I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16894

Using Post-Regularization Distribution Regression to Measure the Effects of a Minimum Wage on Hourly Wages, Hours Worked and Monthly Earnings

Martin Biewen
Pascal Erhardt

MARCH 2024

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16894

Using Post-Regularization Distribution Regression to Measure the Effects of a Minimum Wage on Hourly Wages, Hours Worked and Monthly Earnings

Martin Biewen
University of Tübingen and IZA

Pascal Erhardt
University of Tübingen

MARCH 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

IZA DP No. 16894 MARCH 2024

ABSTRACT

Using Post-Regularization Distribution Regression to Measure the Effects of a Minimum Wage on Hourly Wages, Hours Worked and Monthly Earnings

We evaluate the distributional effects of a minimum wage introduction based on a data set with a moderate sample size but a large number of potential covariates. Therefore, the selection of relevant control variables at each distributional threshold is crucial to test hypotheses about the impact of the treatment. To this end, we use the post-double selection logistic distribution regression approach proposed by Belloni et al. (2018a), which allows for uniformly valid inference about the target coefficients of our low-dimensional treatment variables across the entire outcome distribution. Our empirical results show that the minimum wage crowded out hourly wages below the minimum threshold, benefitted monthly wages in the lower middle but not the lowest part of the distribution, and did not significantly affect the distribution of hours worked.

JEL Classification: J31, C3

Keywords: wage structure, automatic specification search, double machine learning

Corresponding author:
Martin Biewen
School of Business and Economics
University of Tübingen
Mohlstr. 36
72074 Tübingen
Germany

E-mail: martin.biewen@uni-tuebingen.de

1. INTRODUCTION

The introduction of the German statutory minimum wage on the 1st January of 2015 was a major policy experiment. Although there existed a number of industry-specific minimum wages before 2015, Germany was one of the few countries in the world without a general minimum wage. The introduction of the nationwide minimum wage at the level of 8.50 euros/hour in 2015 constituted a major intervention in the German labour market that affected around 4 million workers (over 11% of the workforce) who earned hourly wages below 8.50 euros before its introduction, see Mindestlohnkommission (2020).[1]

Based on different data sets, previous contributions have examined various aspects of the German minimum wage introduction. As to potential employment effects, the literature has reached the consensus that these were non-existent or very small, see, e.g. Caliendo et al. (2019), Dustmann et al. (2022), Bossler and Schank (2023). By contrast, the literature appears to have reached conflicting results about the distributional effects of the minimum wage, i.e., its effects on the distributions of hourly wages, monthly earnings and working hours (the latter including potential shifts between full-time, part-time and marginal part-time work). Using register data, Bossler and Schank (2023) find that the minimum wage significantly reduced inequality in monthly wages. On the basis of survey data, however, Burauel et al. (2019, 2020) and Caliendo et al. (2022) conclude that the minimum wage introduction also reduced working hours neutralizing its effect on monthly wages. Given that German register data do not include information on working hours, Biewen et al. (2022) analyse large-scale data from the statistical offices to conclude that working hours were *not* causally affected by the minimum wage so that increased hourly wages should fully translate into changes in monthly earnings.

Given the inconclusive evidence, the goal of this paper is to reconsider the effects of the minimum wage introduction on the distributions of hourly wages, monthly earnings and working hours using the survey data that were analysed in Burauel et al. (2019, 2020) and Caliendo et al. (2022). Based on modern machine learning methods that allow us to examine the effects of the minimum wage across all points of the distribution, we reach the conclusion that the minimum wage crowded out hourly wages below its threshold, benefitted monthly wages in the lower middle but not the lowest part of the distribution (consistent with Bossler and Schank, 2023), and did not lead to significant changes in the distribution of working hours. Our results help reconcile the conflicting results in the literature obtained from different data sources as described above.

Our analysis is based on the distribution regression approach introduced by Foresi and Peracchi (1995) and developed by Chernozhukov et al. (2013). The method of distribution regression consists of running a large number of binary regressions each modelling the likelihood that an outcome variable falls below a particular threshold in a fine grid of thresholds covering the whole distribution. Compared to alternative methods such as conditional or unconditional quantile regression, distribution regression directly targets nominal points in the outcome distribution. It is therefore ideally suited to study changes in distributions such as hourly wages, working hours or monthly wages whose quantiles typically change over time complicating the interpretation of quantile regression results if more than one time period is involved. Moreover, distribution regression easily deals with discrete mass points, see Chernozhukov et al. (2013), which is particulary relevant

[1]See Caliendo et al. (2019) for a more detailed overview of the institutional details of the minimum wage introduction.

when dealing with highly discrete distributions (working hours) or distributions with severe heaping (hourly wages, especially after the introduction of a minimum wage). This is in contrast to conditional or unconditional quantile regression that are based on the assumption of continuous distributions.

A practical challenge of running distribution regressions is that many separate binary regression models have to be specified and estimated. The different binary regression models should consider potentially varying sets of covariates as different covariates may matter at different points of the distribution. This challenge is particularly pronounced when using a sample with a moderate sample size but a large number of potential covariates as it will become inevitable to select relevant covariates at each distributional threshold to save degrees of freedom. Given that this typically has to be done over a large number of thresholds, a hand-picked approach may lead to a considerable amount of arbitrary specification search with unknown consequences for the potential bias of estimated coefficients and their estimated standard errors. Another aspect is the likely high correlation between the regression results for different distributional thresholds which complicates inferences over more than one point of the distribution. Apart from inferential aspects, the sheer practical task of specification searches for a large number of parallel regressions suggests the use of machine learning techniques such as the lasso to separately predict nuisance terms at the large number of distributional thresholds.

Both the practical and the inferential aspect have been addressed by recent advances in econometric machine learning. In a recent contribution, Belloni et al. (2018a) showed how to employ a sequence of ℓ_1-regularized logistic regressions such that inferences about the coefficients of target regressors are valid both pointwise, i.e., in each regression model separately, as well as uniformly across a large number of models estimated. Their proposed algorithm is closely related to the concept of 'partialling-out' in the econometrics literature, where one removes nuisance terms that are both related to the outcome and the treatment variables (in the present context referred to as 'double selection'). Picking covariates by the lasso method also entails functional form specification as the features offered to the lasso may include arbitrary transformations of variables (logs, polynomials, indicators for particular values, interaction terms). The possibility to obtain valid inference after large-scale automatic specification search is a remarkable achievement of the recent econometric machine learning literature. It represents a major improvement over the often arbitrary and undocumented specification searches carried out by individual researchers, which are typically influenced in unknown ways by the propagation of pre-tested control variables used in previous research.

A key assumption of the ℓ_1-regularized methods used by us is approximate sparsity, i.e., the sequence of coefficients of potential confounder terms sorted by absolute value decays quickly enough, but does not have to be exactly equal to zero. This is a natural assumption in substantive applications where one would like to control for relevant confounding information, but does not rule out the existence of factors whose influence may be negligible for the inferential purpose at hand. In the following, we use Belloni et al. (2018a)'s method but with some modifications to address the fact that our data and research design features observation clusters and sampling weights.[2]

[2]Our paper appears to be one of the first substantive applications of Belloni et al. (2018a)'s method, apart from Chiang (2020) who also considers modifications for observation clusters.

2. ECONOMETRIC METHODS

We aim to measure the effects of the minimum wage introduction across the whole distribution of an outcome variable $Y \in \{\text{hourly wage, hours worked, monthly earnings}\}$. For this purpose, we use the logistic distribution regression model

$$P(Y < u|D, X) = E[1\{Y < u\}|D, X] = \Lambda(D\theta_u + X\beta_u), \quad u \in \mathcal{U} \tag{2.1}$$

which measures the effects of target variables $D = (D_1, \ldots, D_{\tilde{p}})$ (representing treatment indicators of the minimum wage introduction, see below) across a set of grid points $u \in \mathcal{U}$ of the outcome distribution. In order to isolate the effect of the target variables on the likelihood of falling below a particular threshold u, it is necessary to control for confounders X, which may vary across different points of the outcome distribution. This motivates the use of a separate lasso procedure to select a set of relevant control variables at each point $u \in \mathcal{U}$.

For the rest of this article, let the vector of potential features be given by $X = (X_1, \ldots, X_p)$, target coefficients $\theta_u = (\theta_{u1}, \ldots, \theta_{u\tilde{p}})'$, nuisance parameters $\beta_u = (\beta_{u1}, \ldots, \beta_{up})'$ as well as indicators $Y^u = 1\{Y < u\}$. Our application features clusters $g = 1, \ldots, G$ of observations $W_{ig} = (Y_{ig}, D_{ig}, X_{ig}) = (Y_{ig}, D_{1ig}, \ldots, D_{\tilde{p}ig}, X_{1ig}, \ldots, X_{pig})$ assuming that observations are independent across clusters but may be correlated within clusters (see next section for more details). We also make use of deterministic sampling weights v_{ig} which are normalized to sum up to the total number of observations.

2.1. Estimation

The post-double lasso method for the logistic regression model (and other generalized linear models) was developed by Belloni et al. (2016a). Belloni et al. (2018a) extended the method to cover uniform inference for functional parameters, e.g., for coefficients of many parallel logit models as needed in our application. Extensions of the lasso method to clustered data were considered in Belloni et al. (2016b), Chiang (2020) and Ahrens et al. (2020). The idea of applying lasso to clustered data is to conceptualize blocks of observations belonging to the same cluster as 'super-observations' and to apply lasso operations accordingly.

Following Belloni et al. (2018a) and applying modifications for clustering similar to Chiang (2020), as well as for sampling weights, we use the following post-double selection procedure to estimate the process of target coefficients $(\theta_{uj}), u \in \mathcal{U}, j \in \mathcal{J} = \{1, \ldots, \tilde{p}\}$.

STEP 1. Post-lasso logit of Y^u on (D, X)

First run the ℓ_1-penalized logit regression

$$(\hat{\theta}_u, \hat{\beta}_u) \in \arg\min_{\theta,\beta} \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, M_u(W_{ig}, \theta, \beta) + \frac{\lambda_1}{G} \|\hat{\Psi}_u \cdot (\theta', \beta')'\|_1 \tag{2.2}$$

$$M_u(W_{ig}, \theta, \beta) = \log(1 + \exp(D_{ig}\theta + X_{ig}\beta)) - Y^u_{ig} \cdot (D_{ig}\theta + X_{ig}\beta) \tag{2.3}$$

where λ_1 is a penalty parameter and $\hat{\Psi}_u$ a diagonal penalty loading matrix whose entries are chosen according to the procedure explained in Appendix A.1.

Then obtain the post ℓ_1-penalized logit coefficients

$$(\tilde{\theta}_u, \tilde{\beta}_u) \in \arg\min_{\theta,\beta} \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig} \, M_u(W_{ig}, \theta, \beta) : \text{supp}(\theta, \beta) \subseteq \text{supp}(\hat{\theta}_u, \hat{\beta}_u) \quad (2.4)$$

where $\text{supp}(\hat{\theta}_u, \hat{\beta}_u)$ represents the set of indices associated with non-zero coefficients in the logistic lasso solution.

Step 2. Data-dependent orthogonalization

Define the weights to be used in step 3

$$\hat{f}^2_{uig} = \Lambda'(D_{ig}\tilde{\theta}_u + X_{ig}\tilde{\beta}_u). \quad (2.5)$$

Step 3. Weighted post-lasso OLS of $[\hat{f}_u D_j]$ on $[\hat{f}_u D_{\mathcal{J}\setminus j}]$ and $[\hat{f}_u X]$

For each target variable D_j, define $\tilde{X}^j = (D_{\mathcal{J}\setminus j}, X)$ and run the weighted lasso

$$\hat{\gamma}^j_u \in \arg\min_{\gamma} \frac{1}{2G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig} \, \hat{f}^2_{uig} \big(D_{jig} - \tilde{X}^j_{ig}\gamma\big)^2 + \frac{\lambda_2}{G} \|\hat{\Psi}_{uj}\gamma\|_1, \quad (2.6)$$

where λ_2 is a penalty parameter and $\hat{\Psi}_{uj}$ a diagonal penalty loading matrix whose entries are chosen according to the procedure explained in Appendix A.1.

The post-lasso WLS coefficients are defined as

$$\tilde{\gamma}^j_u \in \arg\min_{\gamma} \frac{1}{2G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig} \, \hat{f}^2_{uig} \big(D_{jig} - \tilde{X}^j_{ig}\gamma\big)^2 : \text{supp}(\gamma) \subseteq \text{supp}(\hat{\gamma}^j_u) \quad (2.7)$$

where $\text{supp}(\hat{\gamma}^j_u)$ represents the set of indices associated with non-zero coefficients in the weighted lasso solution.

Step 4. Logit of Y^u on union of variables selected in steps 1 or 3

Obtain the post-double selection logit coefficients

$$(\check{\theta}_u, \check{\beta}_u) \in \arg\min_{\theta,\beta} \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig} \, M_u(W_{ig}, \theta, \beta) : \text{supp}(\theta, \beta) \subseteq \text{union}_u \quad (2.8)$$

for all the variables that were selected in steps 1 or 3, i.e. the indices in

$$\text{union}_u = \text{supp}(\hat{\theta}_u, \hat{\beta}_u) \;\cup\; \bigcup_{j=1}^{\tilde{p}} \text{supp}(\hat{\gamma}^j_u). \quad (2.9)$$

The resulting post-double selection point estimates $\check{\theta}_{u1}, \ldots, \check{\theta}_{u\tilde{p}}$ measure the impact of target regressors $D_1, \ldots, D_{\tilde{p}}$ on Y^u at threshold u.

Note that we differ from the post-double selection algorithm of Belloni et al. (2018a) in that we estimate all target parameters jointly in step 4 above, instead of running $\tilde{p}$ separate post-double logits. In the present study, this turns out to be a computationally attractive modification because we have a small number of target regressors and the cardinality of the union of selected control variables is still substantially smaller than the sample size.

2.2. Inference

In order to compute pointwise and simultaneous confidence intervals for sets of coefficients $(\theta_{uj}), u \in \mathcal{U}, j \in \mathcal{J}$ we use the following procedures.

The Neyman-orthogonal moment equation for target parameter θ_{uj} is given by

$$\psi_j(W, \theta_u, \eta_u) = \left\{Y^u - \Lambda(D\theta_u + X\beta_u)\right\} \cdot \left(D_j - \tilde{X}^j \gamma_u^j\right) \tag{2.10}$$

where the nuisance parameters are collected in $\eta_u = (\beta_u, \gamma_u^1, \dots, \gamma_u^{\tilde{p}})$.

Define

$$\psi(W, \theta_u, \eta_u) = (\psi_1(W, \theta_u, \eta_u), \dots, \psi_{\tilde{p}}(W, \theta_u, \eta_u))' \tag{2.11}$$

and the Jacobian matrix

$$J(W, \theta_u, \eta_u) = \frac{\partial \psi(W, \theta_u, \eta_u)}{\partial \theta_u}. \tag{2.12}$$

As shown in Belloni et al. (2018b), the post-double selection procedure enforces the sample moment condition

$$\frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, \psi(W_{ig}, \check{\theta}_u, \check{\eta}_u) = 0. \tag{2.13}$$

Conditional on the estimates for the nuisance parameters $\check{\eta}_u$, an expansion for the target parameters $\check{\theta}_{u1}, \dots, \check{\theta}_{u\tilde{p}}$ yields a consistent estimate of their asymptotic variance matrix

$$\hat{\Sigma}^u = \hat{J}_u^{-1} \hat{B}_u \hat{J}_u^{-1'} \tag{2.14}$$

with

$$\hat{J}_u^{-1} = \Big[\frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, J(W_{ig}, \check{\theta}_u, \check{\eta}_u)\Big]^{-1} \tag{2.15}$$

$$\hat{B}_u = \Big[\frac{1}{G} \sum_{g=1}^{G} \Big(\sum_{i=1}^{n_g} v_{ig}\, \psi(W_{ig}, \check{\theta}_u, \check{\eta}_u)\Big)\Big(\sum_{i=1}^{n_g} v_{ig}\, \psi(W_{ig}, \check{\theta}_u, \check{\eta}_u)\Big)'\Big], \tag{2.16}$$

where the inner part of matrix $\hat{B}_u$ accounts for the clustering of observations. The estimated asymptotic variance of target parameter θ_{uj} is given by $\hat{\sigma}_{uj}^2 = (\hat{\Sigma}_{j,j}^u)$.

The Neyman-orthogonal moment condition for the target parameters $\check{\theta}_u = (\check{\theta}_{u1}, \dots, \check{\theta}_{u\tilde{p}})'$ was constructed such that

$$\frac{\partial}{\partial \eta} \Big[\frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, \psi(W_{ig}, \check{\theta}_u, \eta)\Big]\Big|_{\eta = \check{\eta}_u} = 0, \tag{2.17}$$

implying that the estimating equations are first-order immune with respect to the nuisance terms, i.e., by constructing an instrument for D_j in (2.10), one has 'partialled-out' the effect of covariates $\tilde{X}^j$.

For the multiplier bootstrap procedure, define the $\tilde{p}$-dimensional estimated influence function

$$\text{influence}(W_{ig}, \check{\theta}_u, \check{\eta}_u) = -\hat{J}_u^{-1} \psi(W_{ig}, \check{\theta}_u, \check{\eta}_u). \tag{2.18}$$

The multiplier bootstrap critical value c_α is computed as the $1 - \alpha$ quantile of the distribution of

$$\hat{s} = \sup_{u\in\mathcal{U}, j\in\mathcal{J}} \frac{1}{\sqrt{G}\hat{\sigma}_{uj}} \sum_{g=1}^{G} \xi_g \cdot \Big[\sum_{i=1}^{n_g} v_{ig} \cdot \text{influence}_j(W_{ig}, \check{\theta}_u, \check{\eta}_u)\Big],\ \xi_g \sim \text{iid } N(0,1) \quad (2.19)$$

where $\text{influence}_j(W_{ig}, \check{\theta}_u, \check{\eta}_u)$ is the j-th component of $\text{influence}(W_{ig}, \check{\theta}_u, \check{\eta}_u)$. The distribution of $\hat{s}$ can be obtained by repeatedly drawing weights ξ_g from the standard normal distribution.

The bootstrap critical value c_α is then used as a scaling factor for the pointwise confidence regions which results in a simultaneous confidence band covering multiple target parameters $(\theta_{uj}), j \in \mathcal{J}$ at multiple distribution thresholds $u \in \mathcal{U}$ with probability $1-\alpha$, i.e.,

$$P\Big(\check{\theta}_{uj} - c_\alpha \frac{\hat{\sigma}_{uj}}{\sqrt{G}} \ \le\ \theta^0_{uj} \ \le\ \check{\theta}_{uj} + c_\alpha \frac{\hat{\sigma}_{uj}}{\sqrt{G}} \ \textit{for all } u \in \mathcal{U}, j \in \mathcal{J}\Big) \approx 1 - \alpha. \quad (2.20)$$

3. DATA AND IMPLEMENTATION

3.1. Data sources and specification

Our empirical analysis is based on the German Socio-Economic Panel Study (SOEP, v35) which is a long-running survey providing representative information about the German population.[3] We use information for the years 2011 to 2018 covering the years around the introduction of the German minimum wage on the 1st January 2015. The strength of a survey like the SOEP is the wealth of information that can be used as covariates. A weakness is the moderate sample size of around 11,000 wage earners per year, which motivates the use of specification selection methods. Our final sample after sample selection criteria is around 90,000 covering the years 2011 to 2018. We exploit only the cross-sectional information in the SOEP as the sample is highly unbalanced due to frequent refreshment samples and permanent sample dropout. We use the sampling weights provided with the SOEP that ensure that cross-sectional information is representative for the German population in the given year. We exclude from our sample individuals who are not subject to the minimum wage (the self-employed, students, apprentices, interns and similar groups).

Following the minimum wage literature after the seminal contribution by Card (1992), we measure the effects of the minimum wage introduction by a continuous treatment indicator *bite* representing the fraction of workers in population subgroups with wages below the minimum wage level of 8.50 euros/hour *before the minimum wage was introduced.* The idea is that the changes induced by the minimum wage should be largest in population subgroups that had the strongest exposure to the new minimum wage level, controlling for other things. As the sample size of the SOEP would be too low to construct reliable bite measures for small population subgroups, we take our bite measure from a larger data set, the German Structure of Earnings Survey (GSES). The bite measure used here is defined at the 2-digit industry level differentiated by East/West Germany. Measuring the bite at the industry level fits well industrial relations in Germany as a large part of wage bargaining takes place at this level. Our bite measure varies between

[3] See Schröder et al. (2020).

.003 and .701, providing large variation to measure the effects induced by the exposure to the newly introduced minimum wage.[4]

We measure the effects of the minimum wage introduction on the distribution of our outcome variables $Y \in \{\text{hourly wage, hours worked, monthly earnings}\}$ by the specification

$$\begin{aligned} P(Y < u|D, X) = \Lambda(\, & \theta_{2013/14} \cdot [bite\#1_{year=2013/14}] \\ & + \theta_{2015/16} \cdot [bite\#1_{year=2015/16}] \\ & + \theta_{2017/18} \cdot [bite\#1_{year=2017/18}] + X\beta_u), \quad u \in \mathcal{U}, \end{aligned} \tag{3.1}$$

where the symbol # represents interaction of the treatment variable *bite* with the year of observation. In order to keep the number of target coefficients low, we combine two adjacent years. Defining yearly coefficients leads to similar results but with higher volatility over years. The coefficients $\theta_{2015/16}$ and $\theta_{2017/18}$ represent the treatment effects of the minimum wage introduction on the likelihood of falling below a particular threshold u in the outcome distribution. They measure to what extent, e.g., hourly wages below a particular level u became more or less frequent *after the minimum wage introduction* per unit of exposure to the newly introduced minimum wage, controlling for other characteristics X that are relevant for explaining that a particular wage observation falls below threshold u (e.g., work experience, education, occupational characteristics, time effects etc., see below). In other words, the coefficients $\theta_{2015/16}$ and $\theta_{2017/18}$ represent the isolated effects of the minimum wage introduction after controlling for factors affecting both the outcome and the exposure to the minimum wage introduction at different points of the distribution. The coefficient $\theta_{2013/14}$ provides a pre-treatment test as it measures to what extent differences already emerged between high and low exposure groups *before the minimum wage was introduced* (compared to the omitted time period 2011/12).

3.2. Variables and feature engineering

Our dependent variables are derived from the survey information on monthly earnings and actual hours worked per week (including overtime). Monthly earnings and actual hours worked per week are taken as they appear in the survey. Hourly wages are computed as monthly earnings divided by monthly hours worked (defined as weekly hours times 4.345).

Table B.1 in the Appendix describes the information that is used to construct the set of control candidates (features) from which the ℓ_1-methods can choose relevant elements for predicting the nuisance terms at each threshold. The total number of features constructed in this way is several thousands as we not only include transformations of continuous variables (polynomial terms, square root, log) and indicators for potentially important individual values of continuous variables (e.g., an indicator for having an unemployment experience of zero years), but interactions and full sets of indicators for all our categorial variables. Take the example of an educational classification with five categories. In this case, we include a full set of five indicators describing the membership in each category (no

[4]For more details on the bite measure used here, see Biewen et al. (2022). An alternative would be to define the bite at the level of labour market regions but this faces the difficulty that the coverage of labour market regions in a survey like the SOEP is patchy and that regional information in the SOEP can only be processed on-site with limited computational facilities.

omitted category). The lasso can then flexibly pick the indicators that help to remove the omitted variable bias for explaining the effect of the treatment variables at a particular threshold.[5] It is important not to omit a reference category when constructing sets of such indicators as exactly the omitted category could be the one prefered by the lasso (the information represented by the omitted category could be re-constructed as a linear combination of other categories, but this runs counter to the idea of finding a sparse approximation for the nuisance term). In a similar way, we offer to the lasso nested or overlapping information from classifications of higher or lower aggregation levels from which it can choose the information that is most suitable to remove omitted variable bias. For example, we include occupation codes at different aggregation levels (1-digit, 2-digit etc.) and nested or partly overlapping education classifications that offer finer or more coarse information (see Table B.1).

In order to arrive at the final set of potential covariates offered to the lasso, we eliminate from the full set of features described in Table B.1 i) constant features, ii) duplicates/multiples of other features, iii) features that uniquely characterize less than 1 percent of our sample. We can relax restriction iii) to a certain extent without changing results in important ways. However, in our experience doing this increases the likelihood of perfect prediction problems and convergence issues in the logit models that counter the motivation of this paper to find a fully automatic way to pick controls at the typically large number of thresholds without having to fix problems or eliminate features by hand at individual thresholds. Applying the above criteria, the final number of features included in our estimations was around 2,500 (the exact number of features depends on the outcome variable as features related to working hours cannot be included for monthly earnings and hours worked due to perfect prediction issues). This is clearly too large for individual specification searches at one given threshold, let alone at the typically around 40-50 thresholds used per dependent variable in our application.

3.3. Details on lasso implementation

As described in section 2, our methods allow for clustering of observations in two ways: i) for statistical inference and ii) for the choice of lasso penalties. As to i), it is well-known that in difference-in-differences-like designs, it is necessary to cluster at the level of the treatment variable, see, e.g. Abadie et al. (2022). Our treatment variable *bite* is based on the combination of 2-digit industries and East/West information giving us 153 population subgroups at the level of which we cluster in all inference procedures (estimation of variance matrices and draws for multiplier bootstrap). We initially also tried to cluster the lasso penalties at this level but found that this led to quite erratic and volatile results across different thresholds and coefficients which did not seem plausible. This behaviour of the lasso is not surprising given the relatively low number of clusters in our application and their sometimes chunky nature. For computing lasso penalty loadings, we therefore clustered at the level of the panel units, which is standard for panel data; see, e.g. Belloni et al. (2016b), Ahrens et al. (2020).

As mentioned above, it is useful to offer potentially multicollinear control variables to the lasso (nested information, full sets of indicators etc.) to allow the lasso to extract the

[5] For categorial variables, we also define a category 'missing value' that may also be picked by the lasso if it helps to predict the treatment or the outcome variable. This also helps to conserve the number of observations as observations with missings in these variables do not have to be discarded.

information that is most suitable for removing omitted variable bias. It is well known that the lasso solution need not be unique if the feature set contains mostly discrete (binary) variables as in our case, see Tibshirani (2013). To evade numerical difficulties, we implemented our post-ℓ_1-estimators using the Moore-Penrose pseudo-inverse.

In our application, the cardinality of the active set of control variables was between 60 and 80 depending on the threshold. The double-selected features consistently included time effects, various information on educational qualifications and work experience as well as additional controls that differed across thresholds in plausible ways (e.g., indicators for low occupational positions/job types at lower thresholds, information on firm characteristics or particular educational/occpuational qualifications at medium or upper thresholds, interactions of such characteristics with gender or East/West Germany at particular thresholds).

4. EMPIRICAL RESULTS

4.1. Econometric analysis

We start with the impact of the minimum wage on the likelihood of falling below certain thresholds in the *hourly wage distribution* after the minimum wage introduction as shown in Figure 1. The subfigures display the coefficients on the bite variable *before* (2013/14) and *after* the minimum wage was introduced (2015/16 and 2017/18). The results show that, as intended by policymakers, the likelihood of having hourly wages lower than 8.5 euros per hour was significantly lower in groups with high minimum wage exposure after the introduction (Figures 1bc) but not before the introduction (Figure 1a). The picture shows how the likelihood of wages below 8.5 euros per hour was gradually reduced in groups with high minimum wage exposure right after the introduction in 2015/16, and even more clearly in later years 2017/18. There is no evidence for spill-over effects over the minimum level of 8.50 euros/hour. This is in line with Caliendo et al. (2022) based on the same data set, but not with Biewen et al. (2022) and Bossler and Schank (2023) who find some evidence for spill-over effects based on other data sources.[6]

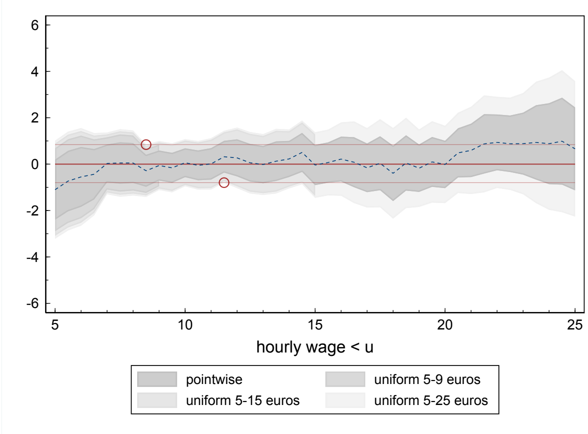
Figure 1 features various simultaneous confidence intervals based on the multiplier bootstrap referring to the coefficient process over increasing parts of the distribution. By construction, the simultaneous bands become wider as the set of coefficients is extended, but the main increase occurs when going from pointwise to simultaneous intervals. The uniform confidence bands can be used to carry out hypothesis test. All three subfigures of Figure 1 display the zero line. If the zero line is not contained in the simultaneous band, we can reject the hypothesis that the minimum wage had no effect on hourly wages. This hypothesis is not rejected in the pre-introduction period 2013/14 (Figure 1a), but rejected in the post-introduction periods 2015/16 and 2017/18 (Figures 1bc). One can also test whether effects are homogenous across the distribution. For this, one has to check whether the lowest value of the upper confidence band lies below the highest value of the lower confidence band (in the graphs, these points are symbolized by small circles). Effect homogeneity can be rejected for the post-introduction period 2017/18 (Figure 1c), but not for periods 2013/14 and 2015/16 (Figures 1ab).

The methodology in Belloni et al. (2018a) allows us to not only include in confidence intervals the whole process of one coefficient but potentially also the processes of other

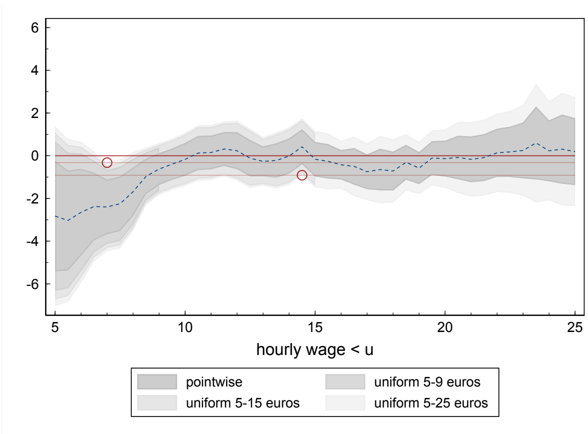
[6]Note that the minimum wage level was initially set at 8.50 euros/hour in 2015, but was increased to 8.84 euros/hour in 2017.

Figure 1: Minimum wage effects in hourly wage distribution

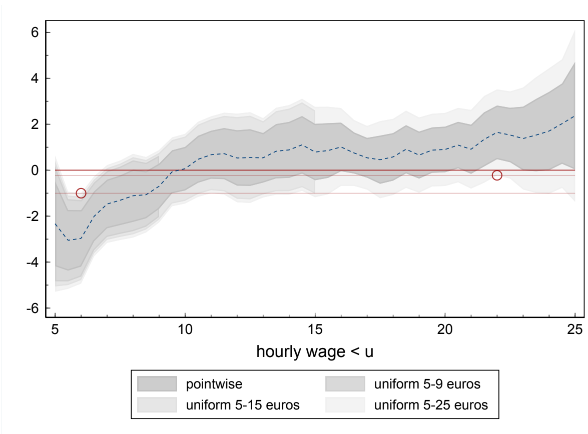
(a) Coefficient process $bite\#1_{year=2013/14}$ (pre-treatment)



(b) Coefficient process $bite\#1_{year=2015/16}$ (post-treatment)



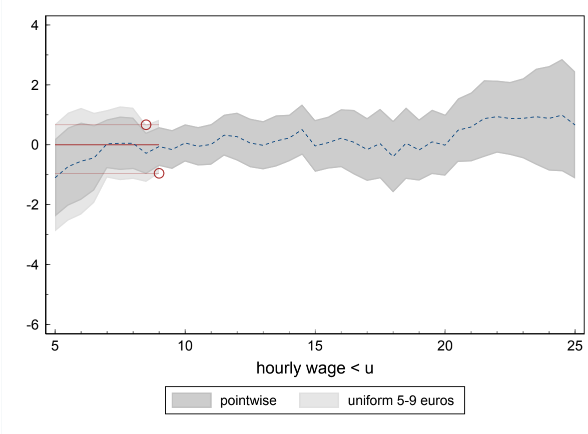
(c) Coefficient process $bite\#1_{year=2017/18}$ (post-treatment)



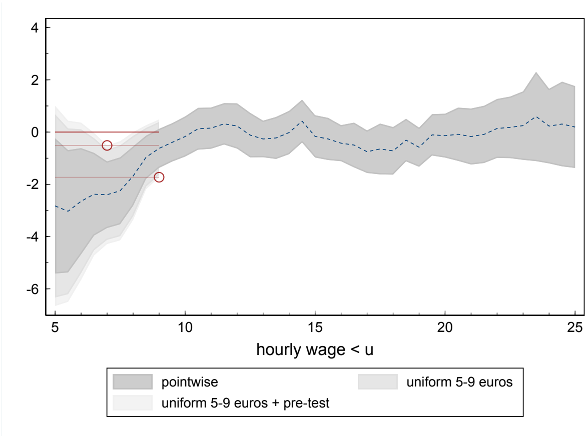
Note: 90 % uniform confidence bands based on multiplier bootstrap (100,000 replications). The intervals refer to increasing ranges of coefficients. The circles mark the lowest value of the upper confidence band and the highest value of the lower confidence band.

Figure 2: Hourly wage effects: confidence bands including pre-test

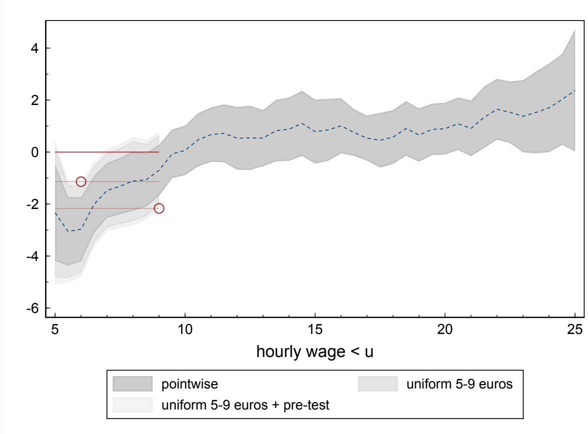
(a) Coefficient process $bite\#1_{year=2013/14}$ (pre-treatment)



(b) Coefficient process $bite\#1_{year=2015/16}$ (post-treatment)



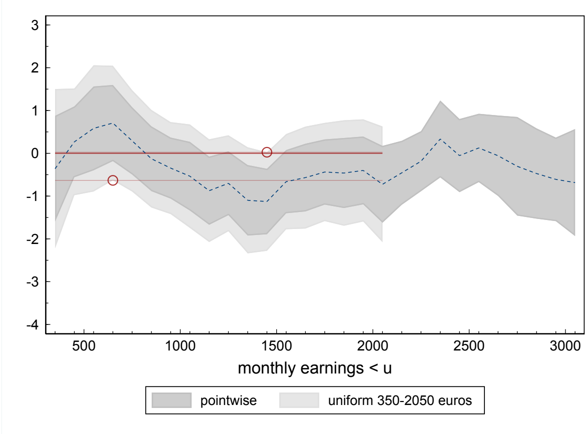
(c) Coefficient process $bite\#1_{year=2017/18}$ (post-treatment)



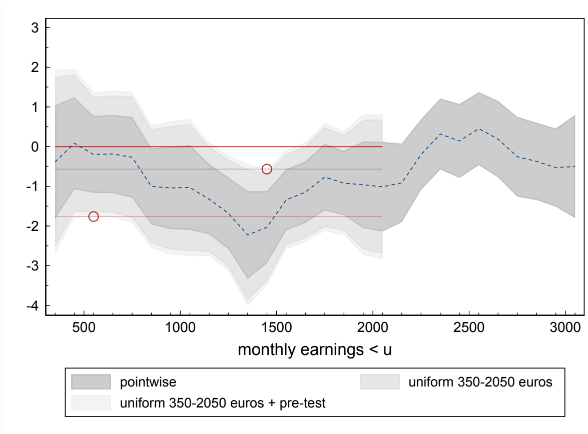
Note: 90 % uniform confidence bands based on multiplier bootstrap (100,000 replications). The additional intervals in the lower two panels simultaneously cover the coefficient process of the respective period and the coefficients of the pre-test period 2013/14 (up to 9 euros/hour). The circles mark the lowest value of the upper confidence band and the highest value of the lower confidence band.

Figure 3: Minimum wage effects in monthly earnings distribution

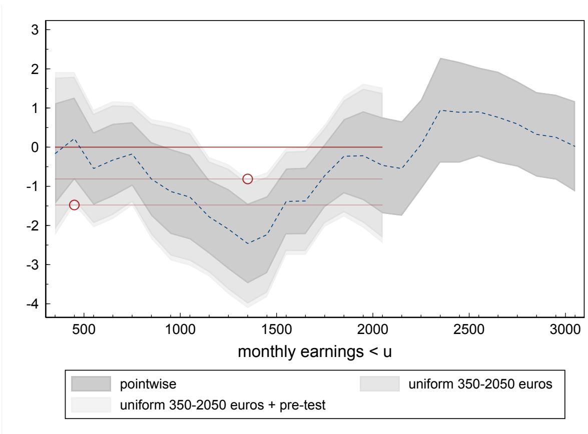
(a) Coefficient process $bite\#1_{year=2013/14}$ (pre-treatment)



(b) Coefficient process $bite\#1_{year=2015/16}$ (post-treatment)



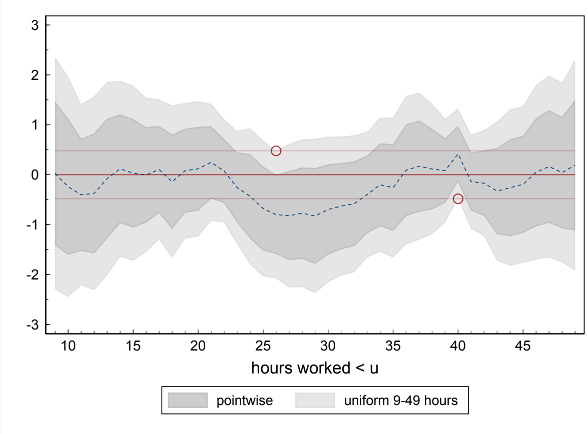
(c) Coefficient process $bite\#1_{year=2017/18}$ (post-treatment)



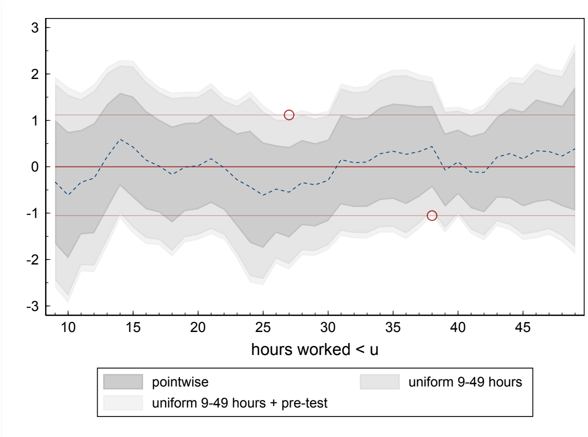
Note: 90 % uniform confidence bands based on multiplier bootstrap (100,000 replications). The intervals refer to the monthly earnings range relevant to minimum wage recipients (350 to 2050 euros). The additional intervals in the lower two panels simultaneously cover the coefficient process of the respective period and the coefficients of the pre-test period 2013/14. The circles mark the lowest value of the upper confidence band and the highest value of the lower confidence band.

Figure 4: Minimum wage effects in working hours distribution

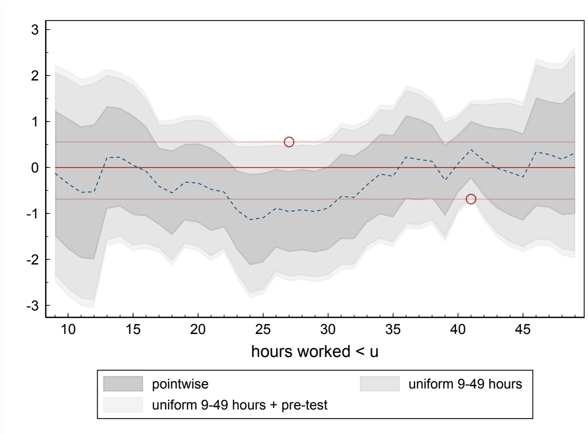
(a) Coefficient process $bite\#1_{year=2013/14}$ (pre-treatment)



(b) Coefficient process $bite\#1_{year=2015/16}$ (post-treatment)



(c) Coefficient process $bite\#1_{year=2017/18}$ (post-treatment)



Note: 90 % uniform confidence bands based on multiplier bootstrap (100,000 replications). The additional intervals in the lower two panels simultaneously cover the coefficient process of the respective period and the coefficients of the pre-test period 2013/14. The circles mark the lowest value of the upper confidence band and the highest value of the lower confidence band.

coefficients, see equation (2.20). In Figures 2bc, we incorporate into the simultaneous interval the cofficients for the potential pre-trends from Figure 2a. For clarity, we focus on the range of hourly wages 5-9 euros/hour at which the introduction of the minimum wage was targeted. The figure illustrates the point made by Roth (2022) that conditioning on the result of a pre-test leads to an undercoverage of confidence intervals. It shows the 'cost' in interval coverage for our coefficients of interest caused by previously testing for the existence of pre-trends (up to 9 euros/hour). In our application, this cost is small but non-negligible.

In Figure 3, we turn to the effects of the minimum wage introduction in the distribution of *monthly earnings*. This analysis is interesting as it allows one to assess which workers in the personal income distribution benefited most from the introduction. Our simultaneous confidence intervals in Figure 3 are constructed such that they cover the range up to 2,050 euros per month as it is unlikely that individuals affected by the minimum wage earn more than this amount per month (wage rates around the minimum wage would require working hours of over 50 hours per week to achieve monthly earnings over 2,050 euros). It turns out that it were not earners of very low monthly wages (i.e., part-time and marginal part-time workers), but individuals with monthly earnings between 1,200 and 1,700 euros/month who gained most. These are most likely full-time workers receiving very low hourly wages before the minimum wage introduction (a full-time worker working 40 hours a week at the minimum wage receives a monthly wage of 8.50 x 40 x 4.345 = 1,477 euros). Our results are consistent with the results in Bossler and Schank (2023) based on administrative data, who also find the largest effects of the minimum wage introduction on the 20th to 30th percentile of their monthly wage distribution (Bossler and Schank, 2023, table 2), roughly corresponding to the range of 1,200 to 1,700 euros as above, but no or very small effects for monthly wages between the so-called minijob threshold of 450 euros and a monthly wage of around 1,000 euros (the latter could be the result of a minimum wage recipient working part-time).[7] Note that, while the confidence bands in Figure 3 reject the hypothesis of no minimum wage effects on monthly earnings, we cannot reject the hypothesis that effects are homogeneous (the lowest point of the upper confidence band is not below the highest point of the lower band).

An open question from Figures 2 and 3 is whether changes in the distribution of monthly earnings were only caused by changes in wage rates or whether the minimum wage also reduced working hours, potentially dampening its effect on monthly earnings as suggested by Burauel (2019,2020) and Caliendo (2022). Our results in Figure 4 suggest *no evidence* for significant changes in the distribution of working hours due to the minimum wage, which is in line with Biewen et al. (2022), but not with Burauel (2019,2020) and Caliendo (2022). Confidence intervals bands consistently cover the zero effect line and the overall pattern of point estimates is stable across pre- and post-treatment periods, indicating no changes post- vs. pre-treatment. As a consequence, our results suggest that the introduction of the minimum wage did not induce significant reductions in weekly working hours in order to keep monthly wage bills constant, or to significant shifts between different types of jobs (part-time, marginal part-time, full-time).

[7] Bossler and Schank (2023)'s analysis based on administrative data find additional significant effects for very low wages *below the minijob threshold of 450 euros per month.* Unfortunately, this range is insufficiently covered in our data and could not not be represented as additional thresholds in our distribution regression model.

5. CONCLUSION

This paper uses a distribution regression model to evaluate the effects of the introduction of the German minimum wage in 2015 on the distribution of hourly wages, hours worked and monthly earnings. Our data source is the German Socio-Economic Panel (SOEP) which is characterized by a moderate sample size but a large number of potential control variables. We measure the effects of the minimum wage at each point of our outcome distributions employing flexible machine learning methods recently developed by Belloni et al. (2018a). These methods allow us to automatically specify a large number of parallel logit models over a fine grid of distributional thresholds, while providing valid statistical inference across ranges of thresholds, after a comprehensive, machine-led specification search which is unaffected by subjective decisions and pre-tested specification choices. Our distribution regression analysis provides a more comprehensive picture about the points of the distribution at which the minimum wage had an effect compared to previous contributions. Our results suggest that the minimum wage displaced hourly wages below its minimum level, benefitted monthly wages in the lower middle but not the very low part of the distribution and did not lead to significant changes in the distribution of working hours. This implies that changes in hourly wages fully translated into increases in monthly wages for minimum wage recipients. Our results help reconcile previously conflicting results in the literature on the effects of the German minimum wage introduction based on different data sources.

ACKNOWLEDGEMENTS

We are grateful to Marco Caliendo, Bernd Fitzenberger, Martin Huber, Michael Knaus, Marian Rümmele, Uta Schönberg, Martin Spindler as well as participants at EALE 2023 and at the 'Machine Learning Meets Quantitative Social Science' Workshop at the University of Tübingen for many valuable comments and discussions. We also thank Nicolas Erhardt for excellent research assistance. This paper uses data from the Socio-Economic Panel study SOEP-Core, v35, doi:10.5684/soep-core.v35. Martin Biewen is a member of the Machine Learning Cluster of Excellence, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2064/1 - Project 390727645.

REFERENCES

Abadie, A., S. Athey, G. Imbens and J. Wooldridge (2022). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, *138*, 1–35.

Ahrens, A., C.B. Hansen, A.B. Schaffer (2020). lassopack - model selection and prediction with regularized regression in Stata. *Stata Journal*, *20*, 176–235.

Belloni, A., V. Chernozhukov and Y. Wei (2016a). Post-Selection Inference for Generalized Linear Models With Many Controls. *Journal of Business and Economic Statistics*, *34*, 606–619.

Belloni, A., V. Chernozhukov, C. Hansen and D. Kozbur (2016b). Inference in High-Dimensional Panel Models With an Application to Gun Control. *Journal of Business and Economic Statistics*, *34*, 590–605.

Belloni, A., V. Chernozhukov, D. Chetverikov and Y. Wei (2018a). Uniformly valid post-regularization confidence regions for many functional parameters in a z-estimation framework. *Annals of Statistics*, *46*, 3643–3675.

Belloni, A., V. Chernozhukov, D. Chetverikov and Y. Wei (2018b). *Supplement to "Uniformly valid post-regularization confidence regions for many functional parameters in a z-estimation framework".*

Biewen, M., B. Fitzenberger and M. Rümmele (2022). Using Distribution Regression Difference-in-Differences to Evaluate the Effects of a Minimum Wage Introduction on the Distribution of Hourly Wages and Hours Worked. *IZA Discussion Paper No. 15534*, IZA Institute of Labor Economics, Bonn.

Bossler, M. and T. Schank (2023). Wage Inequality in Germany after the Minimum Wage Introduction. *Journal of Labor Economics*, *41*, 813–855.

Burauel, P., M. Caliendo, M. Grabka, C. Obs, M. Preuss and C. Schröder (2019). The Impact of the German Minimum Wage on Individual Wages and Monthly Earnings. *Journal of Economics and Statistics*, *240*, 201–231.

Burauel, P., M. Caliendo, M. Grabka, C. Obs, M. Preuss and C. Schröder (2020). The Impact of the Minimum Wage on Working Hours. *Journal of Economics and Statistics*, *240*, 233–267.

Caliendo, M., C. Schröder and L. Wittbrodt (2019). The Causal Effects of the Minimum Wage Introduction in Germany - An Overview. *German Economic Review*, *20*, 257–292.

Caliendo, M., A. Fedorets, M. Preuss, C. Schröder and L. Wittbrodt (2022). The short- and medium-term distributional effects of the Germany minimum wage reform. *Empirical Economics*, *64*, 1149–1175.

Card, D. (1992). Using regional variation in wages to measure the effects of the federal minimum wage. *Industrial Labor Relations Review*, *46*, 22–37.

Chernozhukov, V., I. Fernandez-Val, B. Melly (2013). Inference on Counterfactual Distributions. *Econometrica*, *81*, 2205–2268.

Chiang, H.D. (2020) *Three Essays in Cluster Robust Machine Learning and High-Dimensional Econometrics.* Ph.D. dissertation, Graduate Schoool Vanderbilt University.

Dustmann, C., A. Lindner, U. Schönberg, M. Umkehrer, P. vom Berge (2022). Reallocation Effects of the Minimum Wage. *Quarterly Journal of Economics*, *137*, 267–328.

Foresi, S. and F. Peracchi (1995). The Conditional Distribution of Excess Returns: An Empirical Analysis. *Journal of the American Statistical Association*, *90*, 451–466.

Mindestlohnkommission (2020). *Dritter Bericht zu den Auswirkungen des gesetzlichen Mindestlohns.* Mindestlohnkommission, Bundesamt für Arbeitsschutz und Arbeitsmedizin (BAuA), Berlin.

Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights*, *4*, 305–322.

Schröder, C., J. König, A. Fedorets, J. Goebel, M. Grabka, L. Lüthen, M. Metzing, F. Schikora, S. Liebig (2020). The economic research potential of the German Socio-Economic Panel study. *German Economic Review*, *31*, 335–371.

Tibshirani, R.J. (2013). The lasso problem and uniqueness. *Electronic Journal of Statistics*, *7*, 1456–1490.

APPENDIX A: CHOICE OF PENALTY LEVELS AND LOADINGS

In order to allow for a different choice of observation clusters when computing penalties as opposed to computing standard errors and confidence intervals, we define in the present section clusters as $g^* = 1, \dots, G^*$ with observations $j = 1, \dots, n^{g^*}$ (as opposed to the main text, where we defined clusters as $g = 1, \dots, G$ with observations $i = 1, \dots, n_g$). In addition, define the $(\tilde{p} + p)$-dimensional vector $\tilde{X} = (D, X)$ and the $((\tilde{p} - 1) + p)$-dimensional vector $\tilde{X}^j = (D_{\mathcal{J}\setminus j}, X)$. The algorithms for setting penalty levels and loadings are due to Belloni et al. (2018a) with modifications for clustering of observations as in Chiang (2020). In addition, we incorporate sampling weights v_{ig}.

A.1. Penalty level and loadings for logistic lasso

STEP 1. Initialize procedure

Set $\bar{m}$ = the maximal number of iterations and define constants
$\lambda_1 = c\sqrt{G^*}\,\Phi^{-1}(1 - \gamma/2(\tilde{p} + p)), \quad c = 1.1, \quad \gamma = 0.1/\log(G^*)$.

Set $m = 0$. Starting from a constant-only model, determine the five features that have the maximal *ex-ante* gradient by absolute value. These represent the five features that are most promising for reducing the prediction error. Fill their logit coefficients into $\tilde{\theta}_u^m, \tilde{\beta}_u^m$ (set other entries to zero).

STEP 2. Iterative determination of penalty and loadings

Set

$$\hat{l}_{ujk,m+1} = \Big\{ \frac{1}{G^*} \sum_{g^*=1}^{G^*} \Big[\sum_{j=1}^{n^{g^*}} v_{ig}\,(Y_{ig}^u - \Lambda(D_{ig}\tilde{\theta}_u^m + X_{ig}\tilde{\beta}_u^m))\tilde{X}_{kig} \Big]^2 \Big\}^{\frac{1}{2}}$$

$$\hat{\Psi}_u^{m+1} = \mathrm{diag}\Big(\hat{l}_{ujk,m+1}, k = 1, \dots, (\tilde{p} + p)\Big)$$

Run logistic lasso to obtain refined lasso coefficients $\tilde{\theta}_u^{m+1}, \tilde{\beta}_u^{m+1}$.

If $\max_k |\hat{l}_{ujk,m+1} - \hat{l}_{ujk,m}| <$ tolerance or $m = \bar{m}$ then stop. Otherwise set $m = m+1$ and repeat step 2.

A.2. Penalty level and loadings for WLS lasso

STEP 1. Initialize procedure

Set $\bar{m}$ = the maximal number of iterations and define constants
$\lambda_2 = c\sqrt{G^*}\,\Phi^{-1}(1 - \gamma/2(\tilde{p} + p)(\tilde{p} + p - 1)), \quad c = 1.1, \quad \gamma = 0.1/\log(G^*)$

Starting from a constant-only model, determine the five features that have the maximal *ex-ante* gradient by absolute value. These represent the five features that are most promising for reducing the prediction error. Fill in their WLS coefficients into $\tilde{\gamma}_u^{j,m}$ (set other entries to zero).

Step 2. Iterative determination of penalty and loadings

Set

$$\hat{l}_{ujk,m+1} = \left\{ \frac{1}{G^*} \sum_{g^*=1}^{G^*} \Big[\sum_{j=1}^{n^{g^*}} v_{ig} \, \hat{f}^{2,m}_{uig} (D_{jig} - \tilde{X}^{j}_{ig} \tilde{\gamma}^{j,m}_{u}) \tilde{X}^{j}_{kig} \Big]^2 \right\}^{\frac{1}{2}}$$

$$\hat{\Psi}^{m+1}_{uj} = \text{diag}\left(\hat{l}_{ujk,m+1}, k = 1, \ldots, (\tilde{p} - 1) + p \right)$$

Run WLS lasso to obtain refined lasso coefficients $\hat{\gamma}^{j,m+1}_{u}$.

If $\max_k |\hat{l}_{ujk,m+1} - \hat{l}_{ujk,m}| <$ tolerance or $m = \bar{m}$ then stop. Otherwise set $m = m+1$ and repeat step 2.

APPENDIX B: CONSTRUCTION OF FEATURES FOR LASSO

Table B.1: Variables and transformations included in the double selection algorithm.

Variable	Type	Transformations included
Treatment variable		
Bite	continuous	linear
Time effects		
Year	categorical(8)	indicators for each category
Worker characteristics 1		
Gender	categorical(2)	indicators for each category
East/West Germany	categorical(2)	indicators for each category
Worker characteristics 2		
Age	continuous	1st, 2nd, 3rd, 4th power, square root, log
Years of education	continuous	1st, 2nd, 3rd, 4th power, square root, log
Full-time experience (years)	continuous	1st, 2nd, 3rd, 4th power, square root
Part-time experience (years)	continuous	1st, 2nd, 3rd, 4th power, square root
Ft experience + 0.5 Pt experience	continuous	1st, 2nd, 3rd, 4th power, square root
Tenure (years)	continuous	1st, 2nd, 3rd, 4th power, square root
Overtime (hours/week)	continuous	1st, 2nd, 3rd, 4th power, square root indicator for no overtime
Unemployment experience (years)	continuous	1st, 2nd, 3rd, 4th power, square root indicator for no unemployment experience
Worker characteristics 3		
Type of school degree	categorical(9)	indicators for each category
Type of vocational training degree	categorical(7)	indicators for each category
Type of tertiary degree	categorical(11)	indicators for each category
Fine type of tertiary degree	categorical(23)	indicators for each category
Variants of no educational degree	categorical(4)	indicators for each category
ISCED classification of educational degree	categorical(10)	indicators for each category
5-group categorization of German education system	categorical(5)	indicators for each category
3-group categorization of German education system	categorical(3)	indicators for each category
ISCO08 occupation code (2-digit)	categorical(40)	indicators for each category
ISCO08 occupation code (3-digit)	categorical(121)	indicators for each category
KldB2010 occupation code (1-digit)	categorical(10)	indicators for each category
KldB2010 occupation code (2-digit)	categorical(37)	indicators for each category
Occupational position	categorical(12)	indicators for each category
NACE industry code (1-digit)	categorial(18)	indicators for each category
NACE industry code (2-digit)	categorial(86)	indicators for each category
Full-time/part-time/marginal part-time	categorical(3)	indicators for each category indicator for pt/mpt combined
Minjob contract	categorical(2)	indicators for each category
Firm size categorization I (coarse)	categorical(5)	indicators for each category
Firm size categorization II (finer)	categorical(8)	indicators for each category
Public sector	categorical(3)	indicators for each category

Federal state	categorical(16)	indicators for each category
Urban area	categorical(2)	indicators for each category
Nationality (continents)	categorical(5)	indicators for each category
Nationality (subcontinents)	categorical(12)	indicators for each category
Nationality (countries)	categorical(116)	indicators for each category indicator for German nationality
Household size	count(16)	1st power, indicator each category
Partner lives in household	categorical(3)	indicators for each category
Marital status (single/divorced/widowed etc.)	categorical(9)	indicators for each category
# children in household aged 0-2 years	count	1st power, indicator for zero
# children in household aged 3-5 years	count	1st power, indicator for zero
# children in household aged 6-11 years	count	1st power, indicator for zero
# children in household aged 12-17 years	count	1st power, indicator for zero
Person in need of care lives in household	categorial(3)	indicators for each category
Homeowner/renter (with subcategories)	categorical(5)	indicators for each category
Health indicator	ordinal(6)	1st power, indicator for highest two values indicator for lowest two values
Interactions		
Age # Household size	full expansion of age transformations with household size	
Age # (Worker characteristics 2)	full expansion of features with age transformations	
Household size # (Worker characteristics 2)	full expansion of features with household size	
Gender # (Worker characteristics 2)	full expansion of features with gender indicators	
Gender # (Worker characteristics 3)	full expansion of features with gender indicators	
East/West # (Worker characteristics 2)	full expansion of features with East/West indicators	
East/West # (Worker characteristics 3)	full expansion of features with East/West indicators	

Note: Number of categories in brackets for count and categorical variables.