employment increase than for the overall sample. The increase in employment comes from the changes in participation, which is in line with the idea that more teens are on the participation margin. For older individuals, we find no significant wage effects (though the point estimates are close the overall sample). This makes interpreting the results for older individuals difficult. Still, we find similar responses for them as for the teens in terms of changes in employment and participation rate.

Columns (7) and (8) study the impact on those with lower educational credentials. The labor market impact of the policy on these education groups are very similar to the impact of the policy on the overall sample. These findings suggest that workers with less education seem to benefit from the minimum wage policies.

Table 6 presents additional heterogeneity analysis. Column (2)-(4) show the estimates by the predicted probability of moving in or out from the labor force. We predict the probability that an individual changes their labor market participation status—either from non-participation to participation,—or vice versa applying the boosting tree ML method. We use in this prediction model all the demographic variables that we used before to predict minimum wage exposure, but we also add the number of children, since it is likely to be an important predictor of labor force participation. The relative importance of the predictors shown in Appendix Figure A.4. Similar to the prediction model on minimum wage exposure, age, education and gender are the most important three predictors of changes in participation. In addition, race and number of children also influence substantially the prediction model.

Since the number of children is coded consistently only after 1990 in the CPS, we report estimates on the post 1990 period in Table 6. Column (1) report estimates for all workers using only

that period. Column (2) show the estimates for the group of workers, which has high predicted probability of changing participation rate. The estimated impact for this group and for the overall sample is very similar. Column (3) and (4) show the estimates with lower probability of switching. Again we find similar responses as in the overall sample. In Column (5) and (6) we also report estimates on single and married mothers. In both cases, we find a slight positive (though statistically insignificant) increase on participation rate.

**Additional labor market outcomes.** Finally, in Table 7 we study the impact on further labor market outcomes such as being self-employed, part-time worker (working less than 30 hours per week), or over-time worker (working more than 40 hours per week). Changing working hours or pushing workers to self-employment are often argued to be important margins of adjustment to the minimum wage. At the same time we find little indication that the high-probability or the high-recall groups experienced any changes in these important outcomes. There is a notable decline in part-time jobs for the high-probability group that is statistically significant at the 95th percentile.

## 6 Conclusion

In this paper, we considered the impact of the minimum wage on various labor market outcomes by exploiting 172 prominent minimum wage changes between 1979 and 2019. To provide new estimates on various labour market outcomes, we extended the prediction probability approach that was proposed by [Card and Krueger \(1995\)](#). We applied modern machine learning techniques to estimate the likelihood that someone is a minimum wage worker. This strategy allowed us to study the impact of the minimum wage not just on wages and employment, but also on unemployment and participation rate for groups that cover 75% of all minimum wage workers. These groups include substantially more minimum wage workers than the demographic-based and industry based subsamples commonly used in the literature.

In line with much of the existing evidence in the literature, we find that the minimum wage has a positive and significant impact on wages, while employment effects are modest in the U.S. context. We also show that the slight (statistically insignificant) employment increase comes from a slight drop in unemployment and a slight increase in the participation rate. These responses indicate that the minimum wage is unlikely to have a negative impact on workers by discouraging

them to search for new jobs.

We also find no significant heterogeneities in the responses to the minimum wage. The most likely exposed group and a much boarder group that covers 75% of all minimum wage workers responded very similarly to the policy change. Overall, our results underscore the positive impact of the policy on key labor market outcomes. Furthermore, the approach developed here could be naturally applied to study the impact of the policy on various other important outcomes measured in different data sources. Since the most important predictors (age, education, gender) are available in most data sources, it is straightforward to assess the impact of the minimum wage hikes along the predicted probability spectrum (and show estimates like in Figure 5). A key advantage of doing so is to study directly whether there is some non-linearity in response to the minimum wage by the extent of exposure. Furthermore, studying the impact of the policy on the low-probability group can serve as an additional falsification test and can provide further support for the credibility of the research design.

## References

- Adams, Camilla, Jonathan Meer, and CarlyWill Sloan. 2018. "The Minimum Wage and Search Effort," NBER Working Paper No. 25128.
- Addison, John T, McKinley L Blackburn, and Chad Cotti. 2011. "Minimum wage increases under straightened circumstances."
- Addison, John T and McKinley Blackburn. 1999. "Minimum wages and poverty," *Industrial & Labor Relations Review*, 52(3): 393–409.
- Ahn, Tom, Peter Arcidiacono, and Walter Wessels. 2011. "The Distributional Impacts of Minimum Wage Increases When Both Labor Supply and Labor Demand Are Endogenous," *Journal of Business and Economic Statistics*, 29(1): 12–23.
- Allegretto, Sylvia, Arindrajit Dube, Michael Reich, and Ben Zipperer. 2017. "Credible research designs for minimum wage studies: A response to Neumark, Salas, and Wascher," *ILR Review*, 70(3): 559–592.
- Belman, Dale and Paul J Wolfson. 2014. *What does the minimum wage do?:* WE Upjohn Institute.
- Belman, Dale, Paul Wolfson, and Kritkorn Nawakitphaitoon. 2015. "Who Is Affected by the Minimum Wage?" *Industrial Relations: A Journal of Economy and Society*, 54(4): 582–621.
- Blundell, Richard, Antoine Bozio, and Guy Laroque. 2011. "Labor Supply and the Extensive Margin," *The American Economic Review*, 101(3): 482–486.
- Borgschulte, Mark and HeePyung Cho. 2019. "Minimum Wages and Retirement," *ILR Review*(73(1)): 153–177.
- Breiman, Leo. 2001. "Random forests," *Machine learning*, 45(1): 5–32.
- Card, David. 1992. "Do minimum wages reduce employment? A case study of California, 1987–89," *ILR Review*, 46(1): 38–54.
- Card, David and Alan B. Krueger. 1995. *Myth and Measurement: The New Economics of the Minimum Wage*, New Jersey: Princeton University Press.

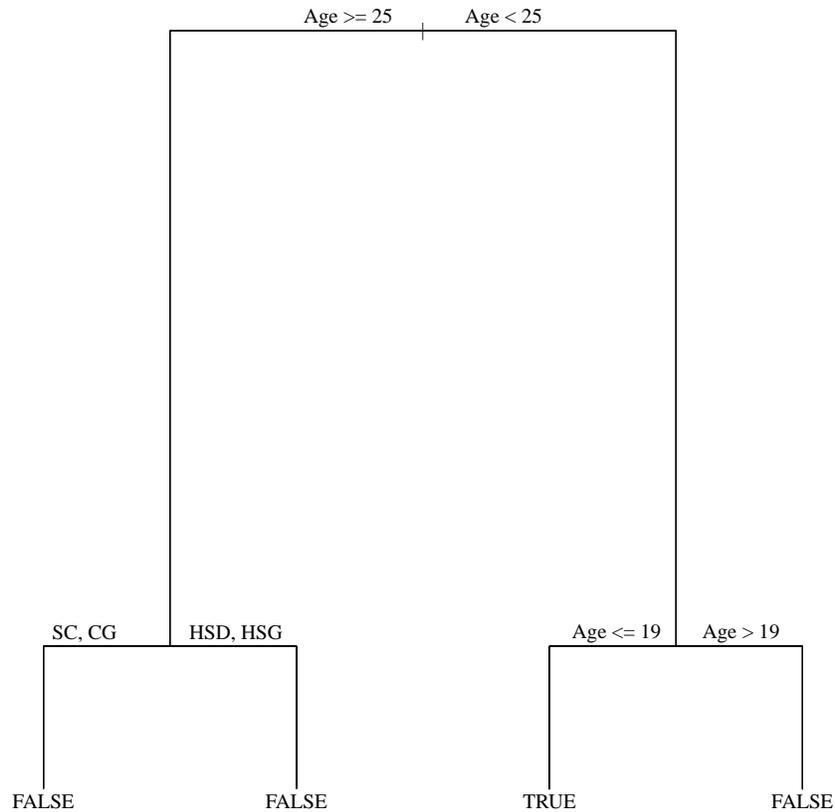
- Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs\*," *The Quarterly Journal of Economics*, 134(3): 1405–1454.
- Clemens, Jeffrey, Lisa B. Kahn, and Jonathan Meer. 2018. "The Minimum Wage, Fringe Benefits, and Worker Welfare," Working Paper 24635, NBER.
- Clemens, Jeffrey and Michael R Strain. 2017. "Estimating the Employment Effects of Recent Minimum Wage Changes: Early Evidence, an Interpretative Framework, and a Pre-Commitment to Future Analysis," NBER Working Paper No. 23084.
- Clemens, Jeffrey and Michael Wither. 2016. "The Minimum Wage and the Great Recession: Evidence of Effects on the Employment and Income Trajectories of Low-Skilled Workers."
- Dube, Arindrajit, T William Lester, and Michael Reich. 2016. "Minimum wage shocks, employment flows, and labor market frictions," *Journal of Labor Economics*, 34(3): 663–704.
- Dustmann, Christian, Attila Lindner, Uta Schoenberg, Matthias Umkehrer, and Philipp vom Berge. 2020. "Reallocation Effects of the Minimum Wage," CReAM Discussion Paper Series 2007, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.
- Feenberg, Daniel and Jean Roth. 2007. "CPS labor extracts 1979–2006."
- Flinn, Christopher J. 2002. "Interpreting Minimum Wage Effects on Wage Distributions: A Cautionary Tale," *Annales d'Economie et de Statistique*(67/68): 309–355.
- 2006. "Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contact Rates," *Econometrica*, 74(4): 1013–1062.
- 2011. *The minimum wage and labor market outcomes*: MIT press.
- Friedman, Jerome H. 2001. "Greedy function approximation: a gradient boosting machine," *Annals of statistics*: 1189–1232.
- 2002. "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, 38(4): 367–378.

- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*: New York, NY: Springer-Verlag New York.
- Godoy, Anna, Michael Reich, and Sylvia A. Allegretto. 2020. "Parental Labor Supply: Evidence from Minimum Wage Changes," IRLE Working Paper No. 103-19.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An introduction to statistical learning*, 6: Springer.
- Laws, Athene. 2018. "Do minimum wages increase search effort?", Cambridge Working Papers in Economics CWPE1857.
- Lundstrom, Samuel M. 2016. "When is a Good Time to Raise the Minimum Wage?" *Contemporary Economic Policy*.
- Manning, Alan. 2016. "The elusive employment effect of the minimum wage."
- Meer, Jonathan and Jeremy West. 2015. "Effects of the minimum wage on employment dynamics," *Journal of Human Resources*.
- Neumark, David. 2017. "The Employment Effects of Minimum Wages: Some Questions We Need to Answer," NBER Working Paper No. 23584.
- Neumark, David, JM Ian Salas, and William Wascher. 2014. "Revisiting the Minimum Wage-Employment Debate: Throwing Out the Baby with the Bathwater?" *ILR Review*, 67(3\_suppl): 608–648.
- Neumark, David and William Wascher. 1992. "Employment effects of minimum and subminimum wages: panel data on state minimum wage laws," *ILR Review*, 46(1): 55–81.
- Neumark, David and William L. Wascher. 2008. *Minimum Wages*, Cambridge, MA: MIT Press.
- Sabia, Joseph J, Richard V Burkhauser, and Benjamin Hansen. 2012. "Are the effects of minimum wage increases always small? New evidence from a case study of New York state," *ILR Review*, 65(2): 350–376.
- Segal, Mark R. 2004. "Machine learning benchmarks and random forest regression," *Center for Bioinformatics & Molecular Biostatistics*.

- Totty, Evan. 2017. "The effect of minimum wages on employment: A factor model approach," *Economic Inquiry*, 55(4): 1712–1737.
- Vaghul, Kavya and Ben Zipperer. 2016. "Historical state and sub-state minimum wage data," *Washington Center for Equitable Growth Working Paper*.
- Zou, Hui and Trevor Hastie. 2005. "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2): 301–320.

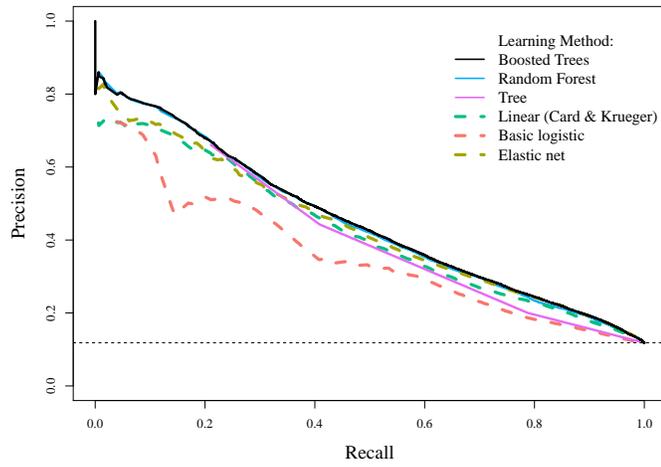
## Figures

Figure 1: Minimum Wage Workers According to Pruned Trees

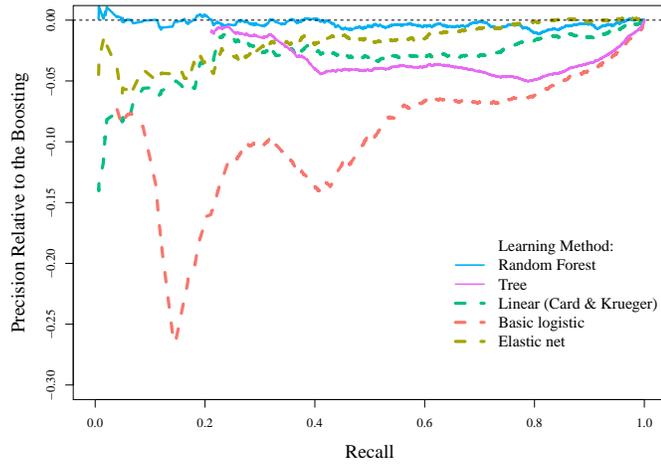


Notes: The figure plots a pruned decision tree produced to predict whether an individual is a minimum wage worker (has an hourly wage of less than 125% of the statutory minimum wage) using the demographic and educational characteristics. In the beginning, the tree tries every possible split to divide the entire sample space into two, and picks the one that diminishes the loss function the most. Then, each subspace is treated as the new feature space and the first step is repeated. "TRUE" indicates that the tree predicts workers in the terminal node are minimum wage workers and "FALSE" otherwise. While the tree explores characteristics such as gender, marital status, veteran status, and rural residency status, it only picks age and education to make splits. LTHS indicate workers with no high school degree, HSG with no college education, SC with no college degree and CG represents college-graduate workers.

Figure 2: Precision-Recall Curves



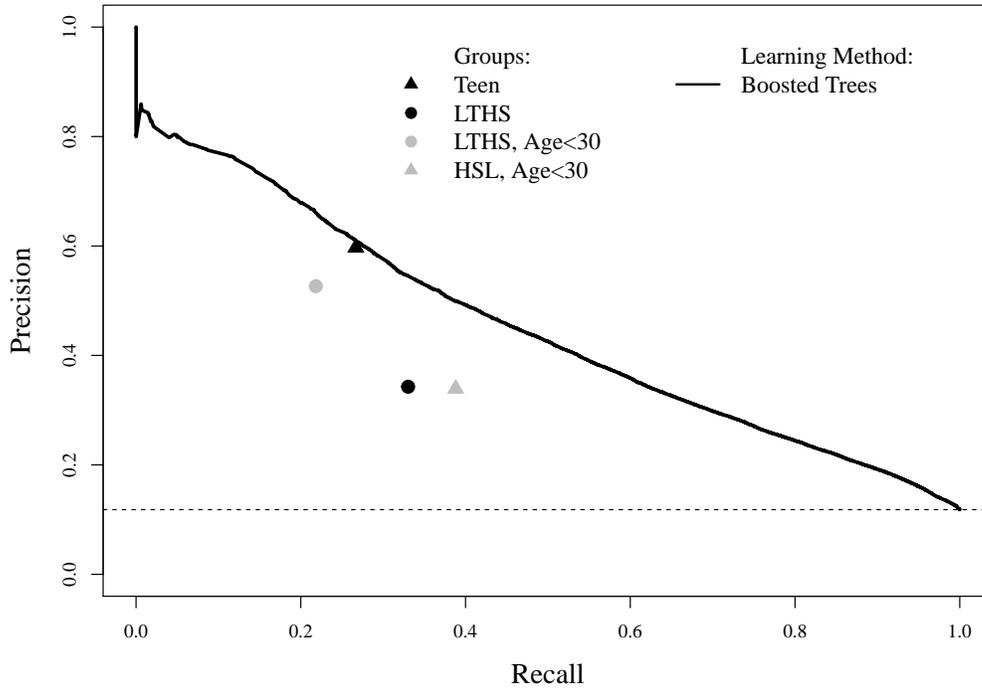
(a) Precision-Recall Curves



(b) Precision-Recall Curves Relative to Boosting Tree Model

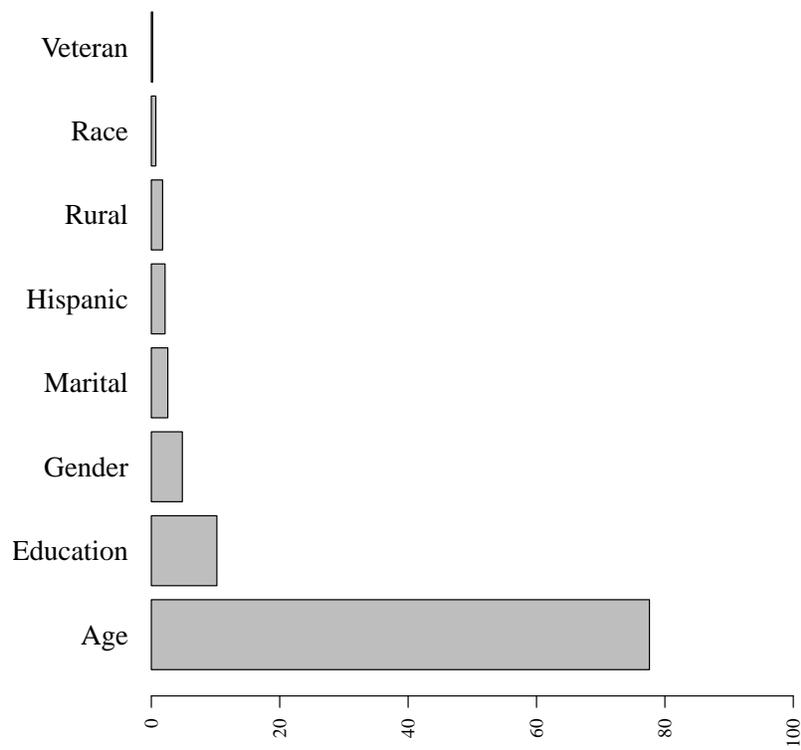
Notes: Panel (a) plots the precision recall-curves for various prediction models described in Section 4.1 and for a basic logistic model which we estimate using (linear) age and categorical education variables. We obtain the precision-recall curves in the following way: we use our prediction model to calculate the probability that someone is a minimum wage worker and we assign all individuals to the predicted group if that probability is above a certain threshold. The figure shows the estimated precision and recall rates obtained when we vary the threshold value. The horizontal dashed line shows the average share of minimum wage workers in the sample. The areas below the precision-recall curves are 0.45 for boosting tree, 0.44 for random forest, 0.43 for elastic net, 0.42 for the Card and Krueger’s linear probability model, 0.33 for the basic logistic, and 0.27 for the single tree. Panel (b) shows the difference in precision rate between the best performing one – the boosted tree – and the other models at each recall rate.

Figure 3: Performance of Boosting Tree Relative to Demographic Subgroups



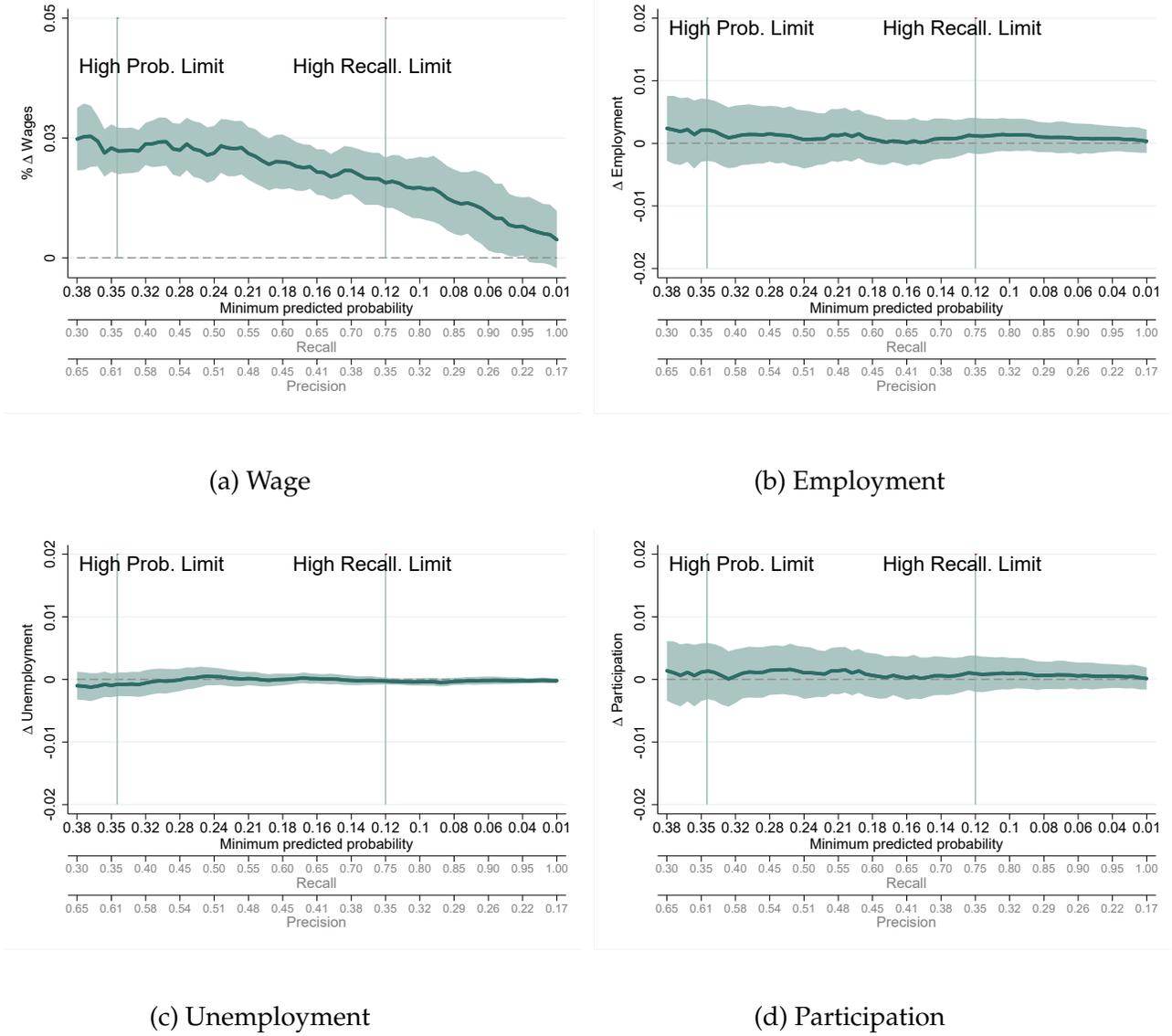
Notes: The figure compares the performance of the best performing prediction model – the boosted tree – relative to the strategy of choosing specific subgroups to proxy minimum wage workers. The black solid line shows the precision recall-curves for the boosted tree model. The black triangle shows the precision and recall level for teens, the black circle for individuals with less than high school (LTHS), the gray circle for individuals with less than high school and younger than 30 and the gray triangle for individuals with high school or less and younger than 30, and the plus signs for the restaurant workers. The horizontal dashed line shows the average share of minimum wage workers in the sample.

Figure 4: Relative Influences of the Predictors in the Boosted Tree Prediction Model



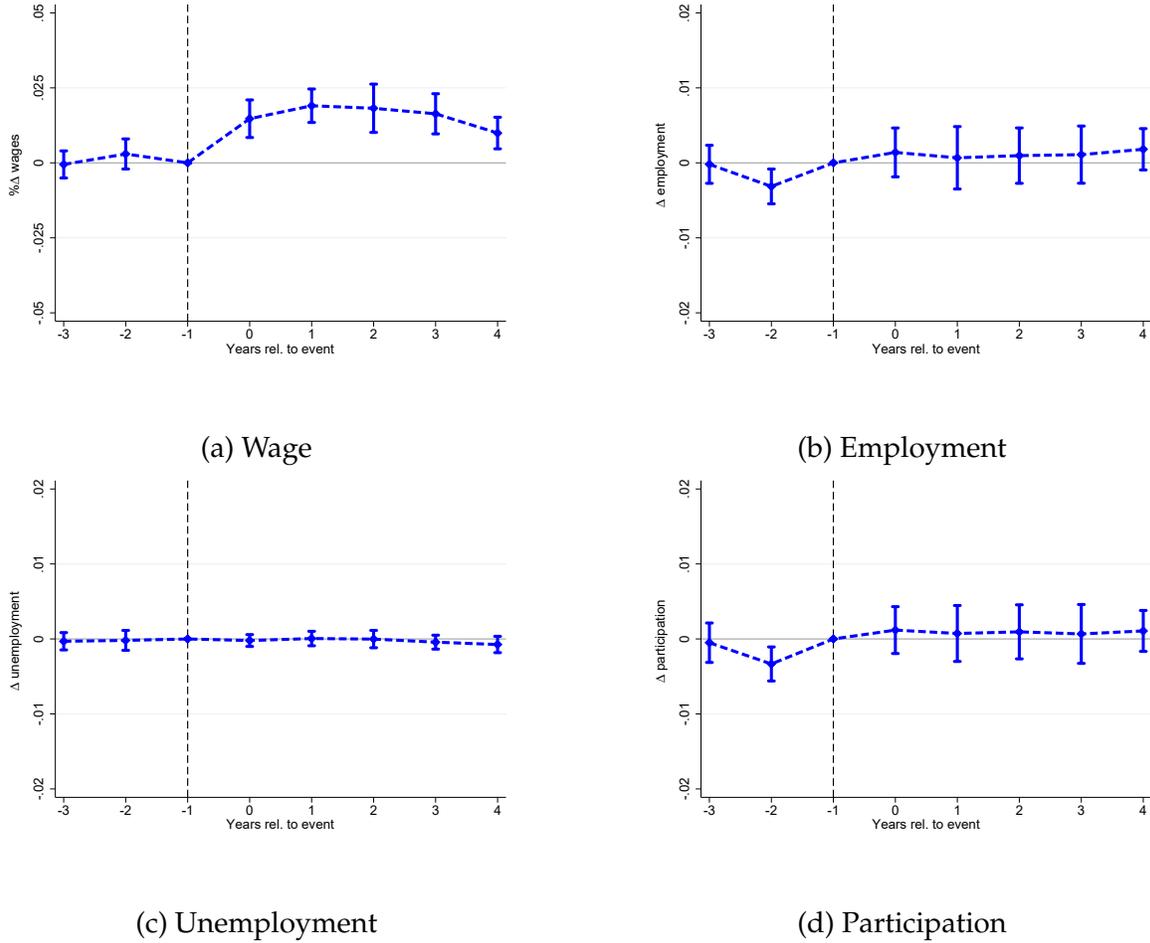
Notes: We plot the relative influences of the variables in the best performing prediction model – the boosted tree model – calculated as in [Friedman \(2001\)](#) (see footnote 13 for the details). The bars, which indicate the decline in the loss function associated with the corresponding variable, are normalized so that they sum up to 100.

Figure 5: Impact of the Minimum Wage for Alternative Predicted Probability Threshold Values



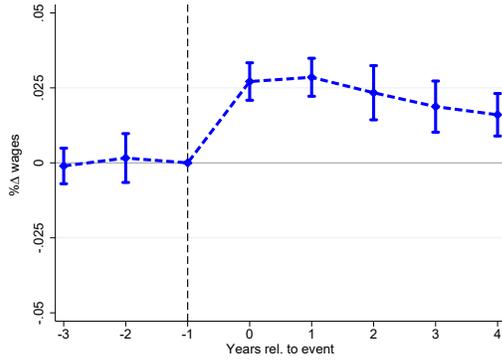
Notes: The figure shows the main results from our event study analysis (see equation 1) using alternative predicted probability threshold values. We exploit 172 state-level minimum wage changes between 1979-2019. Panel (a) shows the impact of the minimum wage on wages, Panel (b) on employment to population, Panel (c) on unemployment to population, and Panel (d) on participation rate. In each panel the green solid line shows the five year averaged post-treatment estimates for individuals whose predicted probability is above the “minimum predicted probability threshold”. On the x-axis we also report the corresponding recall rate (the fraction of minimum wage workers retrieved by the prediction model if the particular minimum predicted probability threshold is applied) and the precision rate (the fraction of minimum wage workers in the predicted group if the particular minimum predicted probability threshold is applied). We also plot the thresholds corresponding to the high-probability group capturing the 10% of the population with the highest predicted probability and to the high-recall group capturing 75% of all minimum wage workers. To calculate the predicted probabilities we use the best performing prediction model – the boosted tree prediction model. The shaded areas show the 95% confidence interval based on standard errors that are clustered at the state level.

Figure 6: Impact of the Minimum Wage Over Time, High-Recall Rate Group

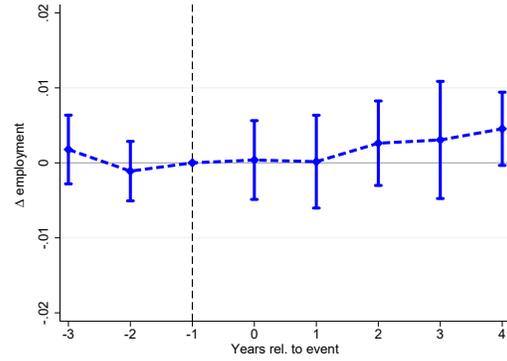


Notes: The figure shows the main results from our event study analysis (see equation 1) exploiting 172 state-level minimum wage changes between 1979-2019. The figure shows the effect of a minimum wage increase on wages (Panel (a)), on employment to population (Panel (b)), on unemployment to population (Panel (c)) and on labor force participation rate (Panel (d)) for the high-recall group. The high-recall group consists of all workers whose predicted probability is above 11.1% – a threshold which corresponds to a 75% of recall rate. To calculate the predicted probabilities we use the best performing prediction model – the boosted tree prediction model. We also show the 95% confidence interval based on standard errors that are clustered at the state level.

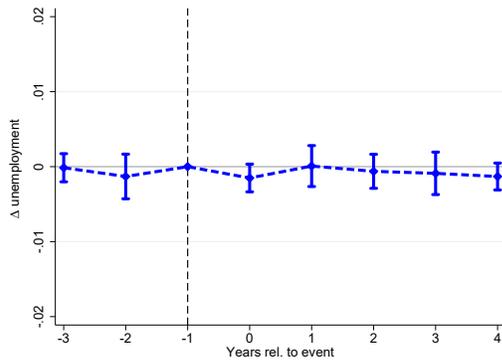
Figure 7: Impact of the Minimum Wage Over Time, High-Probability Group



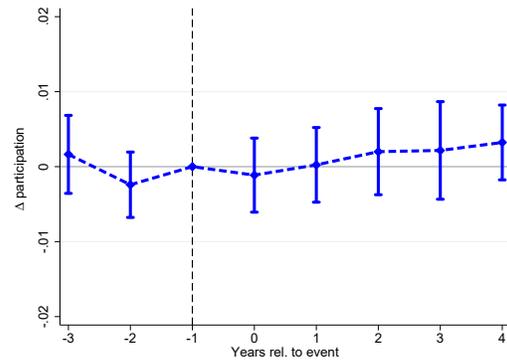
(a) Wage



(b) Employment



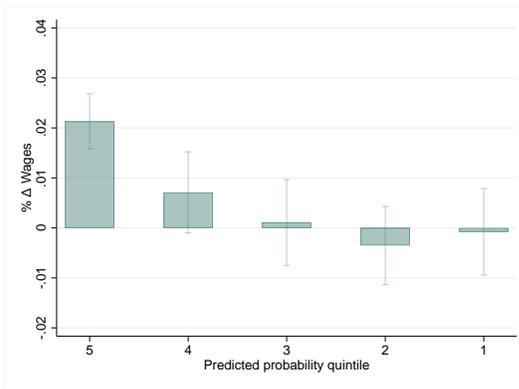
(c) Unemployment



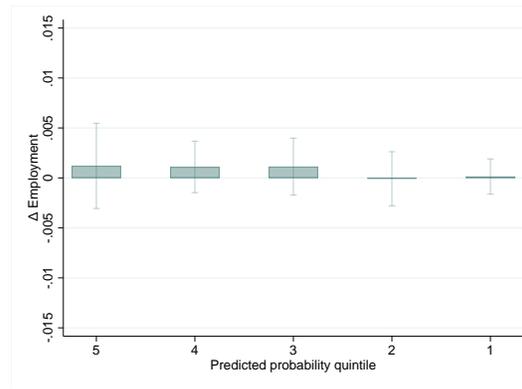
(d) Participation

Notes: The figure shows the main results from our event study analysis (see equation 1) exploiting 172 state-level minimum wage changes between 1979-2019. The figure shows the effect of a minimum wage increase on wages (Panel (a)), on employment to population (Panel (b)), on unemployment to population (Panel (c)) and on labor force participation rate (Panel (d)) for the high-probability group. The high-probability group consists of 10% of the population with the highest likelihood of being affected by the policy. To calculate the predicted probabilities we use the boosted tree prediction model. We also show the 95% confidence interval based on standard errors that are clustered at the state level.

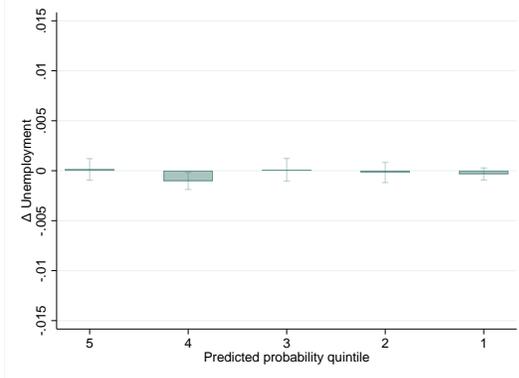
Figure 8: Impact of the Minimum Wage by Predicted Probability Quintiles



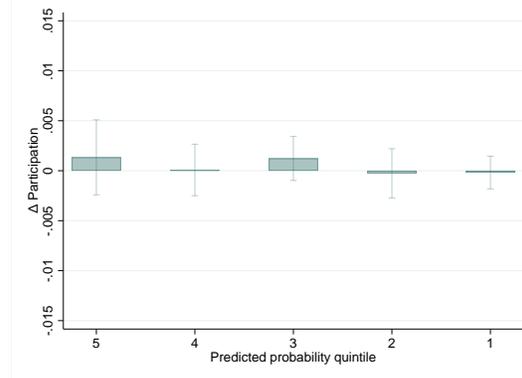
(a) Wage



(b) Employment



(c) Unemployment



(d) Participation

Notes: The figure shows the effect of the minimum wage for each predicted probability quintiles separately. The highest quintile 5 shows the impact of individuals with predicted probability of being minimum wage worker in the top 20%. We estimate equation (1) for each quintiles separately and report the five year averaged post-treatment estimates. We exploit 172 state-level minimum wage changes between 1979-2019. The figure shows the effect of a minimum wage increase on wages (Panel (a)), on employment to population (Panel (b)), on unemployment to population (Panel (c)) and on labor force participation rate (Panel (d)). We also show the 95% confidence interval based on standard errors that are clustered at the state level.

## Tables

Table 1: Demographic Characteristics for Each Predicted Probability Decile

	(1) Teen	(2) 20 ≤ Age <30	(3) LTHS	(4) HSG	(5) Female	(6) White	(7) Black	(8) Hispanic
Most likely decile	0.702	0.128	0.728	0.158	0.636	0.813	0.129	0.174
Probability decile 9	0.056	0.413	0.569	0.220.	0.660	0.834	0.114	0.250
Probability decile 8	0.001	0.366	0.394	0.409.	0.556	0.821	0.125	0.189
Probability decile 7	0.000	0.290	0.151	0.625.	0.541	0.859	0.097	0.169
Probability decile 6	0.000	0.166	0.099	0.628.	0.649	0.861	0.100	0.056
Probability decile 5	0.000	0.143	0.056	0.518.	0.543	0.812	0.130	0.107
Probability decile 4	0.000	0.156	0.057	0.248.	0.461	0.797	0.142	0.088
Probability decile 3	0.000	0.141	0.004	0.308.	0.412	0.873	0.078	0.068
Probability decile 2	0.000	0.099	0.000	0.110.	0.378	0.850	0.090	0.027
Least likely decile	0.000	0.006	0.000	0.028.	0.369	0.734	0.132	0.004

*Notes:* The table shows some demographic characteristics for each predicted probability decile. The predicted probability refers to the probability that an individual have an hourly wage lower than 125% of the minimum wage preceding the minimum wage hike, and is calculated based on the best performing prediction model - the boosted tree prediction model. Each row shows the average characteristics of individuals in the particular predicted probability decile. The top (bottom) row shows the characteristics at the top (bottom) decile, which consists of individuals that are most (least) likely exposed to the minimum wage according to our prediction model. Each cell shows the share of the selected demographic group: column (1) the share of teens (younger than 20); column (2) the share of individuals older than (included) 20 and younger than 30 years old; column (3) the share of whites; column (4) the share of blacks; column (5) the share of Hispanic, column (6) the share of females; c column (7) the share of individuals with less than high school education (LTHS); and column (8) the share of high school graduates with no college education (HSG).

Table 2: Impact of the Minimum Wage on Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta$ wage (%)	0.023*** (0.002)	0.025*** (0.004)	0.016*** (0.003)	0.014*** (0.003)	-0.003 (0.003)	-0.002 (0.003)
$\Delta$ employment (% pt)	0.002 (0.003)	0.001 (0.002)	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	-0.001 (0.001)
$\Delta$ unemployment (% pt)	-0.001 (0.001)	-0.001 (0.001)	-0.000 (0.000)	-0.001* (0.000)	-0.000 (0.000)	0.000 (0.000)
$\Delta$ participation (% pt)	0.001 (0.002)	0.001 (0.002)	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	-0.001 (0.001)
Employment Elas. w.r.t Min. Wage	0.067 (0.079)	0.036 (0.066)	0.030 (0.037)	0.028 (0.032)	0.003 (0.012)	-0.011 (0.012)
Employment Elas. w.r.t Wage	0.280 (0.338)	0.138 (0.245)	0.178 (0.218)	0.191 (0.211)	-0.106 (0.477)	0.567 (0.927)
Number of events	172	172	172	172	172	172
Number of observations	7,854	7,854	7,854	7,854	7,854	7,854
Number of individuals in the sample	571,091	527,106	1,909,060	2,198,894	2,239,793	2,025,324
Mean employment	0.338	0.399	0.424	0.439	0.785	0.807
Mean unemployment	0.057	0.064	0.045	0.043	0.028	0.030
Mean participation	0.395	0.462	0.469	0.481	0.813	0.837
Group	High Prob.	High Prob.	High Recall	High Recall	Low Prob.	Low Prob.
Prediction Model	Boosted Tree	CK Linear	Boosted Tree	CK Linear	Boosted Tree	CK Linear

*Notes:* The table reports the effects of the minimum wage on labor market outcomes based on the event study analysis (see equation 1) exploiting 172 state-level minimum wage changes between 1979 and 2019. The table reports five year averaged post-treatment estimates for each key labor market outcome: percent change in wages and the change in employment to population, unemployment to population, and labor force participation rate. We also report the employment elasticity with respect to the minimum wage and the employment elasticity with respect to the wage, which is the ratio of the percent change in employment and wage. To calculate the percent change in employment we divide the change in employment to population by the mean employment to population rate preceding the minimum wage hikes (reported at the bottom of the table). The line on the number of observations shows the number of quarter-state cells used for estimation, while the number of individuals refers to the underlying CPS sample used to calculate labor market outcomes in these cells. Columns (1) and (2) show estimates for the high-probability group, which captures 10% of the population with highest predicted probability. Columns (3) and (4) show estimates for the high-recall group, which consists of individuals whose predicted probability is above 11.1% - a threshold which leads to a 75% recall rate of minimum wage workers. Columns (5) and (6) show the estimates for workers whose predicted probability is below 11.1%. Columns (1), (3), and (5) use the best performing prediction model — the boosted tree prediction model. Columns (2), (4), and (6) use the Card and Krueger’s linear prediction model. All regressions are weighted by state-quarter population. Robust standard errors in parentheses are clustered by state; significance levels are \* 0.10, \*\* 0.05, \*\*\* 0.01.

Table 3: Impact of the Minimum Wage on Labor Market Outcomes - Robustness to Alternative Specifications (High-Recall Group)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta$ wage (%)	0.016*** (0.003)	0.016*** (0.003)	0.017*** (0.003)	0.015*** (0.002)	0.014*** (0.003)	0.017*** (0.003)	0.012*** (0.003)	0.015*** (0.005)
$\Delta$ employment (% pt)	0.001 (0.001)	0.001 (0.001)	0.001 (0.002)	0.003* (0.001)	0.002 (0.001)	0.001 (0.002)	-0.000 (0.002)	0.002 (0.002)
$\Delta$ unemployment (% pt)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	0.000 (0.001)	-0.001 (0.000)
$\Delta$ participation (% pt)	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	0.003* (0.002)	0.002 (0.001)	0.001 (0.002)	0.000 (0.001)	0.002 (0.002)
Employment Elas. w.r.t Min. Wage	0.030 (0.037)	0.024 (0.036)	0.019 (0.038)	0.064* (0.035)	0.042 (0.033)	0.030 (0.039)	-0.001 (0.039)	0.049 (0.051)
Employment Elas. w.r.t Wage	0.178 (0.218)	0.142 (0.205)	0.106 (0.200)	0.407* (0.237)	0.279 (0.227)	0.165 (0.205)	-0.007 (0.318)	0.366 (0.318)
Number of events	172	172	172	172	172	406	172	99
Number of observations	7,854	7,854	7,854	7,854	7,854	7,854	7,854	7,854
Number of individuals in the sample	1,909,060	1,909,060	1,909,060	1,909,060	1,909,060	4,569,529	1,909,060	1,224,598
Mean employment	0.424	0.424	0.424	0.424	0.424	0.432	0.427	0.434
Mean unemployment	0.045	0.045	0.045	0.045	0.045	0.048	0.044	0.049
Mean participation	0.469	0.469	0.469	0.469	0.469	0.480	0.471	0.483
Controls:								
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y
State Linear Trends		Y	Y		Y			
State Quadratic Trends			Y		Y			
Division-Quarter FE				Y	Y			
State Federal Events						Y		
Unweighted							Y	
No Events After 2014q1								Y

*Notes:* The table reports the effects of the minimum wage on labor market outcomes based on the event study analysis (see equation 1) exploiting 172 minimum wage changes between 1979 and 2019. We assess the impact of the minimum wage on the high-recall group. The high-recall group consists of individuals whose predicted probability is above 11.1% - a threshold which leads to a 75% recall rate of minimum wage workers. The table reports five year averaged post-treatment estimates for each key labor market outcome: percent change in wages and the change in employment to population, unemployment to population, and labor force participation rate. We also report the employment elasticity with respect to the minimum wage and the employment elasticity with respect to the wage, which is the ratio of the percent change in employment and wage. To calculate the percent change in employment we divide the change in employment to population by the mean employment to population rate preceding the minimum wage hikes (reported at the bottom of the table). The line on the number of observations shows the number of quarter-state cells used for estimation, while the number of individuals refers to the underlying CPS sample used to calculate labor market outcomes in these cells. In all the regressions we use the best performing prediction model — the boosted tree prediction model. The first column employs the preferred event-based estimate. Columns (2)-(5) augment the baseline model with linear or quadratic state trends and/or division-by-period fixed effects as indicated in the table. The sixth column reports estimates using 369 state or federal minimum wage increases. All regressions are weighted by state-quarter population except Column (7), where we report unweighted estimates. Column (8) only considers minimum wage events that happened on or before 2014q1 to ensure a full five year post-treatment period. Robust standard errors in parentheses are clustered by state; significance levels are \* 0.10, \*\* 0.05, \*\*\* 0.01.

Table 4: Impact of the Minimum Wage on Labor Market Transitions

	(1)	(2)	(3)
$\Delta$ E-U flow as a share of employment (% pt)	-0.000 (0.001)	0.000 (0.000)	0.000 (0.000)
$\Delta$ E-I flow as a share of employment (% pt)	-0.001 (0.001)	0.000 (0.001)	0.000 (0.000)
$\Delta$ U-E flow as a share of unemployment (% pt)	0.007 (0.005)	0.004 (0.004)	0.002 (0.003)
$\Delta$ U-I flow as a share of unemployment (% pt)	0.002 (0.005)	0.002 (0.004)	0.001 (0.003)
$\Delta$ I-E flow as a share of inactivity (% pt)	0.001 (0.001)	0.000 (0.000)	-0.000 (0.001)
$\Delta$ I-U flow as a share of inactivity (% pt)	-0.000 (0.001)	-0.000 (0.000)	0.000 (0.001)
Number of events	172	172	172
Number of observations	7,854	7,854	7,854
Number of individuals in the sample	395,172	1,323,720	1,568,256
Mean E-U flow as a share of employment (%)	0.028	0.022	0.008
Mean E-I flow as a share of employment (%)	0.107	0.059	0.017
Mean U-E flow as a share of unemployment (%)	0.230	0.247	0.261
Mean U-I flow as a share of unemployment (%)	0.368	0.292	0.178
Mean I-E flow as a share of inactivity (%)	0.057	0.042	0.061
Mean I-U flow as a share of inactivity (%)	0.036	0.024	0.025
Group	High Prob.	High Recall	Low Prob.
Prediction Model	Boosted Tree	Boosted Tree	Boosted Tree

*Notes:* The table reports the effects of the minimum wage on labor market transition rates based on the event study analysis (see equation 1) exploiting 172 state-level minimum wage changes between 1979 and 2019. The table reports five year averaged post-treatment estimates for percent point changes in: the employment-to-unemployment (E-U) transition rate as a share of employment (row 1), the employment-to-inactivity (E-I) transition rate as a share of employment (row 2), the unemployment-to-employment (U-E) transition rate as a share of unemployment (row 3), the unemployment-to-inactivity (U-I) transition rate as a share of unemployment (row 4), the inactivity-to-employment (I-E) transition rate as a share of inactivity (row 5), the inactivity-to-unemployment (I-U) transition rate as a share of inactivity (row 6), and the share of the employed staying in employment (E-E) (row 7). We also report the mean levels of each of these variables for our sample period. The line on the number of observations shows the number of quarter-state cells used for estimation, while the number of individuals refers to the underlying CPS sample used to calculate labor market outcomes in these cells. Column (1) shows estimates for the high-probability group, which captures 10% of the population with highest predicted probability. Column (2) shows estimates for the high-recall group, which consists of individuals whose predicted probability is above 11.1% - a threshold which leads to a 75% recall rate of minimum wage workers. Column (3) shows the estimates for workers whose predicted probability is below 11.1%. Columns (1), (2), and (3) use the best performing prediction model — the boosted tree prediction model. All regressions are weighted by state-quarter population. Robust standard errors in parentheses are clustered by state; significance levels are \* 0.10, \*\* 0.05, \*\*\* 0.01.

Table 5: Impact of the Minimum Wage on Labor Market Outcomes by Demographic Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta$ wage (%)	0.016*** (0.003)	0.012*** (0.003)	0.014*** (0.004)	0.020*** (0.005)	0.027*** (0.003)	0.018 (0.013)	0.013*** (0.004)	0.013*** (0.003)
$\Delta$ employment (% pt)	0.001 (0.001)	-0.003 (0.003)	0.001 (0.002)	0.001 (0.002)	0.003 (0.003)	0.003 (0.003)	0.001 (0.002)	0.001 (0.001)
$\Delta$ unemployment (% pt)	-0.000 (0.000)	-0.001 (0.001)	-0.001 (0.000)	-0.001 (0.001)	-0.001 (0.002)	-0.001 (0.001)	-0.000 (0.000)	-0.000 (0.000)
$\Delta$ participation (% pt)	0.001 (0.001)	-0.004 (0.003)	0.000 (0.002)	0.001 (0.002)	0.003 (0.002)	0.004 (0.003)	0.000 (0.002)	0.001 (0.001)
Employment Elas. w.r.t Min. Wage	0.030 (0.037)	-0.068 (0.055)	0.023 (0.043)	0.036 (0.050)	0.111 (0.087)	0.125 (0.121)	0.023 (0.058)	0.029 (0.038)
Employment Elas. w.r.t Wage	0.178 (0.218)	-0.533 (0.415)	0.157 (0.294)	0.171 (0.229)	0.391 (0.317)	0.647 (0.590)	0.171 (0.397)	0.209 (0.261)
Number of events	172	172	172	172	172	172	172	172
Number of observations	7,854	7,841	7,854	7,854	7,854	7,854	7,854	7,854
Number of individuals in the sample	1,909,060	650,474	1,140,185	778,935	317,538	204,759	734,380	1,444,741
Mean employment	0.424	0.508	0.399	0.409	0.333	0.270	0.363	0.408
Mean unemployment	0.045	0.061	0.039	0.045	0.068	0.015	0.050	0.048
Mean participation	0.469	0.569	0.438	0.453	0.401	0.285	0.412	0.456
Group	High Recall	High Recall	High Recall	High Recall				
Demog. Group	All	Black or Hispanic	Female	Married Female	Teen	Aged 60-70	LTHS	HSL

*Notes:* The table reports the effects of the minimum wage on labor market outcomes by demographic group based on the event study analysis (see equation 1). We exploit 172 state-level minimum wage changes between 1979 and 2019. We assess the impact of the minimum wage on the high-recall group including individuals whose predicted probability is above 11.1% - a threshold which leads to a 75% recall rate of minimum wage workers. The table reports five year averaged post-treatment estimates for each key labor market outcome: percent change in wages and the change in employment to population, unemployment to population, and labor force participation rate. We also report the employment elasticity with respect to the minimum wage and the employment elasticity with respect to the wage, which is the ratio of the percent change in employment and wage. To calculate the percent change in employment we divide the change in employment to population by the mean employment to population rate preceding the minimum wage hikes (reported at the bottom of the table). The line on the number of observations shows the number of quarter-state cells used for estimation, while the number of individuals refers to the underlying CPS sample used to calculate labor market outcomes in these cells. In all the regressions we use the best performing prediction model — the boosted tree prediction model. The demographic subgroups are black, non-white and non-black (NWNB), woman, married woman, teen, aged 60 and older and less than 70, less than high school (LTHS), and high school or less (HSL). All estimates are weighted by the corresponding subgroup’s population in the state. Robust standard errors in parentheses are clustered by state; significance levels are \* 0.10, \*\* 0.05, \*\*\* 0.01.

Table 6: Impact of the Minimum Wage on Labor Market Outcomes by Demographic Groups II

	(1)	(2)	(3)	(4)	(5)	(6)
Δ employment (% pt)	0.000 (0.002)	-0.001 (0.002)	0.003 (0.003)	0.001 (0.001)	0.002 (0.005)	0.005 (0.004)
Δ unemployment (% pt)	0.000 (0.000)	0.000 (0.001)	-0.001 (0.001)	0.000 (0.000)	0.003 (0.003)	-0.000 (0.003)
Δ participation (% pt)	0.001 (0.002)	-0.000 (0.002)	0.002 (0.003)	0.001 (0.001)	0.005 (0.004)	0.004 (0.004)
Number of events	157	157	157	157	157	157
Number of observations	5,609	5,609	5,609	5,609	5,603	5,593
Number of individuals in the sample	1,736,611	872,381	358,834	505,396	43,367	74,065
Mean employment	0.422	0.516	0.541	0.135	0.592	0.666
Mean unemployment	0.045	0.067	0.038	0.006	0.070	0.044
Mean participation	0.467	0.583	0.579	0.141	0.662	0.709
Group	High Recall	High Recall	High Recall	High Recall	High Recall	High Recall
Demog. Group	All	High lfp switch	Medium lfp switch	Low lfp switch	Single mothers	Married mothers

Notes: The table reports the effects of the minimum wage on labor market outcomes by demographic group based on the event study analysis (see equation 1). We exploit 159 state-level minimum wage changes between 1990 and 2019. We assess the impact of the minimum wage on the high-recall group including individuals whose predicted probability is above 11.1% - a threshold which leads to a 75% recall rate of minimum wage workers. The table reports five year averaged post-treatment estimates for each key labor market outcome: the change in employment to population, unemployment to population, and labor force participation rate. The line on the number of observations shows the number of quarter-state cells used for estimation, while the number of individuals refers to the underlying CPS sample used to calculate labor market outcomes in these cells. In all the regressions we use the best performing prediction model — the boosted tree prediction model. Column (1) shows the estimates on overall employment using the 1990 to 2019 periods. Column (2)-(4) shows the estimates for individuals with the different predicted probability on moving in or out from the labor force. Column (5) shows the estimates on single mothers, while column (6) on married mothers (column 7). All estimates are weighted by the corresponding subgroup's population in the state. Robust standard errors in parentheses are clustered by state; significance levels are \* 0.10, \*\* 0.05, \*\*\* 0.01.

Table 7: Impact of the Minimum Wage on Alternative Labor Market Outcomes

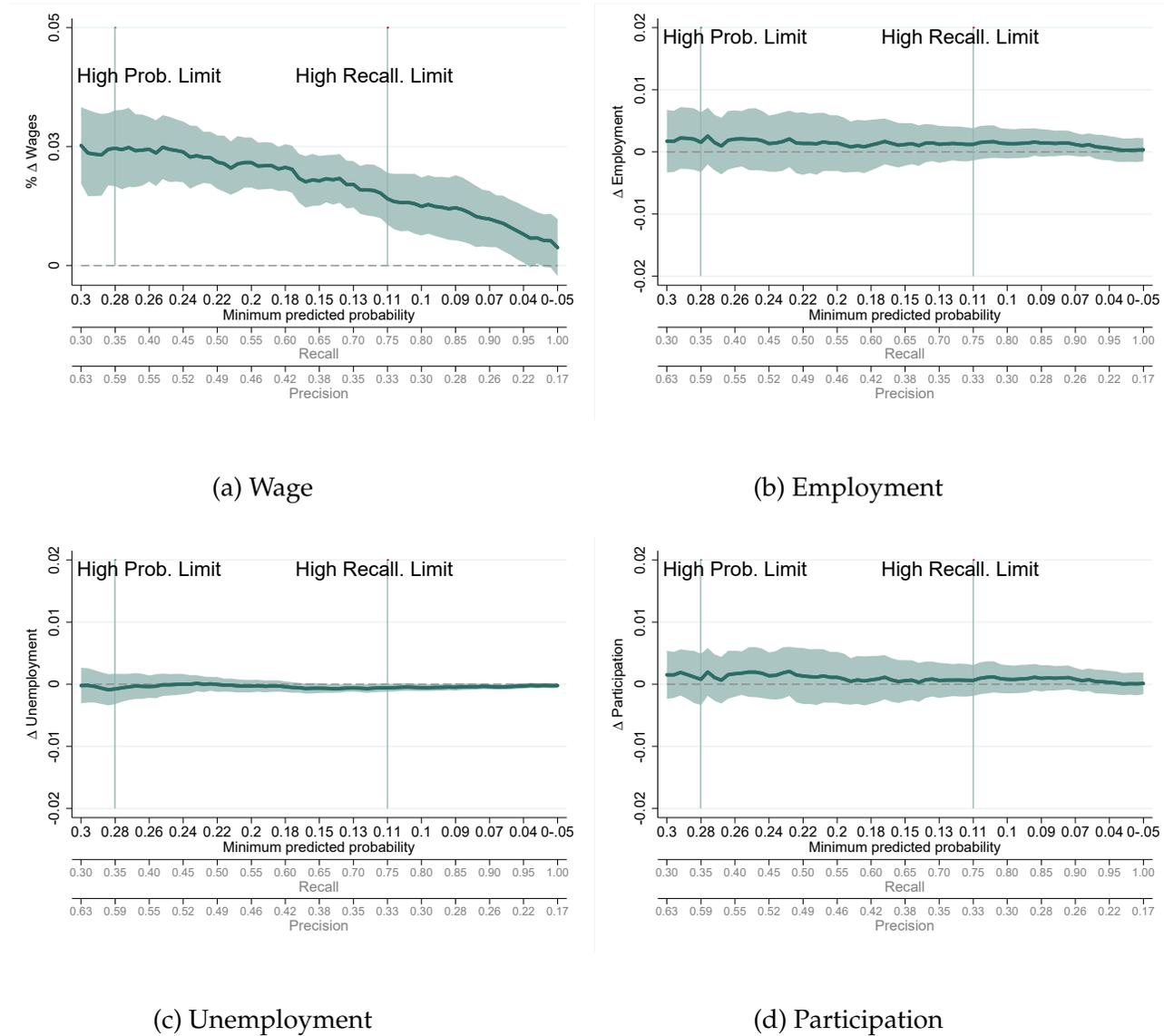
	(1)	(2)	(3)
$\Delta$ self-employment as share of employment (% pt)	0.001 (0.001)	0.000 (0.001)	-0.001 (0.001)
$\Delta$ part-time as share of employment (% pt)	-0.006*** (0.002)	-0.001 (0.001)	-0.000 (0.000)
$\Delta$ over-time as share of employment (% pt)	-0.005 (0.003)	-0.001 (0.002)	-0.000 (0.001)
Number of events	172	172	172
Number of observations	7,854	7,854	7,854
Number of individuals in the sample	571,091	1,909,060	2,239,793
Mean self-employment	0.031	0.068	0.136
Mean part-time	0.409	0.210	0.072
Mean over-time	0.538	0.336	0.150
Group	High Prob.	High Recall	Low Prob.
Prediction Model	Boosted Tree	Boosted Tree	Boosted Tree

*Notes:* The table reports the effects of the minimum wage on alternative labor market outcomes based on the event study analysis (see equation 1) exploiting 172 state-level minimum wage changes between 1979 and 2019. The table reports five year averaged post-treatment estimates for the following labor market outcomes: self-employment, part-time (working less than 30 hours per week) and over-time (working more than 40 hours per week). Each of these variables are expressed as a share of total employment. Column (1) shows estimates for the high-probability group, which captures 10% of the population with the highest predicted probability. Column (2) shows estimates for the high-recall group, which consists of individuals whose predicted probability is above 11.1% - a threshold which leads to a 75% recall rate of minimum wage workers. Column (3) shows the estimates for workers whose predicted probability is below 11.1%. All columns use the best performing prediction model — the boosted tree prediction model. All the regressions are weighted by state-quarter population. Robust standard errors in parentheses are clustered by state; significance levels are \* 0.10, \*\* 0.05, \*\*\* 0.01.

# APPENDIX

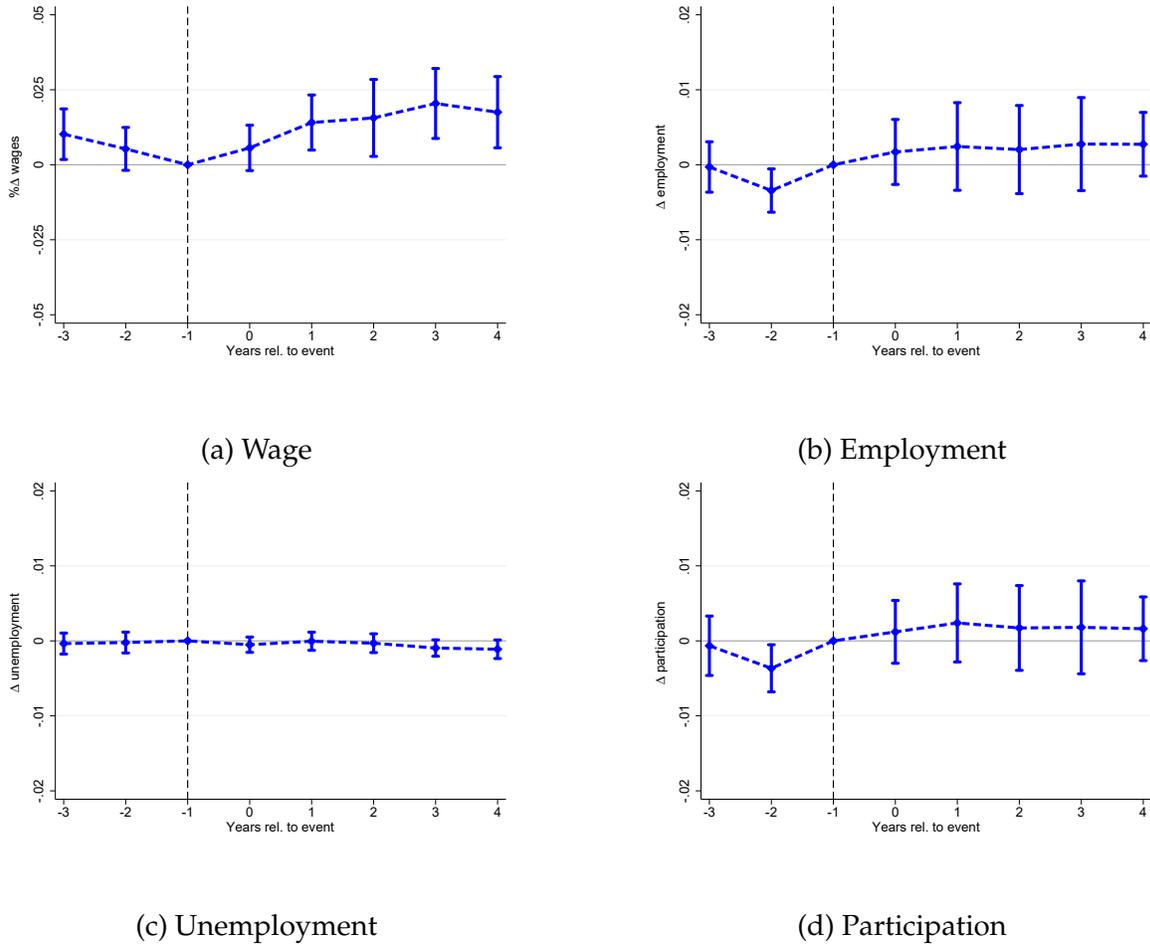
## APPENDIX A: Additional Figures and Tables

Figure A.1: Impact of the Minimum Wage for Alternative Predicted Probability Threshold Values, Card and Krueger Linear Prediction Model



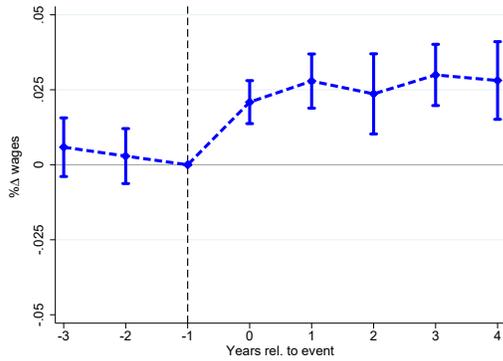
Notes: The figure shows the main results from our event study analysis (see equation 1) using alternative predicted probability threshold values. We exploit 172 state-level minimum wage changes between 1979-2019. Panel (a) shows the impact of the minimum wage on wages, Panel (b) on employment to population, Panel (c) on unemployment to population, and Panel (d) on participation rate. In each panel the green solid lines show the five year averaged post-treatment estimates for individuals whose predicted probability is above the “minimum predicted probability threshold.” On the x-axis we also report the corresponding recall rate (the fraction of minimum wage workers retrieved by the prediction model if the particular minimum predicted probability threshold is applied) and the precision rate (the fraction of minimum wage workers in the sample if the particular minimum predicted probability threshold is applied). We also plot the thresholds corresponding to the high-probability group capturing 10% of the population with the highest predicted probability and to the high-recall group capturing 75% of all minimum wage workers. To calculate the predicted probabilities we use the Card and Krueger linear prediction model. The shaded areas show the 95% confidence interval based on standard errors that are clustered at the state level.

Figure A.2: Impact of the Minimum Wage Over Time (High-Recall Rate Group), Balanced Panel of Events

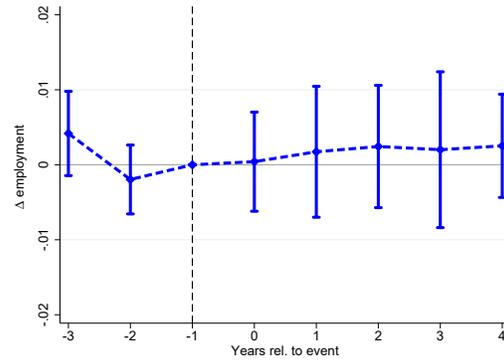


Notes: The figure shows the main results from our event study analysis (see equation 1) exploiting 99 state-level minimum wage changes between 1982-2014. While the estimation sample is between 1979-2019, here we exclude events for which we do not observe responses for the entire event window (3 years before and 5 years after). The figure shows the effect of the minimum wage increase on wages (Panel (a)), on employment to population (Panel (b)), on unemployment to population (Panel (c)) and on labor force participation rate (Panel (d)) for the high-recall group. The high-recall group consists of all workers whose predicted probability is above 11.1% – a threshold which corresponds to a 75% of recall rate. To calculate the predicted probabilities we use the best performing prediction model – the boosted tree prediction model. We also show the 95% confidence interval based on standard errors that are clustered at the state level.

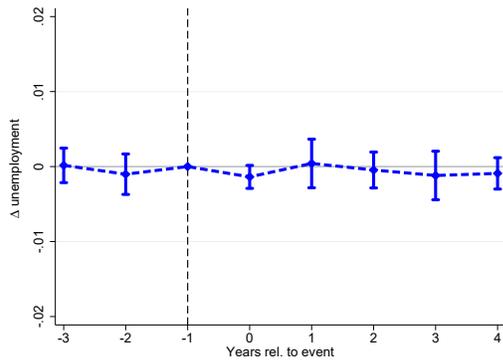
Figure A.3: Impact of the Minimum Wage Over Time (High-Probability Group), Balanced Panel of Events



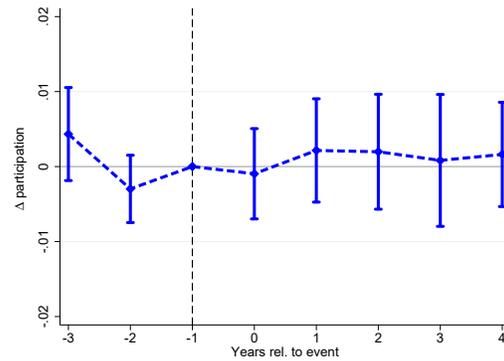
(a) Wage



(b) Employment



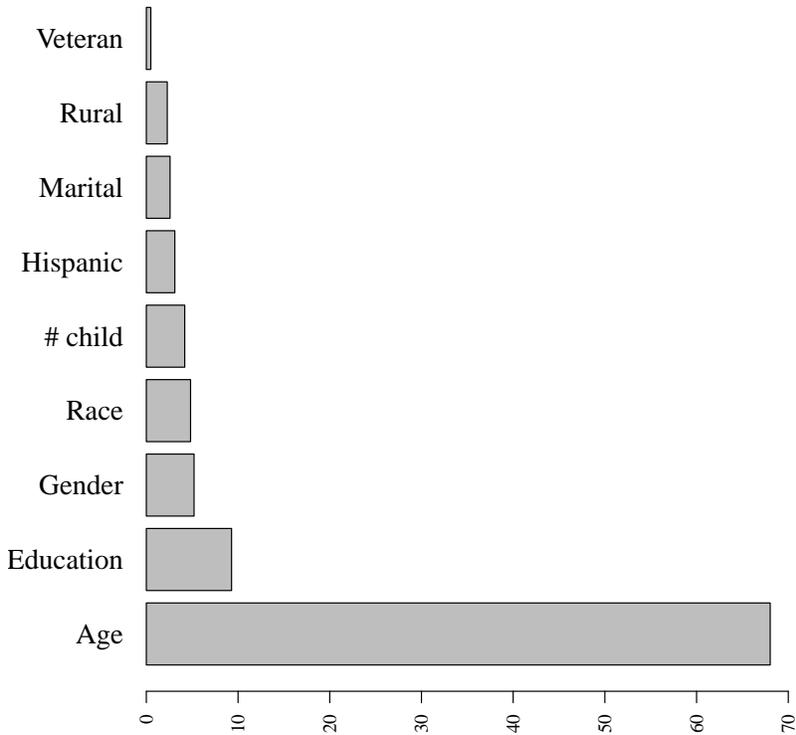
(c) Unemployment



(d) Participation

Notes: The figure shows the main results from our event study analysis (see equation 1) exploiting 99 state-level minimum wage changes between 1982-2014. While the estimation sample is between 1979-2019, here we exclude events for which we do not observe responses for the entire event window (3 years before and 5 years after). The figure shows the effect of the minimum wage increase on wages (Panel (a)), on employment to population (Panel (b)), on unemployment to population (Panel (c)) and on labor force participation rate (Panel (d)) for the high-probability group. The high-probability group consist of 10% of the population with the highest likelihood of being affected by the policy. To calculate the predicted probabilities we use the boosted tree prediction model. We also show the 95% confidence interval based on standard errors that are clustered at the state level.

Figure A.4: Relative Influences of the Predictors in the Boosted Tree Prediction Model for Switching the Labor Force Status



Notes: We plot the relative influences of the variables in the boosted tree prediction model for switching a labor force status. We calculate the relative influence as in Friedman (2001) (see footnote 13 for the details). The bars, which indicate the decline in the loss function associated with the corresponding variable, are normalized so that they sum up to 100.

Table A.1: Impact of the Minimum Wage on Labor Market Outcomes - Robustness to Alternative Specifications (High-Probability Group)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta$ wage (%)	0.023*** (0.002)	0.023*** (0.003)	0.024*** (0.003)	0.019*** (0.004)	0.018*** (0.004)	0.024*** (0.003)	0.020*** (0.004)	0.026*** (0.005)
$\Delta$ employment (% pt)	0.002 (0.003)	0.002 (0.002)	0.003 (0.002)	0.006*** (0.002)	0.006*** (0.002)	0.002 (0.003)	0.001 (0.002)	0.002 (0.004)
$\Delta$ unemployment (% pt)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.000 (0.001)	-0.001 (0.001)
$\Delta$ participation (% pt)	0.001 (0.002)	0.001 (0.002)	0.002 (0.002)	0.005** (0.002)	0.005** (0.002)	0.001 (0.002)	0.001 (0.002)	0.001 (0.003)
Employment Elas. w.r.t Min. Wage	0.067 (0.079)	0.064 (0.075)	0.081 (0.077)	0.184*** (0.061)	0.185*** (0.068)	0.067 (0.079)	0.040 (0.051)	0.047 (0.096)
Employment Elas. w.r.t Wage	0.280 (0.338)	0.267 (0.324)	0.325 (0.319)	0.942** (0.418)	1.008** (0.477)	0.259 (0.311)	0.188 (0.253)	0.198 (0.398)
Number of events	172	172	172	172	172	406	172	99
Number of observations	7,854	7,854	7,854	7,854	7,854	7,854	7,854	7,854
Number of individuals in the sample	571,091	571,091	571,091	571,091	571,091	1,461,676	571,091	381,354
Mean employment	0.338	0.338	0.338	0.338	0.338	0.348	0.352	0.356
Mean unemployment	0.057	0.057	0.057	0.057	0.057	0.062	0.057	0.063
Mean participation	0.395	0.395	0.395	0.395	0.395	0.410	0.409	0.419
Controls:								
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y
State Linear Trends		Y	Y		Y			
State Quadratic Trends			Y		Y			
Division-Quarter FE				Y	Y			
State Federal Events						Y		
Unweighted							Y	
No Events After 2014q1								Y

*Notes.* The table reports the effects of the minimum wage on labor market outcomes based on the event study analysis (see equation 1) exploiting 172 minimum wage changes between 1979 and 2019. We assess the impact of the minimum wage on the high-probability group. The high-probability group consists of 10% of the overall population with the highest likelihood of being affected by the policy. The table reports five year averaged post-treatment estimates for each key labor market outcome: percent change in wages and the change in employment to population, unemployment to population, and labor force participation rate. We also report the employment elasticity with respect to the minimum wage and the employment elasticity with respect to the wage, which is the ratio of the percent change in employment and wage. To calculate the percent change in employment we divide the change in employment to population by the mean employment to population rate preceding the minimum wage hikes (reported at the bottom of the table). The line on the number of observations shows the number of quarter-state cells used for estimation, while the number of individuals refers to the underlying CPS sample used to calculate labor market outcomes in these cells. In all the regressions we use the best performing prediction model — the boosted tree prediction model. The first column employs the preferred event-based estimate. Columns (2)-(5) augment the baseline model with linear or quadratic state trends and/or division-by-period fixed effects as indicated in the table. The sixth column reports estimates using 369 state or federal minimum wage increases. All regressions are weighted by state-quarter population except Column (7), where we report unweighted estimates. Column (8) only considers minimum wage events that happened on or before 2014q1 to ensure a full five year post-treatment period. Robust standard errors in parentheses are clustered by state; significance levels are \* 0.10, \*\* 0.05, \*\*\* 0.01.

Table A.2: Impact of the Minimum Wage on Labor Market Outcomes: Quintile Analysis

	(1)	(2)	(3)	(4)	(5)
$\Delta$ wage (%)	0.021*** (0.003)	0.007* (0.004)	0.001 (0.004)	-0.004 (0.004)	-0.001 (0.004)
$\Delta$ employment (% pt)	0.001 (0.002)	0.001 (0.001)	0.001 (0.001)	-0.000 (0.001)	0.000 (0.001)
$\Delta$ unemployment (% pt)	0.000 (0.001)	-0.001** (0.000)	0.000 (0.001)	-0.000 (0.001)	-0.000 (0.000)
$\Delta$ participation (% pt)	0.001 (0.002)	0.000 (0.001)	0.001 (0.001)	-0.000 (0.001)	-0.000 (0.001)
Employment Elas. w.r.t Min. Wage	0.034 (0.061)	0.023 (0.028)	0.019 (0.024)	-0.001 (0.020)	0.002 (0.011)
Employment Elas. w.r.t Wage	0.150 (0.267)	0.303 (0.445)	1.791 (6.165)	0.035 (0.601)	-0.141 (1.214)
Number of events	172	172	172	172	172
Number of observations	7,854	7,854	7,854	7,854	7,854
Number of individuals in the sample	1,094,498	920,053	951,348	683,129	1,095,178
Mean employment	0.375	0.499	0.633	0.751	0.853
Mean unemployment	0.051	0.038	0.036	0.030	0.023
Mean participation	0.426	0.537	0.669	0.781	0.876
Group	Fifth Quintile	Fourth Quintile	Third Quintile	Second Quintile	First Quintile
Prediction Model	Boosted Tree	Boosted Tree	Boosted Tree	Boosted Tree	Boosted Tree

*Notes:* The table reports the effects of the minimum wage on labor market outcomes based on the event study analysis (see equation 1) exploiting 172 state-level minimum wage changes between 1979 and 2019. The table reports five year averaged post-treatment estimates for each key labor market outcome: percent change in wages and the change in employment to population, unemployment to population, and labor force participation rate. We also report the employment elasticity with respect to the minimum wage and the employment elasticity with respect to the wage, which is the ratio of the percent change in employment and wage. To calculate the percent change in employment we divide the change in employment to population by the mean employment to population rate preceding the minimum wage hikes (reported at the bottom of the table). The line on the number of observations shows the number of quarter-state cells used for estimation, while the number of individuals refers to the underlying CPS sample used to calculate labor market outcomes in these cells. Columns (1) to (5) show estimates for the top through to bottom quintile respectively. Robust standard errors in parentheses are clustered by state; significance levels are \* 0.10, \*\* 0.05, \*\*\* 0.01.

## APPENDIX B: Data Sources and Variable Construction

The data sets in the main text are as follows:

We use the 1979-2019 CPS-Outgoing Rotation Group (CPS-ORG) in building the prediction model, and in estimating the wage effects of the minimum wage. In constructing the hourly wage variable, we exclude self-employed workers as well as observations with imputed wage information. Following [Feenberg and Roth \(2007\)](#)'s recommendation, if the individual is not paid hourly, we calculate the hourly wage by dividing earnings per week by usual weekly hours worked in the job.

We acquire the state-level quarterly statutory minimum wage data from [Vaghul and Zipperer \(2016\)](#).

In building the prediction models, we use the following variables:

**Minimum wage worker indicator:** The outcome variable that takes on the value of 1 if the individual's hourly wage is less than %125 of the statutory minimum wage. We also do not include self-employed workers or those with imputed wages.

**Age:** We use reported age throughout the text.

**Education:** We construct a categorical variable with four categories, less than high school (EDUC=1), high school graduate with no college education (EDUC=2), some college (EDUC=3), and college graduate (EDUC=4), using the variables that report highest degree completed.

**Gender:** We construct a binary variable that takes on the value of 1 if the individual is male, and 0 otherwise.

**Rural residency:** We construct a categorical variable with two categories, resident in a rural area (RURALSTATUS=1) or resident in a small metropolitan area (RURALSTATUS=2).

**Marital:** We construct a binary variable that takes on the value of 1 if the individual is married and the spouse is present, and 0 otherwise.

**Race:** We construct a categorical variable that takes on the value of 1 if the individual is coded as white, and 2 if non-white.

**Hispanic:** We construct a binary variable that takes on the value of 1 if the individual reports of Hispanic ethnicity, and 0 otherwise.

**Veteran:** We construct a binary variable that takes on the value of 1 if the individual is a veteran, and 0 otherwise.

## APPENDIX C: Details for the Prediction Algorithms and the Robustness of the Preferred Prediction Model

In this section, we provide details of the algorithms we employed in the main text and we assess the robustness of the predictions to alternative definitions in determining “minimum wage workers”. In addition, we provide suggestive evidence that the worker misreporting does not severely affect our predictions by comparing models constructed using worker, and employer-reported wages using the January 1977 CPS Supplement.

### Tree-based Machine Learning Tools

#### Random Forests

There is a multiplicity of ways to go beyond a single decision tree. The random forest is one of them (Breiman, 2001). It is a tree-based ensemble learning technique. It provides a way to overcome the bias-variance trade-off of the decision trees. In our case, it constructs a multitude of fully grown decision trees formed using different training bootstrap samples that predict the class of each observation. Using these predictions, it determines the class of the observation according to the majority vote. The final prediction of each observation, therefore, is the average of all the tree predictions.

Since each tree is fully grown, they are unbiased, yet the prediction variance is high. By averaging the predictions, the unbiasedness is preserved and the variance is diminished. To elaborate this point further, assume that the variance of the prediction of the decision trees is  $\sigma^2$ , and  $\hat{f}^b(x)$  is the prediction of the decision tree that is formed using training sample  $b$  for given predictors,  $x$ . The random forest predicts the class of the observation by averaging the predictions  $\hat{f}_{rf} = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x)$ , where  $B$  is the total number of trees. Then, the unbiasedness is retained, and if the predictions were independent from each other, the variance of the random forest predictions would be;

$$\text{var}(\hat{f}_{rf}) = \text{var}\left(\frac{1}{B} \sum_{b=1}^B \hat{f}^b(x)\right) = \left(\frac{1}{B^2}\right) * B * \sigma^2 = \frac{\sigma^2}{B}. \quad (2)$$

However, the trees and the predictions are never uncorrelated. Especially, if one predictor has a very high predictive power, then the top node of all trees use it to split. Therefore, the variance

in equation 2 is, in fact, never achieved. To decrease the correlation of trees, instead of using all the predictors, we employ a randomly selected portion of the predictors at each split.<sup>19</sup> Using a fraction of the predictors might slow the learning process; though, with large number of trees, it outperforms the random forest that uses all of the predictors at each step. In addition, thanks to the averaging, increasing the number of trees does not lead to overfitting; yet the prediction performance does not improve after a certain number of trees<sup>20</sup>.

## Gradient Tree Boosting

The boosting approaches the problem from a different angle. Instead of producing many fully grown trees, it starts by producing a relatively small tree. The initial tree is a weak learner, misclassifies many observations. The proceeding trees, also relatively small and weak learners, alter the data to predict the misclassified observations more accurately.

The gradient tree boosting developed by Friedman (2001) is one of the boosting techniques<sup>21</sup>. In our classification problem with only two classes, we choose the binomial log-likelihood loss function (or one half of the deviance). Hence, the loss function is;

$$L(y, p) = -(y \log(p) + (1 - y) \log(1 - p)), \quad (3)$$

where  $y$  is the class of the observation ( $y \in \{0, 1\}$ ).  $p = P(Y = 1|X)$  where  $X$  indicates predictors. Then using the logistic transformation, we can write the loss function in terms of the model as:

$$L(y, F(X)) = -yF + \log(1 + \exp(F)). \quad (4)$$

Instead of fitting a single tree using 4, the boosting fits many weak trees sequentially. The main trick of the gradient tree boosting is that at step  $m$ , we replace the outcome with the negative gradient, given  $m-1$  boosting steps. To put it differently, the pseudo-response of the observation  $i$ ,

<sup>19</sup>Using all predictors at every split is called “bagging”.

<sup>20</sup>However, although increasing the number of trees does not lead to overfitting, individual trees themselves might overfit severely. As Segal (2004) shows, it is possible to improve the prediction by shrinking the tree size.

<sup>21</sup>For the purposes of this paper, we only describe the gradient tree boosting. Alternatively, one can also employ the AdaBoost algorithm. The gist of the AdaBoost algorithm is that at each step, observations are re-weighted so that misclassified (correctly classified) ones weigh slightly more (less) in the subsequent step. In our study, the AdaBoost performed slightly worse than the preferred model, so we omitted it. A very intuitive description of both the AdaBoost and gradient tree boosting algorithms can be found in Friedman et al. (2009).

$\tilde{y}_i$  is defined as:

$$\tilde{y}_i = -\left[\frac{\partial L}{\partial F}\right]_{F=F_{m-1}} = y_i - p_i. \quad (5)$$

and the tree at step  $m$  is fit to  $\tilde{\mathbf{y}}^{22}$ . Then based on the fit, the new tree is added to the model according to the following formula;

$$F_m(X) = F_{m-1}(X) + \sum_{j=1}^J \gamma_{jm} \mathbf{1}(X \in R_{jm}), \quad (6)$$

where  $j$  is the terminal node of the tree,  $F_{m-1}(X)$  is the model built at step  $m - 1$ , and  $\gamma_{jm}$  is the optimal update coefficient that reduces the loss function the most at step  $m$  for the sub-space  $R_{jm}$ . Concretely,  $\gamma_{jm}$  takes higher values for subspaces that the  $m$ th tree fits relatively well. Note that due to the use of pseudo-responses, the minimization of loss function leads the  $m$ th tree to focus heavily on cases where  $F_{m-1}$  performs poorly. Only the combination of weak learners produce a strong learner, and most of the weak learners are meaningless by themselves.

Since we fit the tree to the negative gradient, the training error rate always decreases as the number of trees increases. Therefore, unlike the random forest, including many trees in the model can lead to overfitting in the gradient tree boosting. For regularization and better predictions, shrinkage techniques are employed. Instead of adding each new tree to the current model as in equation 6, every  $\gamma_{jm}$  is multiplied by a small positive number. The shrinkage parameter renders each learner even weaker and decelerates the learning process. Nevertheless, combined with sufficient number of trees, this can increase the model's predictive power significantly. Furthermore, using only a fraction of training set, known as sub-sampling, one can introduce another stochastic component and de-correlate trees; hence potentially decreasing the variance (Friedman, 2002).

Then, three main parameters that are user-specified in gradient boosting are: the size of each tree, the number of trees, and the shrinkage factor. We employ a 10-fold cross-validation procedure to determine each one of them simultaneously.

Before concluding this section, we note that although both random forests and gradient tree boosting build data-driven models that, in general, outperform a single tree, these models lack the interpretability. In the case of the random forest, due to the use of a fraction of variables at each

---

<sup>22</sup>For the least squares regressions, the negative gradient vector is the residuals at step  $m$ . Hence, the gradient boosting simply fits the regression tree to the current residuals at each step (James et al., 2013).

split, most of the individual trees cannot achieve the performance of a single tree, and the splits are not interpretable. On the other hand, the first tree of the gradient boosting can actually be the same as the decision tree; yet starting from the second tree, the interpretation of each tree is obscured due to the updating of the outcome variable or the loss function.

### **Elastic net**

We also explore the performance of the logistic regression model. We use the elastic net regularization developed by [Zou and Hastie \(2005\)](#). The loss function of the logistic regression is:

$$L = -\frac{1}{N} \sum_{i=1}^N y_i(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) - \ln(1 + \exp(\beta_0 + \mathbf{x}_i \boldsymbol{\beta}))$$

The elastic net regularization adds the penalty term,  $\lambda[(1 - \alpha)\|\boldsymbol{\beta}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_1]$  to the loss function for regularization and decreasing the model complexity.  $\lambda \geq 0$  and  $\alpha \in [0, 1]$  are tuning parameters picked by the cross-validation procedure to prevent overfitting. We purposefully build a very complex model and rely on the regularization to avoid overfitting. The model includes all the predictors, their two-way interactions, all the interactions with the quadratic age variable, and the cubic and quartic terms of the age variable.

### **Card and Krueger's linear probability model**

We also apply the linear probability model analogous to the one employed by [Card and Krueger \(1995\)](#). While the former model has no polynomial or interaction terms, the right hand side variables of the latter model are a set of three-way interaction variables between teen, non-white, and gender indicators; three-way interaction variables between young adult (age 20-25), non-white, and gender indicators; three-way interactions of age, categorical education, and gender variables; quadratic and cubic terms of the age variable; indicator variables for Hispanic, and non-white individuals.

### **Robustness to alternative threshold values**

In the main prediction model, we classified wage workers earning less than 125% of the minimum wage as relevant cases (minimum wage workers), and other wage workers as non-relevant cases. We use the threshold value primarily for description purposes, however it might be the case that

the arbitrarily selected threshold value has a qualitative effect on the predictions since it determines the value of the outcome variable. Taking the concern into account, we build prediction models for alternative threshold values. Since we use the predicted probabilities for sorting, we compare ranks of the predicted probabilities estimated by latter models with the main model using the 1979-2019 CPS-ORG. In addition, we also build a model to predict real wages, sort according to (the negative of) predicted wages, and compare the resulting order with the one obtained by the main prediction model.

In figure C.1 and table C.1, we show that selecting alternative threshold values produce virtually the same ordering. The rank correlation coefficients are always greater than 0.95. The coefficient is always greater than 0.99 for the threshold values (1.03, 1.1, 1.25, 1.5) used in Belman et al. (2015) when they try alternative threshold values in describing demographics and occupations of minimum wage workers. This implies that we obtain highly similar samples when we use alternative threshold values in the prediction model or predicted real wages for sorting. In other words, the high impact and the baseline groups would be essentially the same if we used another threshold wage level in defining the minimum wage workers.

### **Impact of misreporting error on the formation of groups**

One issue that might affect the formation of predicted probability groups is the misreporting error. If the misreporting error of the hourly wage information is larger for certain groups, then the members of the groups might be incorrectly predicted to be more (or less) likely to be a minimum wage worker. An example that illustrates the issue is as follows: consider a case where the minimum wage is \$10, and two demographic groups have the same true wage distributions. The distributions are both normal with mean \$14.5 and standard deviation \$2. This implies that, in truth, 31.7% of both groups are minimum wage workers according to the convention (minimum wage workers are defined as those earning less than 125% of the minimum wage). Say, the first group reports their hourly wages with some error. The error would increase the observed standard deviation without affecting the mean. Then, the ML tools would predict the first group to be more likely to have minimum wage workers than the second group.<sup>23</sup>

---

<sup>23</sup>The large rank correlation coefficient of the main predicted probabilities and the negative of predicted real wages in table C.1 imply that the severity of this kind of error is limited.

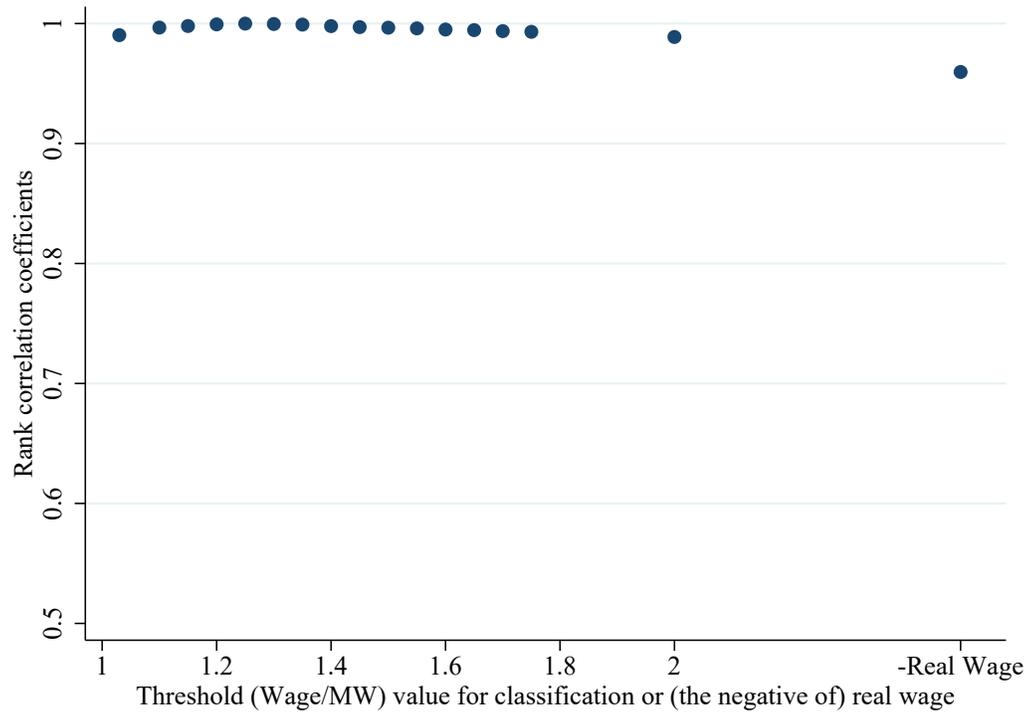
As suggestive evidence on the impact of misreporting error on the predictions, we employ the CPS 1977 January Supplement. In the data set, hourly wages of workers are asked to employers as well as to workers. We assess the severity of the misreporting error by comparing the prediction model constructed using worker-reported wages with that using employer-reported wages.

Specifically, we use 80% of the sample to train two models: the first model employs the worker-reported wage information, whereas the second model employs the employer-reported wage information. The predictors are the same as the model in the paper, except that the CPS 1977 does not have citizenship information, and the small metropolitan areas are defined as those with population less than 1,000,000 (rather than 500,000). After having constructed the models, we test the performances of each model using the remaining 20% of the original sample. Each model predicts the likelihood of being a minimum wage worker for each of the observations in the test sample. The similarity of the predictions suggests that the misreporting error negligibly affects the sorting.<sup>24</sup> The slope, the  $R^2$ , and the rank correlation coefficient are all very close to 1: they are 0.990, 0.989, 0.992, respectively. This suggests that using worker-reported wages results in similar predictions to using employer-reported wages.

---

<sup>24</sup>One underlying assumption here is that the misreporting errors of employers and workers are independent from each other. More precisely,  $w^{worker} = w + e^{worker}$ ,  $w^{employer} = w + e^{employer}$  and  $e^{worker}$  is independent from  $e^{employer}$ .

Figure C.1: Rank Correlation Coefficients



Notes: The 1996-2017 CPS-ORG is employed. The graph shows the rank correlation coefficients of the predicted probabilities obtained from models constructed using alternative threshold  $\frac{wage}{MW}$  values in classifying observations or from the model predicting the negative of real wage and the main predicted probabilities.

Figure C.2: Predictions of Models Based on Employer-Reported, and Worker-Reported Wages



*Notes:* The test sample of the January 1977 CPS Supplement is employed. The graph plots the predicted probabilities of the model using employer-reported wages (y-axis) against those of the model using worker-reported wages (x-axis). Each marker indicates an observation, while the straight line shows the best-fitted. The dashed line shows the 45-degree line. The  $R^2$  of the fit, the slope of the line, and the rank correlation coefficients are reported in the top-right corner.

Table C.1: Rank correlation coefficients

Outcome variable:	Indicator for hourly wage below the threshold						-Real wage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rank correlation coefficient	0.990	0.997	1.000	0.997	0.993	0.989	0.960
Threshold value	1.03	1.1	1.25	1.5	1.75	2	-

*Notes:* The 1996-2017 CPS-ORG is used. Estimated rank correlation coefficients of the main predicted probabilities and the predicted probabilities of the models that use alternative thresholds or the hourly real wage for the outcome variable. Columns (1)-(6) employ a binary outcome that takes on the value of 1 if the ratio of the real hourly wage to the minimum wage is below the specified threshold. Column (7) employs the negative of hourly real wage as the outcome, and predicts the worker's wage. Since the observations more likely to earn low wages are more likely to be minimum wage workers, we sort according to the negative of the real wage.

## APPENDIX D: Importance of Participation in Flinn (2006)

The link between participation and welfare can be seen in a three state model of the labor market, such as Flinn (2006), where workers are either not participating in the labor market, participating and unemployed or participating and employed. We simplify Flinn (2006) by assuming firm homogeneity, and maintain the assumption that workers are identical except in their flow value of non-participation,  $\rho V_i^n$ , which is distributed according to the cumulative distribution function  $F$ . Denote the lifetime value of being employed as  $V^e$  and unemployed as  $V^u$ , then when the minimum wage,  $m_w$ , binds we will have that

$$\begin{aligned}\rho V_i^n &\sim F \\ \rho V^u &= b + \lambda(m_w)(V^e - V^u) \\ \rho V^e &= m_w + \delta(V^u - V^e)\end{aligned}$$

where  $b$  is the replacement rate for unemployed workers,  $\lambda(m_w)$  is the job finding rate which is allowed to vary with the minimum wage, and  $\delta$  is the exogenous job destruction rate. Worker  $i$  participates if and only if  $V^u \geq V_i^n$ , so the participation rate is given by  $F(\rho V^u)$ . If the value of non-participation is invariant to the minimum wage, then an increase in participation following a minimum wage increase is a sufficient condition for an increase in the lifetime value of being unemployed,  $V^u$ , which factors in the joint impact of minimum wage changes on future wages conditional on employment, and on the probability of employment. This follows because the participation rate,  $F(\rho V^u)$ , increases if and only if  $V^u$  increases. Indeed if  $F$  is a twice differentiable continuous function and  $\sup(\rho V_i^n) > \rho V^u$ , i.e. participation is not 100%, then an increase (decrease) in participation is both a necessary and sufficient condition for an increase in the lifetime value of being unemployed. However, in a more general setting where some groups do have close to 100% participation rates then aggregate participation changes are more informative about welfare changes for marginal rather than infra-marginal groups of workers.