

Börschlein, Benjamin; Bossler, Mario

Conference Paper

A new machine learning-based treatment bite for long run minimum wage evaluations

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Börschlein, Benjamin; Bossler, Mario (2021) : A new machine learning-based treatment bite for long run minimum wage evaluations, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/242441>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

A new machine learning-based treatment bite for long run minimum wage evaluations

Benjamin Börschlein*

IAB

Mario Bossler[†]

IAB, LASER

March 2021

Abstract

Empirical evaluations of national minimum wages, such as in Germany or the UK, rely on bite measures that capture treatment variation; measured from the incidence (or intensity) of employees paid below the threshold before the minimum wage was introduced or raised. Bite-dependent estimations face the problem of dynamic selection, implying that even in the absence of the minimum wage the bite may have changed over time. We apply a machine learning method from the field of regularized regression to predict the contemporary bite of the German minimum wage, allowing us to address unobserved dynamic selection in an empirical evaluation of long run effects of the minimum wage. Our LASSO predicted bites show clear improvements over simple forward updating of the initial bite, allowing us to estimate contemporary effects of the German minimum wage from 2015 to 2017.

JEL Classification: J31, J38, C49, C21

Keywords: minimum wage, bite, evaluation, dynamic selection, machine learning, LASSO

*Institute for Employment Research (IAB). Address: IAB, Regensburger Str. 100, 90478 Nuremberg, Germany, e-mail: benjamin.boerschlein@iab.de.

[†]Institute for Employment Research (IAB) and Labor and Socio-Economic Research Center of the University of Erlangen-Nuremberg (LASER). Address: IAB, Regensburger Str. 100, 90478 Nuremberg, Germany, e-mail: mario.bossler@iab.de.

We thankfully acknowledge helpful comments at Workshop on Machine Learning in Labor, Education and Health Economics at the University of Erlangen-Nuremberg. We also acknowledge particularly helpful comments and suggestions from Nicole Gürtzgen, Martin Popp, Claus Schnabel, and Matthias Umkehrer. We thank the "Data- and IT-Management"-department at the IAB for data provision.

1. Introduction

The empirical evaluation of minimum wage legislations in ex-post analyses is highly relevant for policy making, especially since economic theory is not conclusive concerning minimum wage effects on employment. Empirical evaluations are particularly challenging in the case of uniform nationwide minimum wages, such as in the UK or Germany, since there is no experimental variation in the height of the minimum wage within the countries. Evaluations typically rely on a bite measure that captures treatment intensity of the minimum wage across individuals or regions. The intuition of the latter is as follows: Mostly independent labor market regions have different wage structures and hence are differentially affected by a minimum wage. When comparing these differentially treated labor markets over time, a difference-in-differences setting should capture the effect of the minimum wage.

This regional approach was first applied by Card (1992) in an evaluation of different minimum wage levels across federal states in the US. But the identification strategy is equally well important for countries in which there is no variation in the nationwide minimum wage, such as in Germany or the UK. For Germany, bite-dependent evaluations that exploit regional variation to evaluate the minimum wage are presented by Ahlfeldt, Roth, and Seidel (2018); Bossler and Schank (2020); Caliendo, Fedorets, Preuss, Schröder, and Wittbrodt (2018); Garloff (2019); Schmitz (2019). Similarly, in the UK, prominent studies exploiting regional variation are presented by Dolton, Bondibene, and Wadsworth (2012); Dolton, Bondibene, and Stops (2015).

Most of these evaluations concentrate on the estimation of short run effects. An empirical explanation for the focus on short run effects is dynamic selection. Unlike the difference-in-differences-based evaluation of e.g. a training course, where the treatment intervention has a fixed time, the bite of the minimum wage may change over time, which is not observed by the researcher. For example, it remains unknown whether a severely treated region should still be assigned a high bite in the long run three or four years after the minimum wage came in force. If treated regions would have experienced a positive wage development in the absence of the minimum wage introduction, they should no longer be assigned the same bite. Therefore, simple forward updating of the initial bite may result in an inaccurate bite measure and consequently result in biased treatment effects. In fact, Bossler and Westermeier (2020) show in a simulation study that the bias can be severe when the treatment bite is measured inaccurately.

In this article, we use a machine learning method to predict the bite of the German minimum

wage in the long run, allowing us to improve the bite variable compared to simple forward updating. We use two different LASSO regression methods to train and validate a bite prediction model on both the regional and the individual level in the time period before the minimum wage was introduced (2010-2014). First, we compare the predictions of our models and the standard bite-forwarding scheme to measured bite values in those years to assess the added value from our predictions. Second, we use our models to predict the bite over the period 2015-2017, allowing us to estimate wage effects from the LASSO-predicted bite measure. Bite prediction rather than simple forward updating allows us to account for dynamic changes in the composition of the treatment and control group over time, and furthermore, it enables us to include labor market entries of workers who have not been observed as part of the labor force when the bite was initially measured in the analysis.

Our results show that LASSO-based prediction of the bite can improve naive forward updating of the bite from the third year onwards when predicting the regional bite. Short run evaluations of the minimum wage are not improved using LASSO-based prediction of the bite. In the long run, however, dynamic selection between treatment and control group becomes more severe, and thus, the advantages of bite prediction come into play.

The article is structured as follows: Section 2 describes the problem of dynamic selection in long run bite-dependent minimum wage evaluations. Section 3 describes the LASSO estimation and the explanatory variables that we use to predict the bite. In section 4, we present results of the bite predictions on the regional and on the individual level. In section 5, we apply the predicted bite in an estimation of wage effects of the German minimum wage. Section 6 concludes.

2. Dynamic selection in bite-dependent minimum wage evaluations

Why is it relevant to predict the bite of the minimum wage and what is the rationale from the application of machine learning methods in that context? We want to know the contemporary bite to estimate an effect of the minimum wage on contemporaneously affected individuals or regions. If we instead use the bite from simple forwarding of the initially measured bite, i.e. from just before the minimum wage introduction, evaluations identify the effect on individuals that were initially affected by the minimum wage introduction. In the long run, the forwarded bite still identifies the treatment effect on individuals who were initially affected by

the minimum wage introduction, while the policy relevant question concerns the effect on contemporaneously affected individuals but not on those who have been but may not be relevant to the minimum wage legislation anymore.

The larger individual mobility along the wage distribution, the larger the difference between contemporaneously affected individuals and initially affected individuals. Hence, if there are wage dynamics that are not caused by the minimum wage itself, some individuals might have experienced wage growth even in the absence of the minimum wage while others might have faced real wage decreases. Those wage changes create a dynamic selection in and out of the treatment group that is not captured by the initial measured bite.

If we need a contemporary minimum wage bite, why can we not measure the bite at each point in time? In principle, one could characterize individuals who receive exactly the minimum wage as affected by the minimum wage. However, measuring individuals that receive exactly the minimum wage to define the bite has some specific problems, as it neglects wages that are rounded or jobs that pay a fix premium on top of the minimum wage. Moreover, it raises the question how to deal with inaccurate measurement since exact measurement of hourly wages is difficult (Bossler & Westermeier, 2020). Finally, it is not clear how to deal with non-compliance, i.e. workers that should receive the minimum wage but are paid below. These workers could either be included in the bite because they are paid below minimum wage, or alternatively, they could be excluded from the bite because they are likely to remain below minimum wage.

Instead of calculating the share of workers who are remunerated below or exactly at the minimum wage, we propose to predict the contemporary bite using a prediction model. Such a model can be useful for minimum wage evaluations if it outperforms simple forward updating of the initial bite. So far, only few studies use some heuristics to predict the bite of the minimum minimum wage, mostly because no accurate information on hourly wages is available. E.g., Riley and Bondibene (2017) use firm-level average labor costs of firms in the UK – cross-validated with a small sample of workers for which they observe accurate wage information – to predict whether firms are affected by the national minimum wage introduction. Möller and König (2008) predict the individual-level bite for workers of the German main construction sector also because information on hourly wages is missing. They use a probit model to infer the bite from observable characteristics in the administrative employment data. In line with Möller and König (2008), we also use a prediction model for the bite of a minimum wage.

Our goal is however the prediction of the contemporary bite in future periods while Möller and König (2008) aim to estimate the bite at the time of the minimum wage introduction due to the lack of information on working hours.

3. Implementation of the LASSO estimator

A simple prediction model can be formulated as follows:

$$bite_{i,t+r}^{introduction} = \beta_0 + x'_{i,t}\beta_1 + z'_{i,t+r}\beta_2 + \epsilon_{i,t+r} \quad (1)$$

The *bite* of some region or individual i is predicted for some future period $t+r$, for which we want to evaluate a minimum wage which is introduced (or increased) right at (or right after) period t . This $bite_{i,t+r}$ is explained by lagged predictors $x_{i,t}$ and by contemporaneous predictors $z_{i,t+r}$, where β_1 and β_2 are the coefficients of these explanatory variables in the prediction model. Since it is not a priori obvious which variables influence if individuals or regions are affected by the minimum wage, we select the explanatory variables in our model in a purely data driven way based on their predictive power. For that, we use regularized regression to select predictor variables and estimate the coefficients.

The least absolute shrinkage and selection operator (LASSO) was first proposed by Tibshirani (1996) and belongs to a subclass of regularized regression methods. It works as a shrinkage estimator that basically consists of an OLS estimator and adds a penalty term for the absolute sum of coefficient values of all covariates. In contrast to the ridge estimator (Hoerl & Kennard, 1970), which decreases the size of certain coefficients to find an optimal solution for the minimization problem stated below, the LASSO is able to set certain coefficients to zero, thereby excluding them from the model. By doing so, the LASSO aims to select a sparse solution that includes covariates with high predictive performance and excludes others. Hence, LASSO can be used for model selection based on predictive power. The minimization problem of the LASSO estimator in a model setup with N observations and p explanatory variables can be formulated as:

$$\hat{\beta}^{lasso} = \min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (2)$$

While minimizing the first summand in the equation corresponds to the OLS estimator, the

second summand imposes a penalty term for the total sum of coefficient values in the model. The absolute size of each coefficient determines its contribution to the penalty term of the equation.¹ Hence, the LASSO-estimator aims to exclude large coefficients that add much noise in terms of reactivity to value changes of an explanatory variable to the estimator from the model.

The overall penalty level is determined by λ . Obviously, if the penalty level is set to zero the model equals a common OLS estimator. We apply two different approaches for the selection of the penalty parameter λ : First, we determine λ via 5-fold cross-validation and select the penalty level that minimizes the mean squared prediction error (MSE) over all folds. We call this method CV-LASSO in the following. Second, we apply the so-called rigorous LASSO (R-LASSO) proposed by (Belloni, Chen, Chernozhukov, & Hansen, 2012). This approach provides theory-based algorithms for the optimal penalization under heteroskedastic and non-Gaussian errors. It does not solely focus on the predictive performance but selects the penalty level to ensure consistent prediction and parameter estimation. The basic idea here is to control the noise (variance of the estimator around the true values) while keeping bias of the estimator as low as possible. For software implementation we use the Stata package *LASSOPACK* developed by (Ahrens, Hansen, & Schaffer, 2019).

While it has been shown that especially Random Forests can outperform parametric estimation methods and other machine learning methods, especially in the context of propensity score estimation (Pirracchio, Petersen, & van der Laan, 2015; Cannas & Arpino, 2019), we opt to use the LASSO for a few reasons: First, in contrast to studies that use a data generating process with only few covariates, we discuss a setting with potentially very large sets of explanatory variables. In such cases LASSO can outperform Random Forest estimation, as shown in Goller, Lechner, Moczall, and Wolff (2020) for the logistic estimation of propensity scores. Second, we observe a highly skewed distribution with respect to the share of individuals who are affected and who are unaffected by the minimum wage. Again, Goller et al. (2020) demonstrate that LASSO outperforms Random Forests in propensity score estimation with unequally distributed shares of observations in the treatment and control group.

¹To ensure that each coefficient is treated equally w.r.t. scaling of the respective x-variables, all x-variables are mean-standardized.

3.1. Data

Our main data source for the empirical analysis is a 2-percent sample of the administrative employment histories (“Beschäftigten-Historik”, BeH) of the Institute for Employment Research (IAB). The BeH is a rich administrative dataset which contains information on each employee in Germany recorded from compulsory employer reports to the German social security system. The sample is restricted to the years 2010-2014, which is our analysis period to predict and validate the bite of the German minimum wage. Among other variables, the dataset contains information on gross daily and monthly wages of full-time, part-time, and marginal employees. However, it does not include individuals’ working hours which are essential for the determination of hourly wages, which in turn determine the bite of the minimum wage. To calculate hourly wages, we merge data on working hours from the German Statutory Accident Insurance to the BeH observations of each individual. Due to legislative reasons, these data on working hours are only available for the years 2010-2014, as employers had to report working hours of their employees during this period. We therefore restrict our analysis period to the years 2010-2014. Moreover, the German minimum wage was introduced in 2015 such that minimum wage bite can only be determined in the years before the law came into force. Given the data on wages and working hours, we calculate an hourly wage for each employment-spell in the dataset.

All estimations in this paper are presented on the individual level but also on the regional level. For the latter, we aggregate the individual-level data over 50 labor market regions which are categorized according to commuting flows as proposed by Kropp and Schwengler (2011). We differentiate between two bite variables for our estimations. The first captures whether an individual is affected by the minimum wage from a binary categorization (incidence bite). On the individual level, this incidence bite takes the value 1 if we observe an hourly wage below the initial German minimum wage level of €8.50 per working hour and 0 otherwise. On the regional level, the incidence bite represents the share of individuals who are affected by the minimum wage in each of the 50 labor market regions. Hence, the regional incidence bite is a continuous measure on the interval $[0,1]$. The second bite variable is the bite gap between an individual’s wage and the minimum wage level for all individuals affected by the minimum wage. Hence, the bite gap ranges on the interval $[0,8.50]$ with a value of 8.50 for a person earning a zero wage and a value of 0 for all unaffected individuals. On the regional level, we aggregate the individual bite gaps such that the regional measure represents the average bite

gap of employees in the respective region.

Table 1: Descriptive statistics of bite variables

Individual Level		overall	2010	2011	2012	2013	2014
Incidence Bite	Mean	0.13	0.15	0.14	0.13	0.12	0.11
	Std.Dev	0.3372	0.3570	0.1198	0.3374	0.3269	0.3168
	Min	0.00	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00
Bite Gap	Mean	0.33	0.40	0.35	0.32	0.30	0.29
	Std.Dev	1.1466	1.2528	1.1820	1.1218	1.0970	1.0720
	Min	0.00	0.00	0.00	0.00	0.00	0.00
	Max	8.49	8.49	8.48	8.42	8.49	8.46
Observations		2,462,685	479,797	487,991	494,307	498,252	502,338
Regional Level		overall	2010	2011	2012	2013	2014
Incidence Bite	Mean	0.14	0.16	0.15	0.14	0.13	0.12
	Std.Dev	0.0437	0.0471	0.0458	0.0416	0.0388	0.0346
	Min	0.0761	0.0984	0.0910	0.0856	0.0791	0.0761
	Max	0.2940	0.2940	0.2860	0.2585	0.2337	0.2297
Bite Gap	Mean	0.35	0.43	0.37	0.34	0.32	0.30
	Std.Dev	0.1003	0.1138	0.1016	0.0885	0.0742	0.0702
	Min	0.19	0.24	0.24	0.20	0.19	0.20
	Max	0.76	0.76	0.69	0.59	0.51	0.53
Observations		250	50	50	50	50	50

Source: BeH, 2010-2014, analysis sample

Table 1 presents summary statistics for both bite variables. It shows means, standard deviations and extrema as well as the number of observations in our sample for each of the years. The upper part of the table corresponds to our estimation sample on the individual level, while the lower part of the table shows the summary statistics on the regional level. The full sample of all five years comprises about 2.46 million observations for which we can calculate an hourly wage. Hence, the data cover about 500,000 observations each year. The numbers slightly increase over time reflecting increasing employment numbers in Germany over the observation period. By contrast, we observe slightly decreasing bite levels over the years with an overall average of 0.13 for the incidence bite and 0.33 for the bite gap, respectively. Hence, we characterize about 13 percent of the individuals in the sample as affected by the minimum wage introduction and the average bite gap including those who are unaffected is about €0.33 on average. These observations on the individual level are aggregated over the 50 labor market regions, which leaves us with 50 observations per year and 250 observations in total (lower part of Table 1). The size of the incidence bite on the regional level is qualitatively similar compared to the individual level. However, the average values on the regional level are about 1 percent-

age point higher in each of the years, implying that individuals are not randomly distributed across regions. Obviously, we observe much higher variance in the data on the individual level as opposed to the regional level, since for the latter we use aggregated data.

3.2. The set of explanatory variables

The purpose of applying LASSO estimation in our case is the selection of the most powerful explanatory variables in terms of prediction performance of the minimum wage bite. In contrast to traditional regression models, the simple inclusion of all (available) explanatory variables does not lead to multicollinearity problems since the LASSO algorithm selects only the most powerful predictors and excludes explanatory variables that provide only a small contribution to the fit of the model. Although we do not have to deal with multicollinearity problems in the LASSO variable selection framework, there are some restrictions on the set of suitable variables. It mainly concerns problems of endogeneity since predictor variables might be affected by the minimum wage introduction, themselves. Hence, contemporaneously measured values of those variables cannot be used to predict minimum wage affectedness. Essentially, three types of variables seem to be suitable for our prediction problem. An overview of the sets of explanatory variables included in the analyses on the individual level and on the regional level is presented in Appendix Tables B.1 and B.2. The list includes:

- a contemporarily measured variables that are exogenous as these are not endogenously affected by the minimum wage,
- b lagged variables that are strictly exogenous (e.g. not correlated with the error term in any of the periods),
- c lagged variables measured prior to the minimum wage introduction, where the lagged information is exogenous while the contemporary information may be endogenous.

3.3. Estimation strategy

In most of our specifications, the full set of available variables is included allowing the LASSO algorithm to select a subset of the explanatory variables based on their predictive power. In alternative specifications, however, we repeat the prediction process for each of the specifications presented in section 4 omitting wage-based variables. There is no doubt that wage

variables contribute much to the bite prediction since the minimum wage bite is in fact derived from wages. However, it is also obvious that wages are themselves influenced by the minimum wage, thus, all variables derived from wages belong to class (c) type of variables. Hence, we can only include pre-treatment wages due to endogeneity concerns. We expect the pre-treatment wage to exert decreasing predictive power over time when predicting the contemporaneous bite. This is because there may be additional (unexplained) wage variation over longer time spans. With long term bite prediction in mind, it would be a desirable feature of our model to minimize the usage of class (c) variables as these may have a decreasing predictive power in the long run.

Given the outcome variable and a set of explanatory variables, we train and validate our model in a base year while the other years are used for testing external validity of the model. We choose the first year of our dataset (2010) as base year in our baseline specifications and use 2011 in robustness checks. For the baseline specification, this leaves the remaining four years (2011-2014) of our dataset as the test sample, which is used to assess the prediction performance of our model. The model selection process is conducted in the base year and includes all observations in the data for that year (training sample). The prediction (generalization) performance of the selected model is then tested on all observations in the remaining years in our dataset (test sample). Using this procedure, we can ensure both internal validity in the base year and external validity for the years following the base year, for which we also test the prediction accuracy. Hence, we train our model in the base year with only one observation per individual or region and test its performance in the following years, using different observations for each individual or region.

To assess the performance of our estimator, we use the mean squared error (MSE) as a measure for the predictive performance. As this measure describes the average squared deviation of predicted and observed values, it is an intuitive performance measure and easy to interpret. To establish better comparability of the MSE values between different models, we take the root and normalize the Root Mean Squared Error (RMSE) by dividing the RMSE values by the standard deviation of the bite variable in the data. This results in the Normalized Root Mean Squared Error (NRMSE), which is our primary measure for prediction performance. We use this performance measure and the following decomposition to evaluate and compare our LASSO-based predictions in the next section.

Besides the assessment of the predictive performance of our LASSO-selected models, we

would like to gain additional insights about the nature of the prediction error. Thus, we conduct a bias-variance decomposition of the MSE into two parts: the squared bias and the variance. This decomposition formula is formulated in equation 3. Note that this exercise yields a decomposition of the predicted bites relative to the true bites, which we observe in the testing period.

$$\begin{aligned}
MSE &= E[(bite_i^{true} - \widehat{bite}_i^{LASSO})^2] \\
&= Var(bite_i^{true} - \widehat{bite}_i^{LASSO}) + (E[bite_i^{true}] - E[\widehat{bite}_i^{LASSO}])^2 \\
&= (variance) + (bias)^2
\end{aligned} \tag{3}$$

Equation 3 shows that the MSE can be separated in the (squared) bias plus the variance. The first part describes to what extent the mean of the LASSO-predictions differs from the true mean of the observed bites, which is the unexplained systematic error, while the latter part describes the unsystematic random error of the LASSO-predicted bites compared with the observed bites.

4. Results of the bite prediction

The result section structures as follows: Section 4.1 focuses on regional level bite predictions, section 4.2 analyzes bite predictions on the individual level, and section 4.3 presents some additional results. In all sections, we apply the LASSO-based prediction model using 2010 as the base year to train our model and predict the incidence bite and bite gap in the prediction sample which includes the remaining years in the sample, 2011-2014.

4.1. Prediction of the regional bite

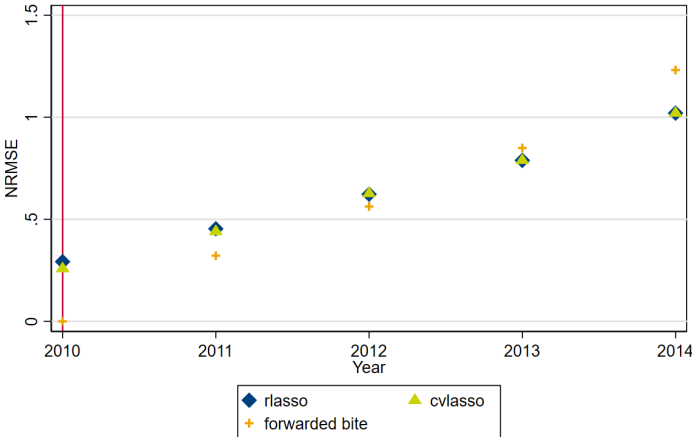
In the base year 2010, we run our LASSO regressions to predict the bite for all regions of 2010 but also to predict the regional bite of the prediction period 2011-2014.² The results from CV-LASSO- and R-LASSO-selected prediction models are illustrated in Figure 1. The graph illus-

²For the LASSO regressions, we use the full sample of 2010. In a robustness check we also conducted the LASSO regressions on a reduced sample of 2010. While this may yield an increased error in the prediction-period (2011-2014) due to a loss of information for the LASSO regressions, it allows us to check the external validity of our predictions already in the base year. For this purpose we tried several sample splits of the sample in the base year, where a 75 percent sample for the estimation and 25 percent sample for the validation maximizes external validity in the base year. While this exercise may be interesting in itself, it is not our main purpose to predict the bite for the base year, but for subsequent years.

trates the NRMSE for our prediction models and - for comparability - of the forwarded bite. The NRMSEs allow us to compare the prediction-errors of LASSO-based predictions to the errors of the forwarded bite. Obviously, the prediction-error of the forwarded bite is zero in the base year as the bite is measured in this year. The NRMSE of the forwarded bite then increases approximately linearly in the subsequent prediction years. The prediction-error of the LASSO-models also shows increasing values that grow with the distance of the prediction year to the base year. However, the slope of the NRMSE curve is lower than the slope of the error-curve for the forwarded bite. After two prediction years, the LASSO-model shows an approximately equal NRMSE and outperforms bite-forwarding in the years thereafter (2013 and 2014). The results show that bite-forwarding is appropriate to approximate the actual bite in the short run but the error exceeds the predicted bites' errors after only three years. This indicates the necessity to account for dynamic changes of the minimum wage bite, which our prediction model is able to do.

Both LASSO-methods show a relatively similar performance. It is remarkable, however, that the R-LASSO produces slightly lower NRMSEs than the CV-LASSO in the long run while the relationship is contrary in the years 2010 and 2011. This may be due to a focus on external validity in the penalization process of the R-LASSO, while the CV-LASSO solely focuses on minimizing the error in the training period.

Figure 1: NRMSE of regional incidence bite prediction, LASSO in comparison with forwarding

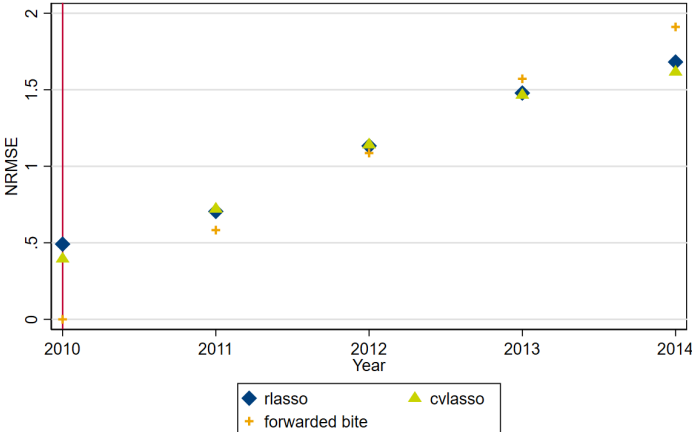


Notes: Normalized root mean squared error (NRMSE) for estimations based on R-LASSO and CV-LASSO compared with bite-forwarding. Base year 2010, prediction years 2011-2014.

We also apply our prediction model to the regional average bite gap. The results are presented in the same fashion as for the incidence bite and are illustrated in Figure 2. The results are qualitatively similar to those presented above. However, the error levels are higher com-

pared to the incidence bite, which is due to a higher variance in the bite gap variable compared to the incidence bite variable. Compared with bite-forwarding we find similar NRMSEs in the second prediction year (2012) and the LASSO-based predictions outperform bite-forwarding in the consecutive years. Divergent to the results from the incidence bite prediction, the NRMSE-values of the CV-LASSO are below the values of the R-LASSO. Although, the differences are rather small.

Figure 2: NRMSE of regional bite gap prediction, LASSO in comparison with forwarding



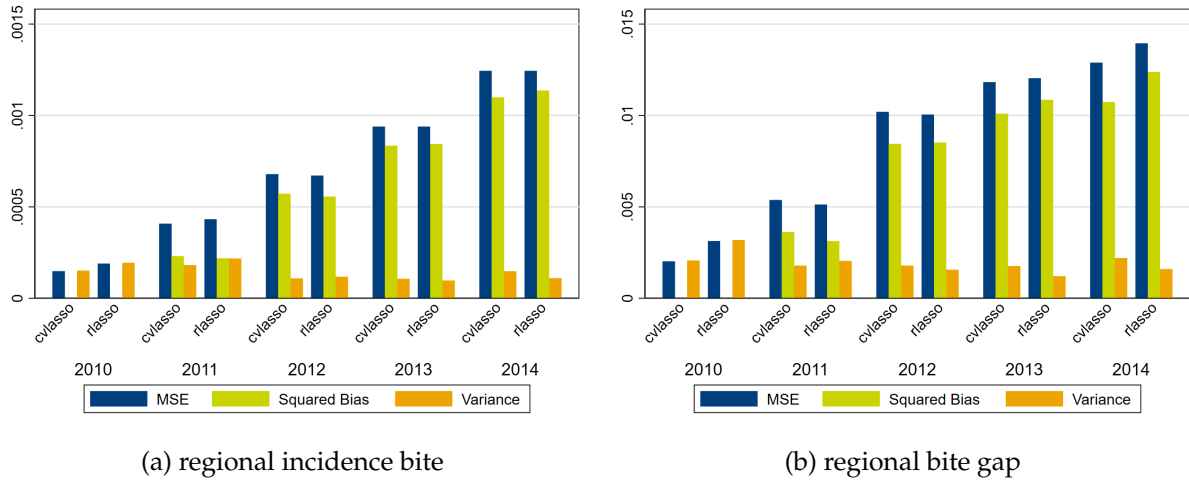
Notes: Normalized root mean squared error (NRMSE) for estimations based on R-LASSO and CV-LASSO compared with bite-forwarding. Base year 2010, prediction years 2011-2014.

The mean squared error can be decomposed into bias and variance as described in section 3.3. Using this decomposition, we can reveal details of the prediction performance and thereby further illuminate the differences between the R-LASSO and the CV-LASSO. In general, it is desirable to achieve low values for both prediction variance and bias, where the first captures the unsystematic error of the bite prediction and the latter describes systematic differences.

Figure 3 illustrates the decomposition for the incidence bite (panel (a)) and the bite gap (panel (b)) separated for the base year and each of the prediction years. In both cases the prediction bias is negligible in the base year, in which we estimate the LASSO regressions. Hence, there is no systematic bias of the LASSO for the estimation year, but of course it leaves some unexplained variation. However, in the prediction years 2011-2014, which we are most interested in most of the mean squared error is due to an increasing systematic difference between the LASSO-predicted bites and the observed bites. However, the random variation that remains unexplained remains fairly constant over time. We observe a similar bias and variance for both LASSO prediction methods. While in the base year, the CV-LASSO seems to be slightly better in terms of explaining additional variance in the bite, both methods are very similar in all the

subsequent prediction years.

Figure 3: Decomposition of the mean squared error of the LASSO-based bite prediction for the years 2010-2014.



Notes: Mean squared error decomposition according to equation 3 for LASSO-based bite predictions on the regional level. Panel (a) shows the decomposition for incidence bite predictions, panel (b) shows the decomposition for bite gap predictions. Base year 2010, prediction years 2011-2014.

The bite prediction on the regional level could possibly have some disadvantages compared to prediction on the individual level. First, the number of observations is quite limited (50 labor market regions) and the variation of many explanatory variables is rather small between regions. In the next section, we use data on the individual level and apply similar variable selection models that allow us to include predictors on the individual level.

4.2. Prediction of the individual bite

4.2.1. Incidence bite prediction

Other than the previous analyses on the regional level that represented a linear regression problem (due to continuous bite variables), the individual level bite prediction of the incidence bite is a classification problem. We use a logit version of the LASSO estimator, which maximizes the penalized log-likelihood function:³

$$L = \frac{1}{N_{training}} \sum_{i=1}^{N_{training}} \{bite_i(\beta_0 + x'_i\beta) - \log(1 + e^{(\beta_0 + x'_i\beta)})\} - \frac{\lambda}{N_{training}} |\beta| \quad (4)$$

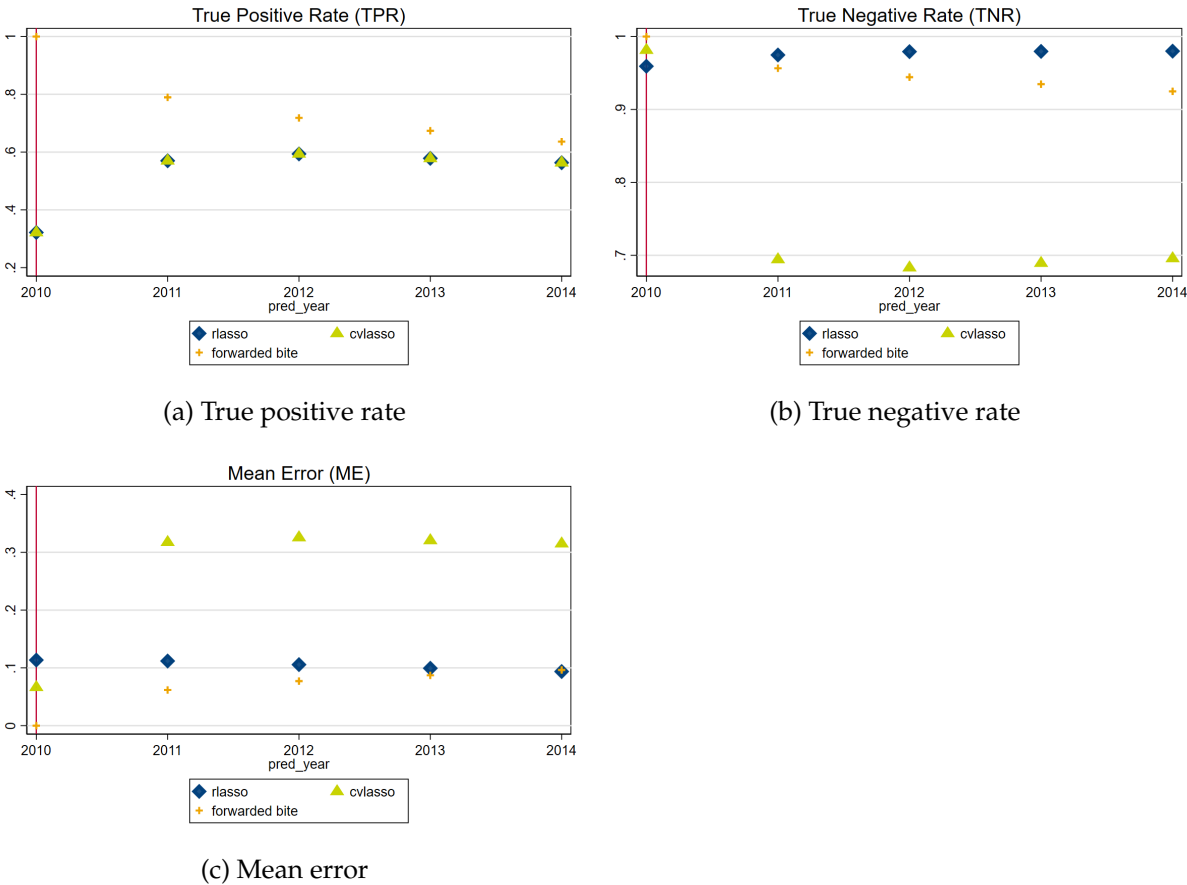
The prediction of the LASSO-logit $\hat{\Lambda}(\beta_0 + x'_i\beta)$ yields a probability to be treated for each individual in the data. However, we ideally want to classify individuals binary either into the

³Deviating from the analysis on the regional level, we use a sample split of 75 percent to receive a training sample to run the R-LASSO and the CV-LASSO.

treatment or the control group. This classification of individuals should ideally yield a large fraction of truly treated individuals which are classified to be treated, so called true positive rate (TPR), and a large fraction of individuals which is actually untreated who are classified in the control group, so called true negative rate (TNR). To classify individuals as treated and non-treated, we apply two different rules to choose a cutoff along the probability distribution ($\hat{\Lambda}$). These cutoff rules determine the probability threshold for each individual to be classified as treated or non-treated. The rules to set the classification cutoff are:

1. Minimization of the mean error (ME), which is the sum of the true positive rate and the true negative rate, i.e. $ME = TPR + TNR$.
2. Setting the error rate (ER) to unity, where the ER is the ratio of TPR and TNR, i.e. $ER = \frac{TPR}{TNR}$. The rationale here is to achieve balance between both classification rates.

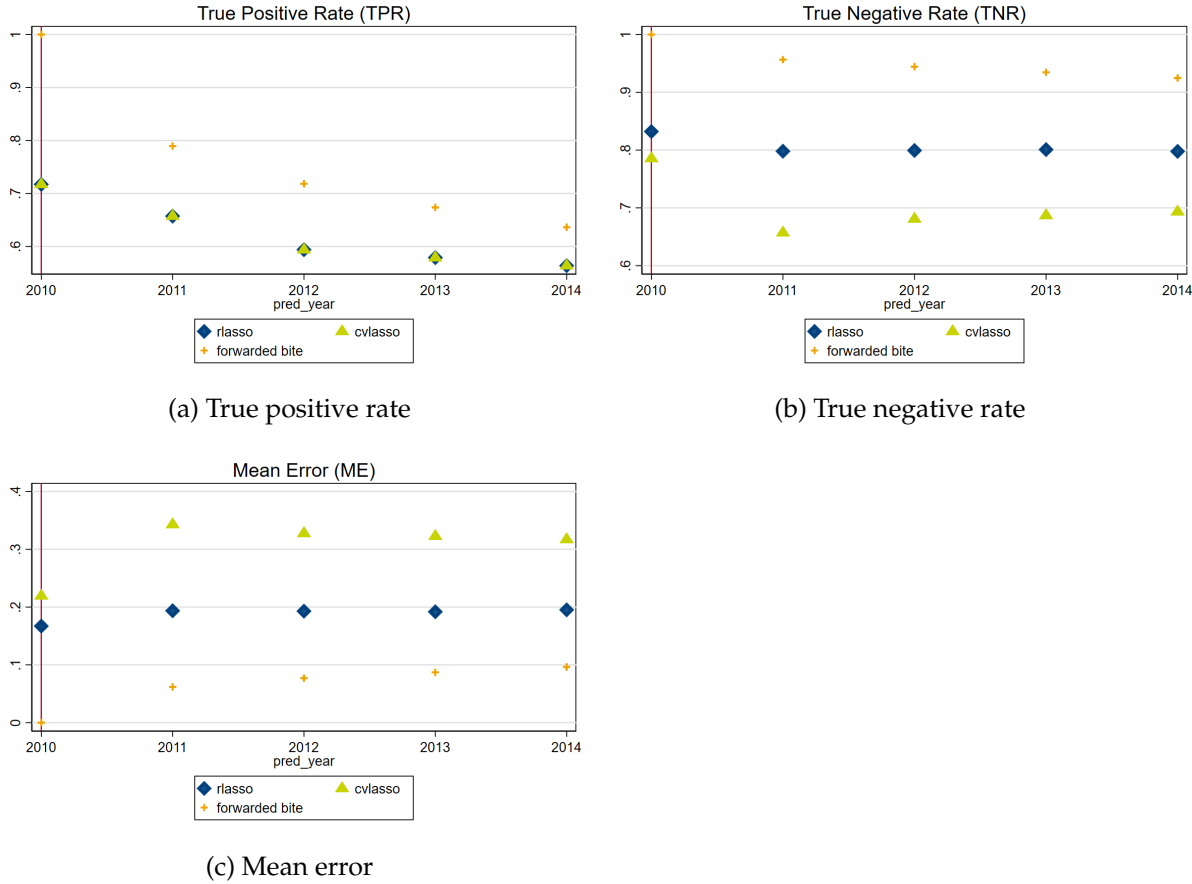
Figure 4: Prediction performance of the Logit-LASSO using the $\min(ME)$ cutoff rule.



Notes: Error measures for estimations based on LASSO-logit predictions and bite-forwarding. The training sample is a random 75 percent sample of the observations in 2010. The validation samples comprise 25 percent of the observations in 2010. The test sample comprises all observations in the years 2011-2014.

Figure 4 shows the prediction performance, when we use the minimization of ME as the cut-off rule. Looking at the true positive rate in Figure 4a, the fraction of true positives dramatically falls when using simple forwarding of the individual bite classification. However, the LASSO-logit prediction is even worse, showing a lower rate of true positives in all LASSO estimation methods. By contrast, Figure 4b shows that the prediction of true negatives is fairly good at least for the R-LASSO, as the LASSO-logit outperforms simple forwarding from $t + 2$ onwards. In combination, the mean error of the LASSO-logit prediction equals simple forwarding in $t + 4$, as shown in Figure 4c, and is likely to decrease below the error of simple forwarding in periods thereafter.

Figure 5: Prediction performance of the Logit-LASSO when using the $ER = 1$ cutoff rule.



Notes: Error measures for estimations based on Logit-LASSO predictions and bite-forwarding. The training sample is a random 75 percent sample of the observations in 2010. The validation samples comprise 25 percent of the observations in 2010. The test sample comprises all observations in the years 2011-2014.

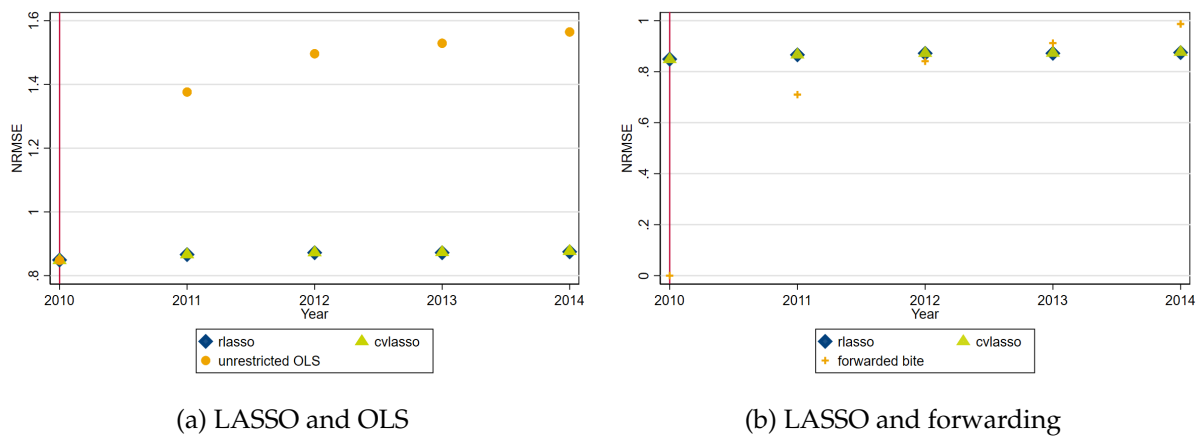
Figure 5 shows the prediction performance, when using $ER = 1$ to set the cutoff in the treatment classification. The results show lower TNR rates while the TPR rates remain similar to the results above. In sum, the mean error rates of the LASSO-predictions are at even higher

levels than in the specification described above and our predictions are not able to beat the forwarded bite over the years in our sample.

4.2.2. Continuous bite gap prediction

In light of the aforementioned difficulties with respect to binary bite prediction in a logistic model, we predict the bite gap as a continuous measure, which allows us to apply the linear (OLS-like) version of the LASSO estimator. This avoids the additional step of cut-off selection in the logistic version. The set of explanatory variables remains unchanged from the previous analyses.

Figure 6: NRMSE of individual bite gap prediction, LASSO, OLS and forwarding



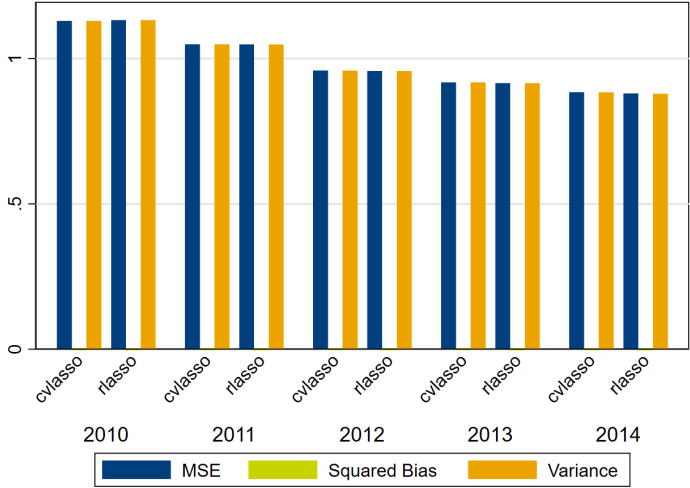
Notes: Normalized root mean squared error (NRMSE) for estimations based on R-LASSO and CV-LASSO as well as unrestricted OLS estimation and bite-forwarding. Base year 2010, prediction years 2011-2014.

As in section 4.1, we compare the NRMSEs of our LASSO-based predictions with the forwarded bite gap. But in addition, the large sample size on the individual level also allows us to compare the LASSO-predictions with an unrestricted OLS regression. Just like in chapter 4.1, we use the observations in the base year for model training and the remaining four years for predictions. Figure 6 shows that the LASSO-predictions by far outperform the predictions from an unrestricted OLS-model (panel (a)). Comparing the LASSO-predictions with simple forwarding in panel (b), the NRMSE is equal in the second year after the training period but the LASSO-predictions outperform bite forwarding from the third prediction year (2013) onwards. In comparison to the regional level predictions (see figure 2) it is noticeable that there is barely any difference in the NRMSE across prediction years. Moreover, the predictions on the individual level are more accurate since we observe lower NRMSE-values in each of the prediction

years. Hence, the results indicate that bite prediction on the individual level is particularly promising using our approach.

To get a more detailed picture of our LASSO-selected model, we again evaluate the bias-variance trade-off. The decomposition results are presented in Figure 7 for the two LASSO estimation methods. It is interesting to see that the squared bias is negligible as the MSE is entirely explained by the variance. This result indicates that there is lots of variance in the individual-level microdata which is not explained by the LASSO regressions but not meaningful bias. This is very different from the regional level, where most of the MSE is due to a systematic difference between the predictions and the observed bites (see Figure 3).

Figure 7: Decomposition of bias, variance and residual variation of the LASSO-based bite prediction for the years 2010-2014.



Notes: Mean squared error decomposition according to equation 3 for LASSO-based bite gap predictions on the individual level. Base year 2010, prediction years 2011-2014.

4.3. Additional results

Additional results are presented in this section. In a first step, we abstain from wage variables, which should only be used with a time-lag in the set of potential explanatory variables. In a second step, we use 2011 as an alternative base year to check whether results are sensitive to the year of training.

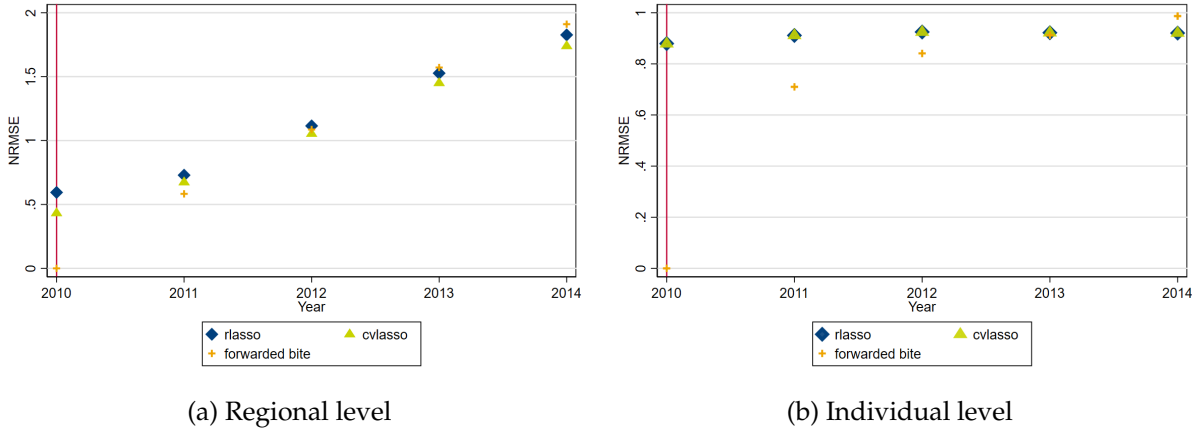
4.3.1. Omitting wage variables

We assess the LASSO-based bite prediction when abstaining from initially measured wage variables. The idea is to use predictors that are observed contemporarily but are themselves not

endogenously affected by the minimum wage. Wage variables might be good predictors of the bite since the bite is directly defined from wages. At the same time, wages are highly endogenous to minimum wages since they are directly targeted by the legislation. Hence, we use only lagged wage information in our main specifications. While this procedure ensures that wages are exogenous to the minimum wage treatment, these variables may lose prediction power over time. Hence, we also test a prediction model that omits wage information and uses contemporaneous macro variables only.

Figure 8 presents the respective results for the base year and the prediction years. The left panel (a) of Figure 8 shows the NRMSE for the regional-level bite and the right panel (b) shows the NRMSE for the individual-level bite. In both cases we restrict the analysis to the bite gap, as this bite measure appeared slightly superior in our main analyses.

Figure 8: NRMSE of bite gap predictions, no wage-variables, LASSO and bite-forwarding



Notes: Normalized root mean squared error (NRMSE) for estimations based on R-LASSO and CV-LASSO and bite-forwarding. Base year 2010, prediction years 2011-2014.

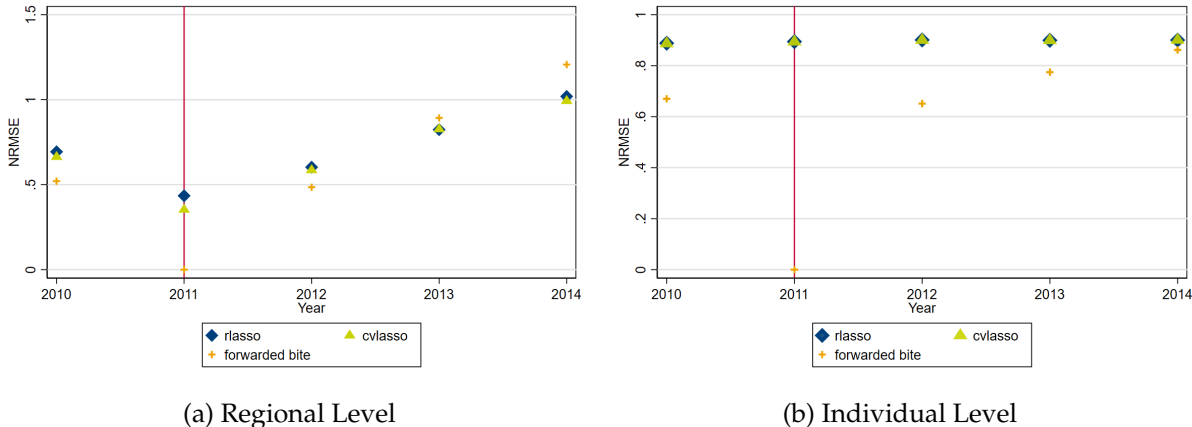
Compared to the main results, Figure 8 shows identical results w.r.t. the NRMSE of simple bite forwarding. While the error is zero in the initial period by definition, it strongly increases over time, leading to a large inaccuracy of the forwarded bite to represent the observed contemporaneous bite. The predictions from the LASSO show larger error rates in the short run and the results for the prediction years differ in the regional and the individual analysis. On the regional level, the LASSO predictions can hardly beat simple forwarding as the error rates increase with roughly the same slope over the prediction years (panel (a)). On the individual level, however, we observe very little variance in the error rates over the prediction years such that the predictions outperform simple bite-forwarding from the year 2013 onwards. (panel (b)). However, the improvement (compared to bite-forwarding) is slightly lower than in the

main analysis, demonstrating that wage variables result in a significant improvement of the prediction. Hence, we recommend to use wage variables in the bite prediction, albeit losing predictive power in the long run, simply because they are measured with an increasing time lag. This seems to be especially important on the regional level since the prediction performance is much lower than in section 4.1.

4.3.2. Using an alternative base year 2011

Next, we change our initial specification with respect to the base year of the prediction exercise. We use 2011 as the base year instead of 2010 and again assess the prediction performance in the test years 2010 and 2012-2014. This iteration of the base year serves as a robustness check to test whether similar patterns are observed from a different base year. The test allows us to check whether the base year is specific in its wage structure or whether there are specific changes between the base and test years that may affect the prediction performance.⁴ At the same time, this change of the base year limits the number of subsequent validation years to check for the prediction performance.

Figure 9: NRMSE of bite gap predictions, base year 2011, LASSO and bite-forwarding



Notes: Normalized root mean squared error (NRMSE) for estimations based on R-LASSO and CV-LASSO and bite-forwarding. Base year 2010, prediction years 2011-2014.

The results in Figure 9 show similar patterns as before, again showing results for the bite gap on the regional level (panel (a)) and on the individual level (panel (b)). The NRMSE of the forwarded bite is again zero in the base year, which is 2011, but it strongly increases in the years thereafter. Again, the LASSO-based prediction improves relative to bite forwarding over

⁴We are aware of a structural breaks of some explanatory variables, including the occupational classifier and the part-time information in the administrative data between 2010 and 2011. Hence, it is possible that the prediction performance in our test sample is negatively affected by those changes (Fitzenberger & Seidlitz, 2020).

time: In the prediction on the regional level, it outperforms simple bite-forwarding already in the second period. In the individual-bite prediction, where there is more variation, it improves on the forwarded bite not before the third prediction year. But again, it is likely to improve the longer run over simple bite-forwarding.

5. Application: Wage effects of the German minimum wage

introduction

While the previous sections solely focused on testing the LASSO-based bite prediction using the years 2010 to 2014 for which our sample includes data on working hours, we extend the time frame in this section and forecast minimum wage bites for the years after 2014, based on our LASSO-selected prediction models.⁵ This allows us to analyze the 2015 minimum wage introduction in Germany.

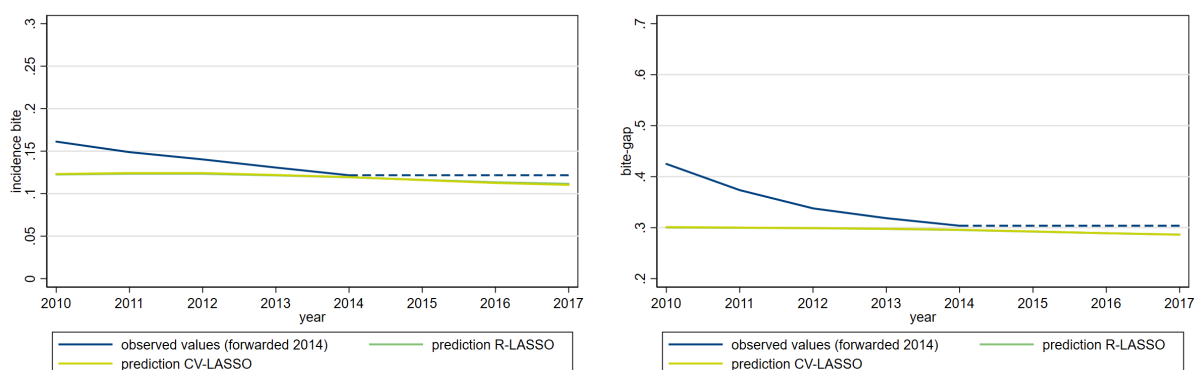
First, we descriptively compare our bite predictions and forecasts with the forwarded values of the bite for the post-2014 period. In contrast to the previous section, we now focus on the development of the mean bite rather than the prediction-error. This perspective enables us to assess the ability of our models to capture changes in the mean over time. We argue that bite prediction is particularly superior to simple forwarding in times of changing economic circumstances. I.e., our models is able to capture changes in the bite of the minimum wage.

Second, we analyze the impact of the minimum wage on wages, in treatment effect regressions of log wages on the predicted bite. We thereby compare the results of conventional difference-in-differences specifications using forwarded bites with regressions that use the LASSO-based predicted bites. As before, we separately assess predictions on the regional level and on the individual level. However, the models are now trained based on the observations in 2014 since we do not desire a testing period and instead want to train the model based on the most recent information and forecast the bite from 2014 onwards.

Figure 10 shows the observed values for the incidence bite (panel (a)) and the bite gap (panel (b)) on the regional level, as well as the respective predicted (and forecasted) values from both LASSO penalization schemes (R-LASSO and CV-LASSO). The observed values are forwarded from 2014 onwards. It shows that the predicted bite follows a slight downward trend after

⁵We use the term "prediction" to indicate predicted bites within the sample period (2010-2014) that can be tested by comparison with observed bites from the data, whereas we use the term "forecast" to indicate predicted bites outside the sample period which are the years after 2014 that cannot be compared to observed bites, and thus, cannot be validated.

Figure 10: Average bite forecasts - regional level, base year 2014



(a) incidence bite

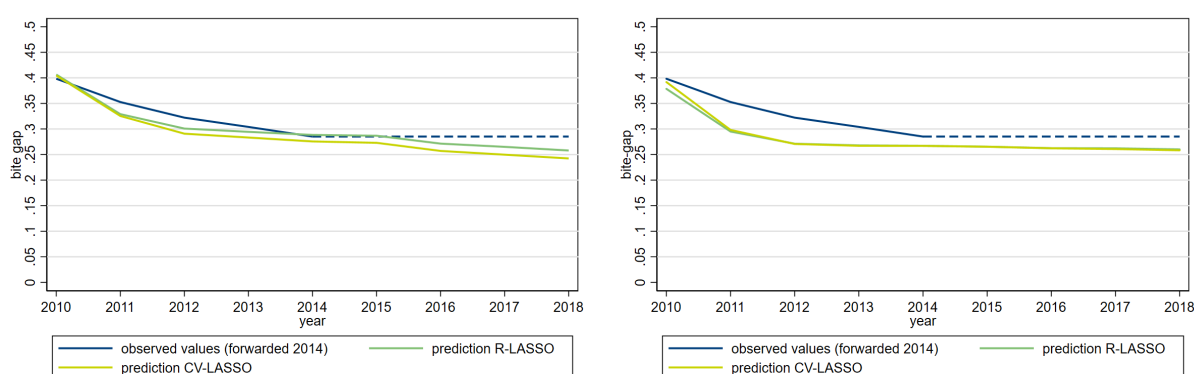
(b) bite gap

Notes: Forecasts of the mean values (trends) for the incidence bite (a) and the bite gap (b) from R-LASSO and CV-LASSO-based predictions on the regional level. base year for the predictions: 2014. The blue line shows the measured bite values from the BeH dataset which are forwarded for 2015-2017 with the value from 2014.

2014, while the forwarded bite remains constant by definition.⁶

The bite forecasts on the individual level show a different picture (Figure 11). We focus on prediction of the bite gap as outcome and illustrate the results of LASSO-based forecasts including wage variables (panel (a)) and excluding wage variables (panel (b)). The forecasted bite-gaps follow the observed trend quite well. This is not only true for the specifications that include wage-variables but also for estimations without wage information, as shown in panel (b).

Figure 11: Average bite forecasts - individual level, base year 2014



(a) with wages

(b) without wages

Notes: Forecasts of the mean values (trends) for the bite gap including wages as explanatory variables (a) and excluding wages (b) from R-LASSO and CV-LASSO-based predictions on the individual level. base year for the predictions: 2014. The blue line shows the measured bite values from the BeH dataset which are forwarded for 2015-2017 with the value from 2014.

⁶For forecasts based on 2010 as training year, see Appendix Figure C.1.

In order to apply the bite-forecasts in a causal analysis, we estimate a standard Difference-in-Differences model. This allows us to quantify the effect of the predicted bite. Moreover, we can compare this effect with conventional treatment effects that use the forwarded bite. We estimate the following model for the post-treatment years 2015-2017 separately, where 2014 serves as the reference-year of the Difference-in-Differences specification:

$$\ln(wage)_{it} = year_t * \gamma + bite_{it} * \theta + year_t * bite_{it} * \delta + \epsilon_{it} \quad (5)$$

The dependent variable of interest is the log monthly wage. $year_t$ is a dummy for the respective post-treatment year and $bite_{it}$ is either the forwarded bite or the predicted bite. The interaction of the post treatment dummy and the bite captures the treatment effect of interest δ .

Tables 2 and 3 present the estimated treatment effects for the estimation on the regional level, where the first uses the incidence bite and the latter uses the bite gap. The first column shows effects of the conventional bite forwarding, i.e. effects on regions that were affected by the minimum wage in 2014. The wage effect of the regional bite gap is 13 log-points in 2015 and 14 log-points in 2016. In 2017, the effect increases to 20 log-points. In columns (2) and (3) we use the predicted bite gap of the respective post-treatment years. When using the R-LASSO-based bite in column (2) the wage effect increases in the first post treatment year. But the effect decreases in the years thereafter. When using the CV-LASSO-based bite (column 3), the effects are slightly below these numbers, but again, in subsequent treatment years the treatment effect slightly decreases.

The result of an over time decreasing treatment effect is in line with our expectations. Initially affected regions experience an upward trend in wages, e.g. due to a convergence of regions or mean-reversion, which leads to an upwards bias of the effect in subsequent years which is included in the conventional estimates in column (1). By contrast, the contemporary affected regions, which should be more closely identified by the LASSO-based bites, show a somewhat smaller treatment effect in years after the minimum wage introduction. This pattern holds when we estimate a wage effect of the bite-gap on the regional level (see Table 3).

Table 4 presents wage effects on the individual level using the bite gap. Column (1) shows the wage effect for the initially affected individuals from a forwarded bite. The wage effect increases over time: It is 0.06 log-points in 2015 and grows to 0.10 log-points in 2016 and 0.14 log-points in 2017. When using the predicted bites in columns (2) and (3), we observe a similar

Table 2: Minimum wage effects on wages - regional level - bite gap

	Forwarding		Prediction	
	(1)	R-LASSO	CV-LASSO	(3)
	Log real wage	Log real wage	Log real wage	Log real wage
2015 - 2014	0.134*** (3.74)	0.219** (3.35)	0.108* (2.39)	
2016 - 2014	0.141*** (3.54)	0.140 (1.45)	0.102 (1.44)	
2017 - 2014	0.205*** (5.28)	0.125 (1.30)	0.0857 (1.42)	
Observations	100	100	100	

Notes: The table shows treatment effects of the minimum wage on log wages from difference-in-differences estimations on the regional level. The independent variables are the forwarded and predicted bite gaps in the years 2015, 2016 and 2017, respectively. The average size of the predicted bite gaps is about €0.29 across the years 2015-2017, see Table A.1. Base year of the LASSO model: 2014. T-statistics are in parentheses. Asterisks indicate the statistical significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Minimum wage effects on wages - regional level - incidence bite

	Forwarding		Prediction	
	(1)	R-LASSO	CV-LASSO	(3)
	Log real wage	Log real wage	Log real wage	Log real wage
2015 - 2014	0.339*** (5.57)	0.372*** (3.95)	0.334*** (4.04)	
2016 - 2014	0.337*** (4.93)	0.263* (2.28)	0.248* (2.40)	
2017 - 2014	0.443*** (6.24)	0.288* (2.28)	0.276* (2.46)	
Observations	100	100	100	

Notes: The table shows treatment effects of the minimum wage on log wages from difference-in-differences estimations on the regional level. The independent variables are the forwarded and predicted incidence bites in the years 2015, 2016 and 2017, respectively. Base year of the LASSO model: 2014. T-statistics are in parentheses. Asterisks indicate the statistical significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

wage effect of 0.06 log-points in 2015 and a fairly constant wage effect of 0.07 to 0.09 in the two consecutive treatment years. Noting that the treatment effect does not over time increase by the same margin.

The results demonstrate that the conventional estimation from a forwarded bite is over-estimated, which reflects that initially low paid individuals experience a wage growth that would also emerge in the absence of a minimum wage introduction. This wage growth of low paid individuals is commonly known as mean-reversion. The wage effect of contemporary af-

Table 4: Minimum wage effects on wages - individual level

	Forwarding	Prediction	
	(1)	R-LASSO (2)	CV-LASSO (3)
	Log real wage	Log real wage	Log real wage
2015 - 2014	0.0629*** (36.08)	0.0673*** (18.48)	0.0649*** (18.61)
Observations	941,993	1,140,668	1,140,668
2016 - 2014	0.101*** (56.08)	0.0742*** (20.36)	0.0715*** (20.48)
Observations	918,806	1,145,560	1,145,560
2017 - 2014	0.142*** (77.91)	0.0922*** (25.28)	0.0889*** (25.44)
Observations	902,659	1,148,629	1,148,629

Notes: The table shows treatment effects of the minimum wage on log wages from difference-in-differences estimations on the individual level. The independent variables are the forwarded and predicted bite gaps in the years 2015, 2016 and 2017, respectively. The number of observations between forwarding and prediction differs because in the forwarding scheme only initially observed individuals are included while the prediction includes labor market entrants. Base year of the LASSO model: 2014. T-statistics in parentheses. Asterisks indicate the statistical significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

affected individuals seems much lower, i.e., individuals who would have been affected by the minimum wage in 2017 receive about 9 percent higher wages than they would have received without minimum wage.

Another discrepancy between columns (2) and (3) and the conventional estimation in column (1) is that the predicted bites include individuals that entered employment after the minimum wage was introduced, while bite forwarding only allows estimating an effect on individuals that were initially affected and employed, creating an increasingly selective sample over time. The results suggest that such stable jobs create a positive selection in terms of wages, which may again explain an overestimation of the true wage effect. By contrast, the group of contemporary affected individuals also includes labor market entrants, including young individuals, who tend to earn relatively low wages.

6. Conclusion

The purpose of this article is to provide an alternative approach to forward updating of the bite in empirical minimum wage evaluations. We predict the treatment bite of the German minimum wage introduction on the individual and regional level in the medium to long run. We

use the LASSO estimator for the selection of explanatory variables with high predictive power and fit the LASSO-selected model to data from before the minimum wage introduction 2010 to 2014. The set of explanatory variables contains microeconomic variables from administrative datasets as well as macroeconomic data from public data sources. Our LASSO-selected prediction models show improvements in terms of prediction error compared to bite forwarding at least in the long run (after 3-4 years) on the regional level as well as on the individual level. The results indicate that forwarding of the pre-defined minimum wage bite is appropriate to approximate the actual bite in the first years after the minimum wage introduction. In the long run, however, wage changes that occur independently of the minimum wage can lead to changes in the composition of affected and unaffected individuals/regions such that the error from bite forwarding increases over time.

Using the LASSO-based bite yields two improvements over conventional estimations: First, it addresses dynamic selection in and out of the treatment group. Second, it allows us to define a bite for labor market entrants, while in the conventional bite forwarding, the bite is only defined for initially employed individuals; again, leading to an increasingly selective treatment group over time.

Applying the LASSO-based bite to wage effects of the German minimum wage introduction confirms the literature in terms of significantly positive wage effects. It demonstrates, however, that simple forwarding of the initial bite over-estimates the wage effect from the second treatment year onwards, which can be due to mean-reversion or a convergence of regions. It demonstrates that the use of an initially defined individual bite yields an increasingly selective treatment group over time.

However, the presented approach also comes along with a number of shortcomings: First, good predictor variables in the sample period (2010-2014) may not be good predictors in the future after the minimum wage was introduced. Second, it is a challenging task to find good predictors of the bite, which are not themselves (endogenously) affected by the minimum wage. Hence, we have to make assumptions on the variables that are considered as exogenous predictors of the bite, especially if these are measured contemporaneously.

References

- Ahlfeldt, G. M., Roth, D., & Seidel, T. (2018). The regional effects of Germany's national minimum wage. *Economic Letters*, 172, 127–130.
- Ahrens, A., Hansen, C. B., & Schaffer, M. E. (2019). *LASSOPACK: Stata module for lasso, square-root lasso, elastic net, ridge, adaptive lasso estimation and cross-validation*. (Tech. Rep.).
- Belloni, A., Chen, D. L., Chernozhukov, V., & Hansen, C. B. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6), 2369–2429.
- Bossler, M., & Schank, T. (2020). Wage inequality in Germany after the minimum wage introduction. *IZA Discussion Paper No 13003*.
- Bossler, M., & Westermeier, C. (2020). Measurement error in minimum wage evaluations using survey data. *IAB-Discussion Paper 11/2020, Nuremberg*.
- Caliendo, M., Fedorets, A., Preuss, M., Schröder, C., & Wittbrodt, L. (2018). The short-run employment effects of the German minimum wage reform. *Labour Economics*, 53, 46–62.
- Cannas, M., & Arpino, B. (2019). A comparison of machine learning algorithms and covariate balance measures for propensity score matching and weighting. *Biometrical Journal*, 61(4), 1049–1072.
- Card, D. (1992). Using regional variation in wages to measure the effects of the federal minimum wage. *Industrial and Labor Relations Review*, 46(1), 22–37.
- Dolton, P., Bondibene, C. R., & Stops, M. (2015). Identifying the employment effect of invoking and changing the minimum wage: A spatial analysis of the uk. *Labour Economics*, 37, 54–76.
- Dolton, P., Bondibene, C. R., & Wadsworth, J. (2012). Employment, inequality and the uk national minimum wage over the medium-term. *Oxford Bulletin of Economics and Statistics*, 74(1), 78–106.
- Fitzenberger, B., & Seidlitz, A. (2020). The 2011 break in the part-time indicator and the evolution of wage inequality in Germany. *Journal for Labour Market Research*, 54(1), 1–14.
- Garloff, A. (2019). Did the German minimum wage reform influence (un)employment growth in 2015? Evidence from regional data. *German Economic Review*, 20(3), 356–381.

- Goller, D., Lechner, M., Moczall, A., & Wolff, J. (2020). Does the estimation of the propensity score by machine learning improve matching estimation? the case of germany's programmes for long term unemployed. *Labour Economics*, 101855.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 1(12), 55–67.
- Kropp, P., & Schwengler, B. (2011). Abgrenzung von arbeitsmarktregionen - ein methodenvorschlag. *Raumforschung und Raumordnung, H. 1*(Jg. 69), 45–62.
- Möller, J., & König, M. (2008). Mindestlohneffekte des entsendegesetzes? eine mikrodatenanalyse für die deutsche bauwirtschaft. *Zeitschrift für ArbeitsmarktForschung-Journal for Labour Market Research*, 41(2/3), 327–346.
- Pirracchio, R., Petersen, M. L., & van der Laan, M. (2015). Improving propensity score estimators' robustness to model misspecification using super learner. *American journal of epidemiology*, 181(2), 108–119.
- Riley, R., & Bondibene, C. R. (2017). Raising the standard: minimum wages and firm productivity. *Labour Economics*, 44, 27–50.
- Schmitz, S. (2019). The effects of Germany's statutory minimum wage on employment and welfare dependency. *German Economic Review*, 20(3), 293–329.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Appendix

A. Descriptive statistics of predicted bites

Table A.1: Descriptive statistics of bite forecasts (regional)

incidence bite		(2010)	(2011)	(2012)	(2013)	(2014)	(2015)	(2016)	(2017)
R-LASSO	Mean	0.12	0.12	0.12	0.12	0.12	0.12	0.11	0.11
	Std.Dev	0.0259	0.0254	0.0257	0.0259	0.0257	0.0249	0.0243	0.0233
	Min	0.09	0.09	0.09	0.09	0.09	0.08	0.08	0.08
	Max	0.19	0.19	0.20	0.19	0.19	0.19	0.18	0.18
CV-LASSO	Mean	0.12	0.12	0.12	0.12	0.12	0.12	0.11	0.11
	Std.Dev	0.0281	0.0274	0.0278	0.0280	0.0279	0.0269	0.0261	0.0249
	Min	0.09	0.09	0.09	0.09	0.08	0.08	0.08	0.08
	Max	0.20	0.20	0.20	0.20	0.20	0.19	0.18	0.18
bite gap		(2010)	(2011)	(2012)	(2013)	(2014)	(2015)	(2016)	(2017)
R-LASSO	Mean	0.30	0.30	0.30	0.30	0.30	0.29	0.29	0.29
	Std.Dev	0.0354	0.0351	0.0353	0.0347	0.0347	0.0341	0.0333	0.0326
	Min	0.25	0.25	0.25	0.25	0.25	0.24	0.24	0.24
	Max	0.40	0.40	0.40	0.40	0.40	0.39	0.38	0.38
CV-LASSO	Mean	0.30	0.30	0.30	0.30	0.30	0.29	0.29	0.29
	Std.Dev	0.0410	0.0404	0.0406	0.0397	0.0400	0.0395	0.0386	0.0379
	Min	0.24	0.24	0.24	0.24	0.24	0.24	0.23	0.23
	Max	0.42	0.42	0.42	0.42	0.41	0.41	0.40	0.39
Observations		50	50	50	50	50	50	50	50

Notes: The table shows descriptives statistics of LASSO-based forecasts for the incidence bite (upper part) and the bite gap (lower part) on the regional level. base year for the prediction: 2014.

Table A.2: Descriptive statistics of bite forecasts (individual)

Bite gap		Including wages							
		(2010)	(2011)	(2012)	(2013)	(2014)	(2015)	(2016)	(2017)
R-LASSO	Mean	0.34	0.29	0.28	0.27	0.27	0.26	0.25	0.25
	Std.Dev	0.3810	0.4184	0.4256	0.4212	0.4135	0.4122	0.3941	0.3850
	Min	-0.06	-0.16	-0.15	-0.16	-0.16	-0.16	-0.16	-0.16
	Max	2.55	2.64	2.64	2.61	2.64	2.64	2.64	2.61
CV-LASSO	Mean	0.34	0.30	0.28	0.27	0.27	0.26	0.25	0.24
	Std.Dev	0.4086	0.4439	0.4497	0.4448	0.4371	0.4357	0.4179	0.4085
	Min	-0.15	-0.24	-0.24	-0.23	-0.24	-0.23	-0.23	-0.25
	Max	2.57	2.65	2.66	2.61	2.66	2.65	2.66	2.60
Observations		544,177	556,436	561,775	564,032	568,747	571,921	576,813	579,882
Bite gap		without wages							
		(2010)	(2011)	(2012)	(2013)	(2014)	(2015)	(2016)	(2017)
R-LASSO	Mean	0.38	0.30	0.27	0.27	0.27	0.27	0.26	0.26
	Std.Dev	0.2409	0.3345	0.3499	0.3468	0.3453	0.3422	0.3380	0.3336
	Min	0.06	-0.18	-0.18	-0.18	-0.18	-0.18	-0.18	-0.18
	Max	2.04	2.19	2.19	2.19	2.19	2.19	2.19	2.19
CV-LASSO	Mean	0.39	0.30	0.27	0.27	0.27	0.27	0.26	0.26
	Std.Dev	0.2607	0.3552	0.3695	0.3660	0.3643	0.3611	0.3567	0.3521
	Min	-0.02	-0.27	-0.27	-0.26	-0.26	-0.26	-0.27	-0.27
	Max	2.13	2.27	2.27	2.28	2.28	2.28	2.27	2.27
Observations		544,177	556,436	561,775	564,032	568,747	571,921	576,813	579,882

Notes: The table shows descriptives statistics of LASSO-based forecasts for the bite gap including wages as explanatory variables (upper part) and excluding wages (lower part) on the individual level. base year for the prediction: 2014.

B. Explanatory Variables

Table B.1: Explanatory Variables - Individual Level

Group		Variables	Measurement Level	Time
(1)	Wage	Deciles of the distribution of log. monthly wages (10)	individual	Lagged 1 year
(2)	Social	Eastern Germany (2), gender (2), age-categories (12), foreign (2)	individual	contemp.
(3)	Education	School education (5), professional education (4)	individual	contemp.
(4)	Tenure	Deciles of the distribution of tenure (current establishment affiliation) in days (10), Deciles of the distribution of experience (total labor market participation) in days (10)	individual	contemp.

Notes: The numbers in brackets in the variables column show the number of categories for categorical variables.

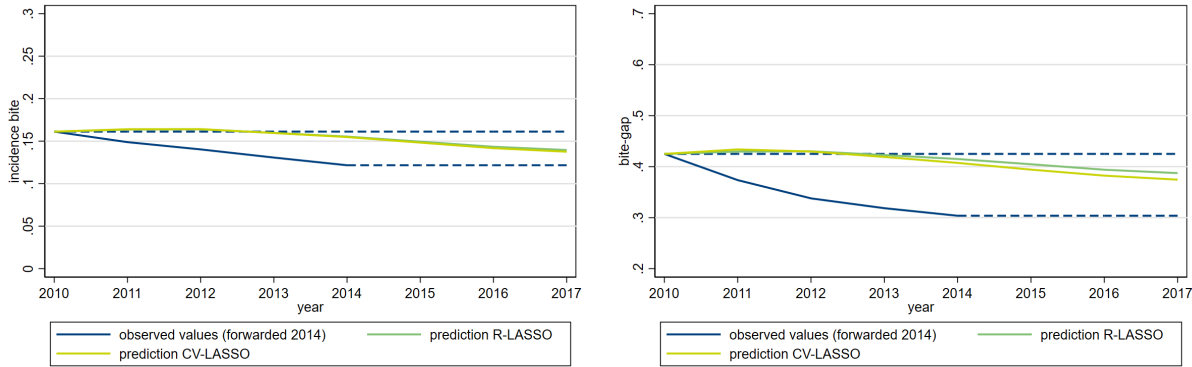
Table B.2: Explanatory Variables - Regional Level

Group		Variables	Measurement Year(s)
(1)	Wage Distribution	Percentiles of the wage distribution) (level/ change)	2010-2014
(2)	Regional Characteristics	Percentile/Percentile Ratios (level/change) dummy east Germany, share of female workers (level/change), mean educational level (level/change), share of foreigners(level/change)	
(3)	Wage and ind. Variables	(1) and (2)	
(4)	Macro Variables I	GDP, Employment, Employees Gross Value Added; (levels for each)	2009,2010
(5)	Macro Variables II	GDP, Employment, Employees Gross Value Added, (changes for each)	2009,2010 2010
(6)	Structural Variables I	size, population, type I, type II, type III, type IV	2017
(7)	Structural Variables II	Geburtensaldo, Wanderungssaldo, Alo quote Anz Betr verarbgew, Gew St, Gew Anmeldungen, Gew Abmeldungen, Insolvenzen, Beschaeftigte sozV, Alo Leistungsempfaenger	2013
(8)	Working Pop. by Branches	Number of workers in each branch	2010
(9)	Election Variables	Election outcomes of the 2009 parliamentary elections, vote shares of parties	2009

Notes: All variables are aggregated over 50 labor market regions. Data sources: BeH and various public macroeconomic data sources.

C. Bite forecasts on the regional level with base year 2010

Figure C.1: Average bite forecasts - regional level. Base year 2010



(a) incidence bite

(b) bite gap

Notes: Forecasts of the mean values (trends) for the incidence bite (a) and the bite gap (b) from R-LASSO and CV-LASSO-based predictions on the regional level. Base year for the predictions: 2010. The blue line shows the measured bite values from the BeH dataset. The dashed lines indicate forwarded values from 2010 (2014) onwards with the value from 2010 (2014).