1230 2025

SOEP papers
on Multidisciplinary Panel Data Research

Random Forests for Labor Market Analysis: Balancing Precision and Interpretability

Daniel Graeber, Lorenz Meister, Carsten Schröder, Sabine Zinn

DIW BERLIN

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)
DIW Berlin
Anton-Wilhelm-Amo-Strasse 58
10117 Berlin, Germany
Contact: soeppapers@diw.de

Random Forests for Labor Market Analysis: Balancing Precision and Interpretability

Daniel Graeber, Lorenz Meister, Carsten Schröder & Sabine Zinn

September 11, 2025

Machine learning is increasingly used in social science research, especially for prediction. However, the results are sometimes not as straight-forward to interpret compared to classic regression models. In this paper, we address this trade-off by comparing the predictive performance of random forests and logit regressions to analyze labor market vulnerabilities during the COVID-19 pandemic, and a global surrogate model to enhance our understanding of the complex dynamics. Our study shows that, especially in the presence of non-linearities and feature interactions, random forests outperform regressions both in predictive accuracy and interpretability, yielding policy-relevant insights on vulnerable groups affected by labor market disruptions.

JEL codes: C45, C53, C25, J08, I18, C83, J21,
Keyword: Machine learning, interpretability, labor market, random forests

Daniel Graeber: SOEP/DIW Berlin, IZA Bonn and CEPA, Germany. Email: dgraeber@diw.de.
Lorenz Meister: SOEP/DIW Berlin and Freie Universtät Berlin, Germany. Email: lmeister@diw.de.
Carsten Schröder: SOEP/DIW Berlin and Freie Universtät Berlin, Germany. Email: cschroeder@diw.de.
Sabine Zinn: SOEP at DIW Berlin and HU Berlin, Germany. Email: szinn@diw.de.
We thank attendants of the yearly congress of the German Association of Health Economics and the BSE Workshop on Machine Learning in Economics. The authors acknowledge funding for the project "Groups Put at Particular Risk by COVID-19 (GaPRisk)," from the German Research Foundation (SCHR 1498/7-1). The authors declare no conflict of interest. The data used in this study are drawn from the German Socio-Economic Panel (SOEP), which is publicly available for scientific research. Access to the SOEP data requires registration and adherence to the data use agreement. Ethical compliance: This study did not involve any experiments with human participants or animals conducted by the authors.

1. Introduction

Machine learning (ML) is increasingly being used in social science research, primarily for prediction, allowing researchers to forecast human behavior, social trends, and policy outcomes by analyzing complex patterns in large datasets. Among ML methods, random forests are particularly useful in situations where theoretical guidance for model specification is lacking. Random forests correspond to an ensemble learning method for classification that operates by constructing a multitude of decision trees by binary data splitting, hence enabling the study of non-linearities and variable interaction (Breiman 2001). Ensemble estimators are usually superior to individual regression tree estimators by achieving lower variance (Mullainathan and Spiess 2017). On the other hand, however, their interpretability remains a challenge (Haddouchi and Berrado 2019). For example, in contrast to traditional regression analysis, it does not yield coefficients and confidence intervals.

This paper uses and compares the predictive performance of random forests and logit regressions to identify groups that were most vulnerable to the COVID19-related labor market shock in 2020. Specifically, we identify social groups that were most likely to enter short-time work, a policy measure in Germany that allowed employees to remain employed with only small income declines, even if working hours were (markedly) reduced. Adams-Prassl et al. (2020) show that, because of the short-time work scheme, German employees faced a significantly lower risk of being impacted by the COVID19 crisis than employees in other countries. This study explores the factors that determine the use of short-time working: Is it primarily the sector of employment or the type of job, or do individual characteristics (e.g. family composition or educational attainment) matter?

We systematically investigate several potential predictors of short-time work. Our analysis considers individual characteristics such as gender, age, migration status, marital status, education, and number of children. In addition, we examine employment-related factors, including occupational position, sector of employment, full-time and part-time work experience, task type, working hours, wages, and firm size. Finally, we include regional characteristics, such as federal state, county population sector sizes, and a policy index capturing the strength of COVID-19 containment measures.

Throughout the analysis, we evaluate the performance of the random forest in our setting and employ methods to enhance its interpretability: Adjusted feature importance, an H-statistic for the relevance of feature interactions, partial dependence functions, and a global surrogate model. At the center of these methods are partial dependence functions. A partial dependence function returns the predictions of a focal outcome (i.e. being in short-time work) as a function of one or two explanatory variables of interest (e.g. by sector of employment and job type). It is defined by the related marginal density function that integrates out explanatory variables that are not of interest including all kinds of possible (higher-order) interaction effects. Thus, by construction, partial dependence functions account for all types of possible non-linearities that may exist between the outcome and all explanatory variables of a model. Thus, by examining the shape and magnitude of the partial dependence function, we can assess the relevance and effect size of the factor under consideration for predicting the outcome.

For our analysis, we utilize and combine two data sources. Our first source is the German Socio-economic Panel (SOEP), one of the largest and longest-running household panel surveys worldwide. The data contain a broad set of information on the respondents' employment situation, health, education, family type, etc. (Goebel et al. 2019). Our second source is SOEP-CoV (Kühne et al. 2020), a random sample of about 6,700 SOEP respondents, surveyed in nine staggered tranches from early April to the end of July 2020, the first year of the pandemic. SOEP-CoV collected real-time data on the following topics: a) prevalence, health behavior, and health inequality; b) labor market and gainful employment; c) social life, networks, and mobility; d) mental health and well-being; and e) social cohesion. Thus, with SOEP-CoV data, we can describe the situation of SOEP respondents during the pandemic. At the same time, we can use SOEP's broad set of respondents' characteristics before the pandemic as predictors.

Our results show that the most important determinant of entering short-time work is an individual's sector of employment. This is an expected finding, as some sectors were hit more severely by the pandemic, such as the sector for accommodation and food service activities or the manufacturing sector. The hourly wage is the second-most important determinant of short-time work. Here, we find a u-shaped partial dependence function, highlighting that low-wage earners are at risk. Personal characteristics like educational attainment, gender, and migration background are not informative about the outcome. A classification tree, based on a global surrogate, points to the interaction of sector and wage and, therefore, to related risk groups. That is, the random forest approach helps in detecting a non-linearity and interaction that a corresponding logit model cannot easily find. In terms of accuracy, the random forest marginally outperforms a logit model.

Our study adds to the existing literature in two fields: economic inequality and the application of interpretable machine learning methods in economics. In the field of economic inequality, it contributes to the context of the COVID-19 pandemic (see Stantcheva (2022) for a review). Inequality in household income generally increased during the first months of the pandemic, but this effect subsequently reversed due to government compensation schemes (Clark, d'Ambrosio, and Lepinteur 2021; O'Donoghue et al. 2020; Li et al. 2020). Closures of companies and layoffs played an important role and employees' risks of being affected were rather heterogeneous. In the United States, mostly low-wage workers were laid off, particularly as a response to declining revenues in small businesses (Chetty et al. 2020). Thus, our study builds upon and advances the existing literature by providing a comprehensive quantitative assessment of the role of employees' characteristics for short-time work.

Our paper also contributes to a growing of body literature on machine learning in economics. This literature covers a broad range of topics including the determinants of subjective well-being (Clark et al. 2022; Oparina et al. 2199), judicial bail decisions (Kleinberg et al. 2017), physicians' diagnoses (Mullainathan and Obermeyer 2022), and labor market matching (Mühlbauer and Weber 2022). However, most of these studies primarily address prediction accuracy, without delving into the functional form relationship between the individuals' characteristics and the outcome of interest. One exception is the work of Clark et al. (2022), who use Shapley Values for exploring the direction and magnitude of relationships between determinants and outcome variables. In contrast, our study uses random forests and partial dependence functions to identify risk groups that were severely affected by the short-term measures used during the COVID-19 pandemic. With this, we are among the first to highlight the potential of interpretable machine learning for economic research.

The remainder of this paper is structured as follows. Section 2 introduces the random forest algorithm and the decision trees it comprises. Section 3 discusses various methods to interpret machine learning algorithms. Section 4 contains the application of the aforementioned methods to the context of the pandemic labor market shock. Section 5 concludes.

2. Random Forests

Random forests were developed in the context of statistical classification methods (Breiman and Cutler 2001; James et al. 2013). Due to their intuitive appeal, random forests are widely used, constituting an important method in the machine learning toolbox. Since they are typically trained to predict an outcome, associated with a range of explanatory variables, they belong to the class of supervised learning approaches (James et al. 2013). Subsequently, we briefly describe the general model of random forests and a common method to train these models, i.e., optimize the hyperparameters that govern this model. Thereafter, we describe partial dependence functions and how they are used to estimate the functional relationship between the target and the features in the final random forest model. In our description, we use machine learning terminology, i.e., explanatory variables are denoted as features, the outcome variable as target, and estimating a model is called training.

Random forests are decision tree ensembles (James et al. 2013). Decision trees are tree-structured models for classification and regression. They are generated by an algorithm that recursively splits the sample along features such that the observations in the resulting subsamples, or *nodes*, are as similar as possible with respect to the target and as dissimilar as possible between nodes.

Formally, at every level k, the data D are divided into two nodes $D_{k,L}$ and $D_{k,R}$. The exact split is characterized by one feature X_j and the threshold associated with this split, $\gamma(k, j)$. Consequently, nodes $D_{k,L}$ and $D_{k,R}$ are defined as follows:

$$D_{k,L} := \{x|x_j < \gamma(k, j)\};\ D_{k,R} := \{x|x_j \geq \gamma(k, j)\}.$$

The optimal splitting feature X_j^* and associated threshold $\gamma(k, j)^*$ for each node are determined by minimizing the impurity across nodes:

$$\min_{\forall\, x_j \in x \text{ and } \gamma(k,j) \in R_{x_j}} I_{k,L} + I_{k,R},$$

where $I_{k,m}$, with $m \in \{R, L\}$, is the respective impurity measure. For regression tasks, impurity is usually measured by the mean squared error, a natural candidate for continuous targets. For binary or categorical targets, as in our application, the Gini impurity or Entropy index is used.

Realizing optimal splits at each stage of the recursion results in a tree consisting of a root node, parent nodes, and child nodes. The nodes on the last level of the tree are called *leaves*. These leaves contain the predictions of the target with respect to the trained tree: The predictions are the averaged values of all target values in a leaf. For categorical outcomes, to assign a label to the observation, a threshold must be determined.

We use random forests for estimation due to their advantages over single decision trees in usability and performance. Single trees risk overfitting, with low bias but high variance. Random forests mitigate this by fitting multiple trees to random, bootstrapped samples (typically 75% of the data), using the remaining 25% for validation. Predictions are averaged across these trees to reduce variance. Additional techniques, such as limiting features at each split, reducing the sample fraction, and restricting tree depth, further control variance (James et al. 2013). These hyperparameters are usually optimized through cross-validation.

3. Interpretable Machine Learning

Originally developed for prediction tasks, random forests have gained popularity for explanatory studies due to their intuitive appeal and straightforward implementation. However, applying random forests to explanatory problems requires new methods to interpret the interrelationships they model. While this field is still nascent, a few viable approaches exist (see Molnar, 2022), though they remain underutilized in economics. In the following sections, we outline the four methods we later apply to detect individual vulnerabilities in the labor market.

3.1. Variable Importance Measure

Our first tool to assess the explanatory power of features in random forests is the variable importance measure (VIM), which allows us to assess the importance of each feature in predicting short-time work. For each X_j, it returns the increase in purity moving from the parent node to the child nodes, averaged over all trees in the forest (James et al. 2013). Thus, the VIM for a feature corresponds to the average change of impurity if the respective feature is chosen as splitting feature at the respective nodes. The VIM might yield biased results if the set of features contains continuous and categorical variables (Sandri and Zuccolotto 2008; Nembrini, König, and Wright 2018). The reason is that continuous variables, by their nature, offer more splitting possibilities than categorical variables. Thus, due to the heuristic nature of tree-method optimization procedures, continuous variables have a higher probability of being considered a splitting variable. A way to correct for this is the adapted variable importance measure described by Sandri and Zuccolotto (2008) and Nembrini, König, and Wright (2018). Intuitively, each change in the variable importance, which results from a split, comprises the true change and a change that results from the bias. The bias emerges due to the characteristics of the feature under consideration. Sandri and Zuccolotto (2008) and Nembrini, König, and Wright (2018) postulate to augment the feature matrix with pseudo-features that are completely uninformative but share the same characteristics as the original feature. As a result, the true change of the impurity measure for these pseudo-features is in expectation equal to zero and the bias due to the characteristics, which the pseudo-feature shares with the original feature, remains. The bias correction then works by correcting the VIM of the original variable by the VIM of the pseudo-feature and averaging over all permutations.

The advantage of the VIM is that it is usually readily available after estimating the random forest. However, the measure does not convey any information about interactions between variables nor any information about non-linearities in the association between the features and the target. Furthermore, it does not provide any information about the direction of the association between the features and the target (i.e., whether it is positive or negative). Therefore, the VIM is of limited use in gaining insights into more complex structures between features and the target. For instance, economic theory dictates decreasing marginal utility from consumption and diminishing returns from labor in the production function, just to name a few examples. The latter is mapped in classical production functions. If one is interested in the returns to scale or about complementarities according to certain variables (such as labor market sector or company size), variable importance measures are not very informative. However, partial dependence functions offer a very intuitive way do address these issues.

3.2. Partial Dependence Function

The second method we employ is the computation of partial dependence (PD) functions. They visualize the functional relationship between our predictors and short-time work. They allow to infer whether the relationship between features and the target is linear, monotonic, or non-linear (Friedman 2001; Molnar 2022). It is specified by the marginal density function f of the target depending on one or two features X_S while integrating all other features X_C out:

$$PD_S(x_S) := f(x_S) = \int f(x_S, X_C) d\mathbb{P}(X_C). \tag{1}$$

This yields an average marginal effect of X_S on the target, similar to average marginal effects in binary regression models. The PD function comprises main effects, but also all kinds of higher order interaction effects included in the trained tree ensemble. The empirical estimator of PD is

$$\hat{f}(x_S) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_S, x_C^{(i)}), \tag{2}$$

where $\hat{f}(x_S, x_C^{(i)})$ are the predictions from the trained random forest model for the feature tuple $[x_S, x_C^{(i)}]$ and n is the number of distinct value combinations of x_C, i.e., the number of observations. That is, the estimator of the PD function results from averaging over all predicted targets for the observed values of the features X_C at the distinct values of X_S.

The PD function allows a convenient illustration of its values by plotting it for each realization of X_S. The resulting PD plot then visualizes the relationship between the features of interest and the target.

Essential for the PD to work is its assumption that all features X_C are uncorrelated to the features X_S. If this assumption is not met, the interconnected parts of X_C and X_S will enter the calculation twice, leading to a biased result. One way to get around this problem is to calculate conditional densities instead of marginal densities (as it is the case with the PD function). This results in Accumulated Local Effect (ALE) plots (Molnar 2022).

While partial dependence functions provide an intuitive way to visualize complex relationships, their interpretability depends on the assumption that the focal features are not strongly correlated with the rest of the feature set. In practice, however, socioeconomic variables—such as sector, wages, or firm size—are often correlated. As a result, interpreting PD functions as reflecting clean marginal effects may be misleading when this assumption is violated. We acknowledge this limitation and discuss results with caution, focusing on patterns rather than strict causal interpretations. To address these dependencies more systematically, we complement PD plots with additional techniques that do not rely as strongly on the independence assumption. In particular, the H-statistic provides a useful way to quantify interaction effects directly and helps to validate whether observed patterns stem from genuine interactions or collinearity-related distortions.

3.3. H-statistic

The third tool we use to increase the random forest's interpretability is a statistic to quantify the relevance of feature interactions is the so-called H-statistic. The definition of the H-statistic builds on the condition that the PD function is additively separable in its arguments, i.e., the features do not interact (Friedman and Popescu 2008). That is, for two features j and k that do not interact, it holds that

$$PD_{jk}(x_j, x_k) = PD_j(x_j) + PD_k(x_k),$$

where PD_j and PD_k are the respective PD functions of X_j and X_k. Generalizing the concept yields

$$PD_{SC}(x_S, x_C) = PD_S(x_S) + PD_C(x_C). \tag{3}$$

Here, the joint distribution of the features X_S and X_C decomposes into a univariate part $PD_S(x_S)$ for the single feature X_s and multivariate part $PD_C(X_C)$ for the remaining feature set X_C. If the assumption that X_S and X_C are uncorrelated does not apply, inequality arises due to unaccounted interactions. In other words, the difference between the left hand and right hand side of Equation 3 quantifies the unconsidered interaction effects between the two focal feature sets. Relating this difference to the overall variation in the PD function yields the H-statistic (Friedman and Popescu 2008). For two single features X_j and X_k, it is defined as:

$$H^2_{jk} := \frac{\sum_{i=1}^{n}(PD_{jk}(x_j^{(i)}, x_k^{(i)}) - PD_S(x_j^{(i)}) - PD_k(x_k^{(i)}))^2}{\sum_{i=1}^{n} PD_{jk}(x_j^{(i)}, x_k^{(i)})^2}, \tag{4}$$

where n is the number of all distinct value combinations of the tuple $[x_j, x_k]$. For a single feature X_S, related to all other features X_C, the H-Statistic is (Molnar 2022):

$$H^2_S := \frac{\sum_{i=1}^{n}(PD_{SC}(x_S^{(i)}, X_C^{(i)}) - PD_S(x_S^{(i)}) - PD_C(X_C^{(i)}))^2}{\sum_{i=1}^{n} PD_S(x_S^{(i)})^2}. \tag{5}$$

We estimate the H-statistic using the empirical counterparts of the respective PD functions. It is bounded between zero and one. The form of interaction can be inferred via multivariate PD functions. A multivariate PF function is a PD function that returns the joint prediction for two or more features, after all other features have been marginalized out. If we plot these, we can infer the exact details of the interaction of the features involved.

3.4. Global Surrogate Model

Our final tool is a global surrogate model, which reduces the complex structure of the random forest. The outcome will be a simpler model that can be read and interpreted. For example, a linear regression model for continuous targets, a logit regression for binary targets, or a decision tree.

The creation of a global surrogate model involves three steps (Molnar 2022):

(a) A random forest is trained on the data.

(b) Predictions for the targets are derived from this random forest.

(c) The global surrogate (e.g., a random forest) is fitted on these predictions and the associated features.

A global surrogate model only constitutes a simplified version of the original machine learning model, mapping parts of its complex model structure into an interpretable framework.

4. The Risk of Working in Short-time during COVID-19

4.1. Background

To compensate employers and employees for the economic hardships associated with containment measures and the COVID-19 pandemic, the German government initiated an economic relief package in April 2020. The main instrument in this relief package to keep employees in employment was Kurzarbeit, or short-time work. Short-time work in the employment relationship means the temporary reduction of regular working hours due to a significant loss of work in a company. Short-time employees work less or not at all, and the arrangement may affect all or only some of the company's employees. To partially compensate for the employees' earnings loss, short-time allowance can be claimed at the Federal Employment Agency as a compensation payment. This compensation amounts to 60% of the net pay difference for the month(s) in short-time work, or 67% for employees with children (§105 and §106 Sozialgesetzbuch III). In April 2020, the allowance was temporarily increased through the end of 2020—later extended to 2021—with rates rising to 70% (77% with children) from the fourth month and 80% (87%) from the seventh month for those experiencing at least a 50% reduction in working hours.

Importantly, participation in short-time work typically required agreement between employers and employees, often negotiated through collective bargaining structures or works councils. This means access to the scheme was not purely automatic or unilateral; it depended on institutional arrangements at the firm level and the capacity of firms and workers to coordinate. As such, characteristics like firm size, sector, and union presence may reflect not only economic exposure to the crisis but also the ability to engage in this negotiation process—an important consideration when interpreting the patterns our machine learning models uncover.

In this environment, it is of utmost importance for policy makers to understand how economic hardships in the population emerge. For this sake, we are concerned of validating and understand how employee's characteristics are predictive of short-time working.

4.2. Data

Our estimation relies on SOEP and SOEP-CoV data to predict at risk groups and to understand the relationship between the features and the likelihood of working in short-time during the COVID-19 pandemic. The SOEP is a representative panel of German households and its household members, with the first annual wave collected in 1984. Data is collected on a wide range of individual and household related topics such as individuals' labor market histories, education, household compensation, and health, among others. Overall, nearly 15,000 households and 30,000 individuals participate in the SOEP survey (Goebel et al. 2019).

SOEP-CoV (Kühne et al. 2020) is an add-on survey conducted in the second quarter of 2020. It provides information on the experiences of SOEP households and their members during the onset of the COVID-19 pandemic. Most importantly for our study, the SOEP-CoV asked labor market participants if they work in short-time or not. Thus, our target is an indicator that is equal to one if respondents indicate that they work in short-time and zero otherwise. The description and operationalization of the target and features is displayed in Table 1.

This add-on survey includes items from many domains that seem relevant for the prediction of short-time work. These include labor market experiences, the individuals' health situation, attitudes, and family situation, among others. Our dataset is ideally suited for identifying and researching risk groups in the context of the COVID-19 pandemic. This is because it enables extrapolation to the population level, given its foundation on a random sample, thereby facilitating generalization. The two SOEP datasets include a rich set of predetermined characteristics of households and their members that are entered as features in our random forest model. Notably, a major advantage over administrative data is the fact that it also contains information on household characteristics, such as details about cohabitation or the presence of children, as well as individual-level characteristics, such as migration background or education.

TABLE 1. Description of features and labels

Feature	Description
Occupational position	Contains nine categories based on occupation (worker vs. salaried employee) and skills (e.g. unskilled, skilled, untrained, trained)
Gender	Indicator that is equal to one if the individual is female and zero otherwise.
Age	Difference between survey year and year of birth.
Migration background	Indicator that is equal to one if the individual has a direct migration background.
Sector	Categorical feature indicating the sector of the individual according to the Statistical Classification of Economic Activities in the European Community (revision 2, also called NACE). The sector corresponds to the first NACE level. A detailed description of all sectors can be found in Appendix Table A1.
Share of experience in part-time	The time spent in working part-time divided by the sum of the full-time, part-time, and unemployment spells.
Share of experience in full-time	The time spent in working full-time divided by the sum of the full-time, part-time, and unemployment spells.
Marital status	Indicator that is equal to one if the individual is married.
Education	Categorical feature of the highest educational degree according to the Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) scheme. Includes 10 categories.
Tasktype	Main tasktype. One of the following five potential realizations: Analytic non-routine tasks, interactive non-routine tasks, cognitive routine tasks, manual routine-tasks, manual non-routine tasks. Includes five categories.
Working hours	Actual weekly working hours
Children	Categorical features for number of children
Federal state	Categorical feature for the federal state. Includes 16 categories.
Hourly wage	Weekly gross labor earnings, divided by weekly working hours.
Firm sizes	Categorical feature with the following realizations für the number of employees in the firm the respondent is working in: "$<$ 20", "20 – 199", "200 – 1999", "$\geq$ 2000".
County level population incidence	County level population over county size in 2019.
Share of the secondary sector on overall county GDP in 2019	Secondary sector output over total output (county level).
Share of the tertiary sector on overall county GDP in 2019	Tertiary sector output over total output (county level).
Policy index	Summary index reflecting the overall strength of the policy measures to contain the COVID-19 pandemic (federal state level).

Note: Table 1 displays the description of the features and associated labels for categorical features.

4.3. Predictive performance and interpretation

To ensure that the random forest approach yields at least as good results as a logit model with only main effects, we compare initially both approaches with respect to their predictive accuracy. Then, we proceed with the application of the methods from the domain of interpretable machine learning.

4.3.1. Comparing predictive performance of logit model and random forest

Tuning. Training a random forest model requires specifying its hyperparameters *number of features per split*, the *minimal node size*, and the *sample fraction per tree*. This step is called tuning. We tune these hyperparameters for the random forest via sequential model-based optimization with 30 warm-up iterations and 70 main iterations. As evaluation criterion, we used the Brier score. The Brier Score measures the accuracy of probabilistic predictions by calculating the mean squared difference between predicted probabilities and actual outcomes, with lower scores indicating better performance. The resulting hyperparameters are displayed in Table 2. These hyperparameter values are employed in all subsequent analyses.

TABLE 2. Optimal hyperparameters for random forest

Hyperparameter	Optimal value
Number of features per split	13.00
Minimal node size	95.00
Sample fraction per tree	0.67

Note: Table 2 displays the optimal hyperparameters based on sequential model based optimization with 30 warm up iterations and 70 main iterations. The evaluation criterion was the Brier score. The number of features per split indicates the optimal number randomly drawn features at each split in each decision tree of the respective random forest. The minimal node size indicates the minimal number of observations in the terminal nodes. The sample fraction indicates the size of the randomly drawn sample, with replacement, which is used to grow the decisions trees on.

Evaluation. To evaluate the predictive performance of a random forest model with the above hyperparameters fixed and an equivalent logit model with only main effects, we use three standard measures of the machine learning toolbox: the confusion matrix as well as the receiver operator characteristic (ROC) curve and its area under the curve (AUC). All three measures allow assessing how well binary classification models, such as a random forest with a binary target and a logit model, perform in their task to assign data points to the correct bin. For this, a confusion matrix juxtaposes the predicted and observed values in the form of a 2×2 contingency table. The ROC is a graphical illustration of the true positive rate against the false positive rate at a number of different candidate cut-off values between 0 and 1, and the AUC is the corresponding integral. Beware that each instance of the confusion matrix represents one point in the ROC space.

To be able to quantify predictive performance, we first retain a random subsample corresponding to 25% of our analytic sample as test data and then train both models on the remaining data. Note that there exists no unambiguous rule on how to split the data between test and training data. An increase in training data increases variance in the evaluation statistics will have more variance. On the other hand, the increase in test data will increase variance in the model will show higher variance in training. We believe that our split is a good middle ground of recommendations, that typically include splits of size 1/5 and 4/5 as well as 1/3 and 2/3 of test and training data, respectively.

Overall, we find that the random forest performs marginally better than the logit model. First, we select the optimal cut-point to allocate the respective predictions for each model such that it maximizes the sum of the specificity and sensitivity, i.e., minimizes false positive and false negative predictions jointly. The results are displayed in Table A3. For the random forest, the optimal cut-point is 0.223. For the logit model, the optimal cut-point is 0.200. Considering sensitivity and specificity separately no model is superior to the other. Based on the associated sensitivity and specificity, the ordering of the two models is ambiguous. However, considering the overall sum of the two quantities, the random forest clearly outperforms the logit model.

Table A2 displays the confusion matrix of both models with the associated optimal cut-points. Overall, the accuracy of the logit model is 0.71. That is, about 71% of the individuals in the test data are correctly classified. For the random forest, this figure is equal to 0.72.

TABLE 3. Optimal cut-point

	Opt. cut-point	Sensitivity	Specificity
Random forest	0.223	0.673	0.727
Logit	0.200	0.609	0.730

Note: Table A3 displays the optimal cut-point, which maximizes the sum of sensitivity and specificity, to classify the predictions based on the random forest and Logit model as well as the associated sensitivity and specificity.

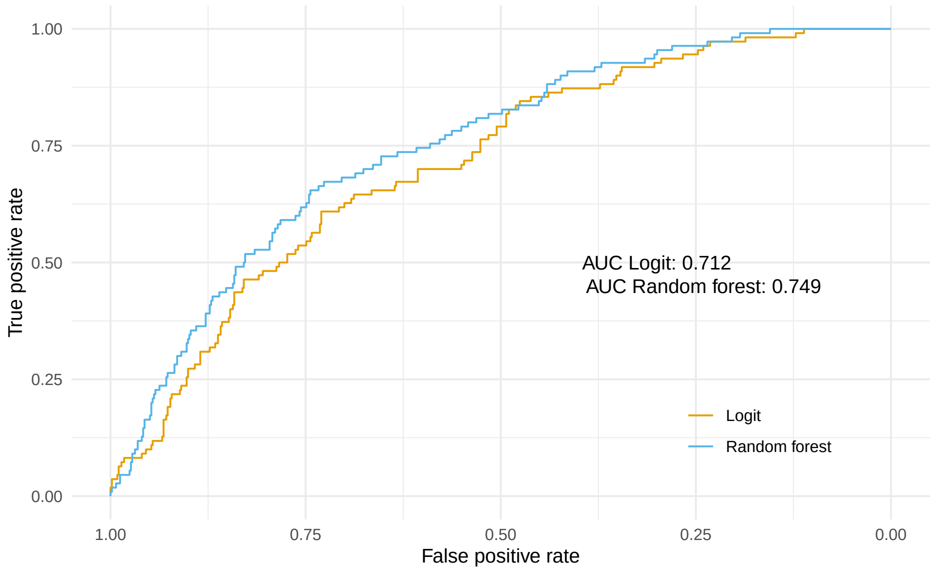
The models' ROC curves, displayed in Figure 1, indicate that the random forest is more efficient than the logit model. Concretely, compared to the logit model, we find that for almost all cut-offs considered, the true positive rate of the random forest model is larger than its false positive rate. This is also reflected in the respective values for the AUC, which is 0.71 for the logit model and 0.75 for the random forest.

TABLE 4. Confusion matrices

	Opt. cut-point	True positive	False negative	False positive	True negative
Random forest	0.223	74	36	157	417
Logit	0.200	67	43	155	419

Note: Table A2 displays the optimal cut-point to classify individuals in the test data based on the predictions. The optimal cut-point is such, that it maximizes the sum of sensitivity and specificity. True positives are individuals which are predicted to be in short-time work and worked in short-time. False negatives are individuals which are predicted not to work in short-time work but worked in short-time. False positives are individuals that are predicted to work in short time but did not work in short-time. True negatives are individuals that are predicted not to work in short-time and did not work in short-time.

FIGURE 1. Receiver operator characteristic curve



Note: Figure 1 displays the receiver operator characteristic curve (ROC) for the logistic regression (Logit) model and the random forest, and the related area under the curve (AUC).

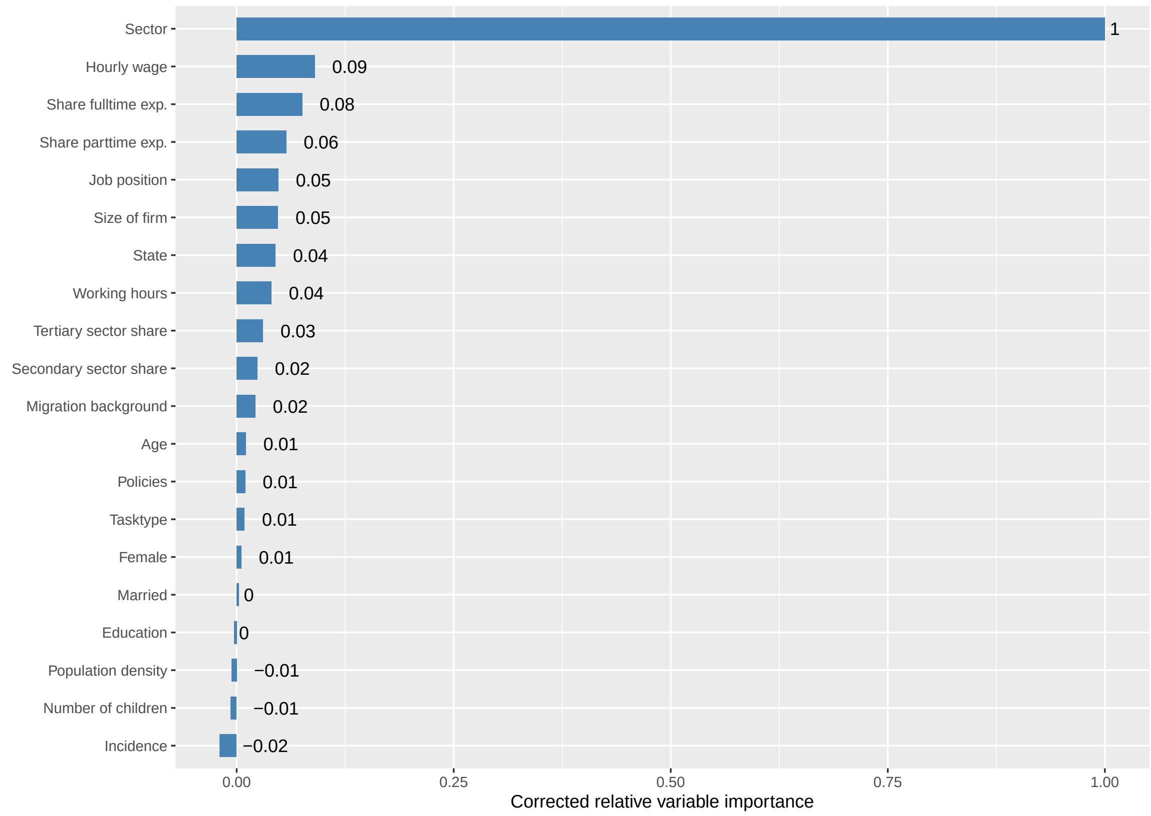
4.3.2. Interpretation

Variable Importance. Figure 2 shows the (adapted) VIM for each explanatory variable considered in our random forest model. To make the values comparable and easier to interpret, the VIMs are standardized with the highest VIM being 1. VIMs can also be negative; this happens whenever a feature contributes noise to the model and thus its presence hurts performance.

The main conclusion is that the sector in which the respondent works is the most important feature in predicting whether respondents work in short-time. Furthermore, there exists a substantial difference in the VIM of the respondents' sector and that of the other features. For example, the importance of the hourly wage, the second most important feature, relative to the VIM of the respondent's sector, amounts to 0.09.

A second conclusion is that it is mainly the job characteristics that matter, in contrast to individual-level attributes. For instance, the five most important features in explaining whether respondents work in short-time are the respondent's sector, hourly wage, the share of full-time and part-time experience, and their job position. In contrast, migration background, age, gender, and education, to name a few individual socio-demographic characteristics, are ranked in the bottom half of the ranking.

FIGURE 2. Relative feature importance related to short-time work during the onset of the Corona pandemic.



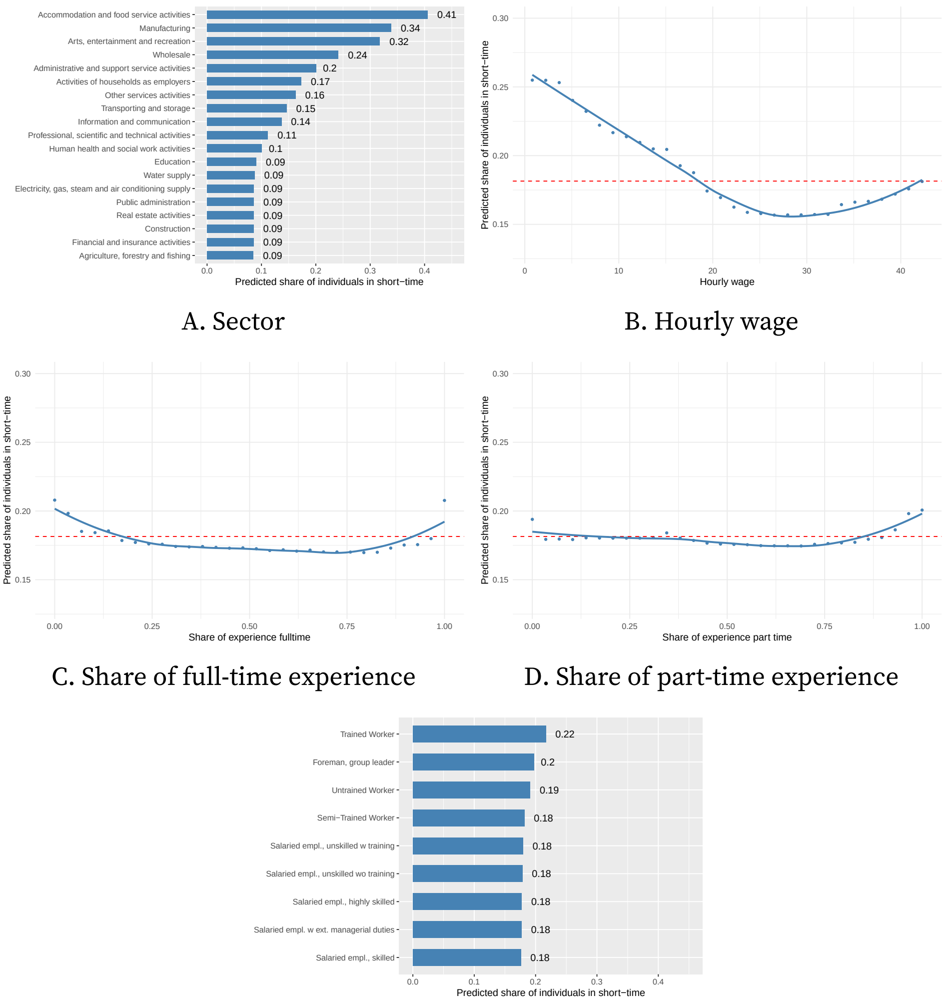
Note: Figure 2 displays the relative corrected feature importance. The feature importance is defined as the average decrease in the target function associated with the feature. If adding a feature leads to an increase in the target function, the VIM turns negative. All values have been standardized to be denoted in relative terms, in comparison to the variable importance with the highest realization.

Functional Form between Features and Target. PD plots gauge the degree of non-linearity in the associations between the model's features and the target "short-time work during the onset of the COVID-19 pandemic." For reasons of clarity, we focus on the top five features of the feature importance statistic.

The Panel 3 displays the predictions based on the related PD functions. For the two categorical variables "respondent's sector" and "occupational position," the predicted share of individuals in short-time work is predicted across their possible categories. For the three continuous features "hourly wage," "share of full-time experience," and "share of part time experience," the PD plots show the predicted share of individuals in short-time work within a bin of the feature (blue dots) along the predictions based on a

LOESS[1] regression (blue lines). The dashed red lines correspond to the unconditional average probability of working in short-time in our sample.

FIGURE 3. Partial dependence plots of the random forest model for mapping short-time work during the onset of the COVID-19 pandemic.



A. Sector

B. Hourly wage

C. Share of full-time experience

D. Share of part-time experience

E. Occupational position

Note: Figure 3 displays the realisations of the empirical PD functions for the features “Sector”, “Hourly wage”, “Share of full-time experience”, “Share of part-time experience” and “Occupational position”. The bar plots display the realisation of the PD functions for categorical features. For continuous features, these values are displayed with dots associated with each bin of the support. Solid blue lines depict the fit of a LOESS estimation. The dashed red line corresponds to the unconditional sample average of the target.

Note that the more variation a PD function displays, the more important the feature

[1] Locally Estimated Scatterplot Smoothing.

is for predicting the target. Therefore, for instance, the PD function for employment sectors and hourly wage, displayed in Figures 3A and 3B, display significant variation. In contrast, the PD function for the share of full-time and part-time experience as well as the occupational position display considerably less variation.

Beyond this, the PD plots provide information about the impact of distinct values of features as well as the functional form of the underlying relationship. In particular, the share of people working in short-time varies strongly over sectors in a non-linear way. The sector "Accommodation and food service activities" exhibits a predicted share of individuals working in short-time during the onset of the pandemic of approximately 38%. In contrast, the share for "Human and health and social work activities" only is 10%. Thereafter, the same statistic is 9% for the remainder of the sectors. Notably, since all other features are marginalized out, this can be interpreted as an independent effect of sectors.

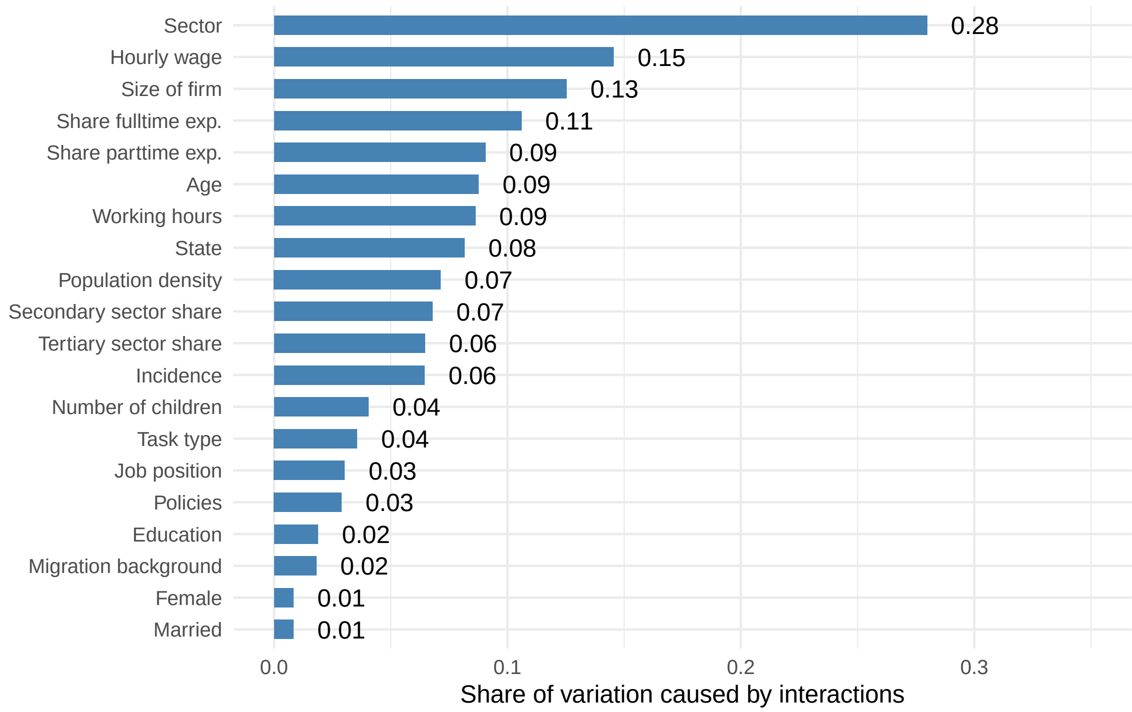
Turning to the the hourly wage, the association between hourly wages and the probability of being assigned to short-time work is strongly negative for a wide range of the distribution of hourly wages, depicted in Figure 3B. For low hourly wages, the share of individuals working in short-time is close to 28%. From there, the share decreases to approximately 16% for hourly wages up to 30 euros. Above 30 euros, this relationship fades out and turns positive for very high hourly wages. Clearly, this is indicative of a highly non-linear relationship between hourly wages and the probability of working in short-time.

Turning to labor market experiences, individuals with very low labor market attachment before the COVID-19 pandemic are more likely to be predicted to work in short-term jobs (see Figures 3C and 3D, respectively). For individuals who spend up to 25% of their time on the labor market in full-time employment, the predicted share of individuals working in short-time during the onset of the COVID-19 pandemic is above average and declining from 0% to 25% experience in full-time. For the range above 25%, there exists almost no variation in the predicted likelihood of working in short-time. Similarly, individuals who spend more than 75% of their time in the labor market in part-time employment, experience an above average risk of working in short-time during the onset of the COVID-19 pandemic. For individuals who spend less than 75% of their labor market experience in part-time, the variation in the probability of working in short-time does not vary in a significant way.

Turning to the PD function for the occupational position, Figure 3E shows that there exists almost no variation in the predicted likelihood of being in short-time across occupational positions. Trained and untrained workers were at somewhat higher risk of being in short-time during the onset of the COVID-19 pandemic. However, the differences only range from one to three percentage points.

Feature interactions. Next, we investigate the relevance of interactions between features. Figure 4 displays the Friedman's H-statistic for all features. The Friedman H-statistic for the features "Sector" and "Hourly wage" are 0.28 and 0.15, respectively. That indicates that about 28% and 15% of the variation that is explained by these two variables can be explained by interactions. The corresponding values are 0.13 and 0.11 for the "Size of firm" and the "Share of full time experience". The same figures are below 0.1 for the other features.

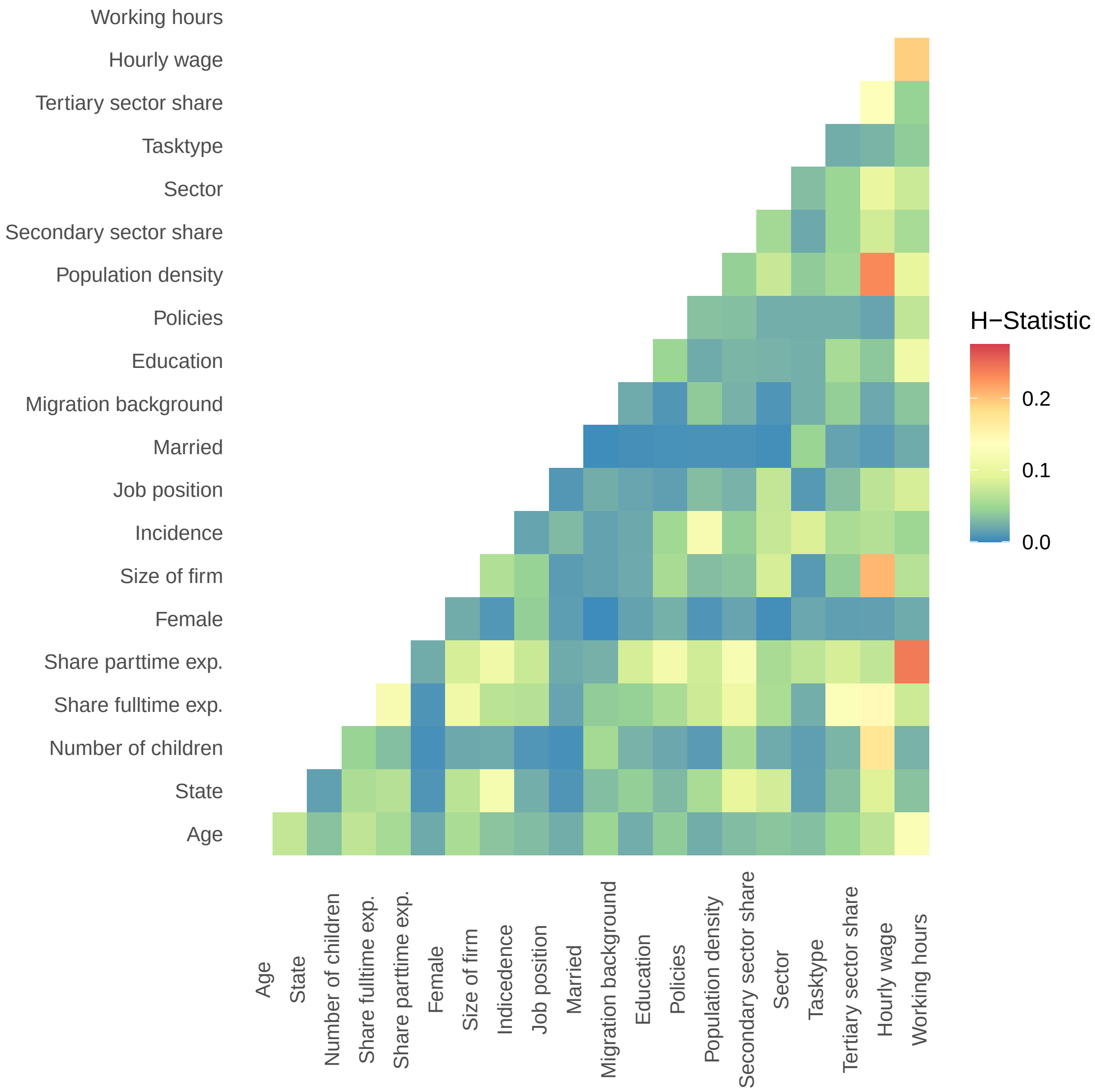
FIGURE 4. Friedman's H-statistic for mapping the relevance of interactions of features.



Note: Figure 4 displays Friedman's H-statistic for all of the features in our sample. The H-statistic corresponds to the sample version of Equation 5.

The strengths of the pairwise feature interactions, according to the bivariate H-statistics, are illustrated in Figure 5. A closer look reveals that none of the socio-demographic attributes interact in any relevant way with job characteristics. In contrast, we find that job characteristics mainly interact with each other. For instance, the hourly wage, and working hours interact strongly with other features. The working hours interact strongly with hourly wages and the size of the firm. The hourly wage interacts strongly with the individuals part-time experience and the population density.

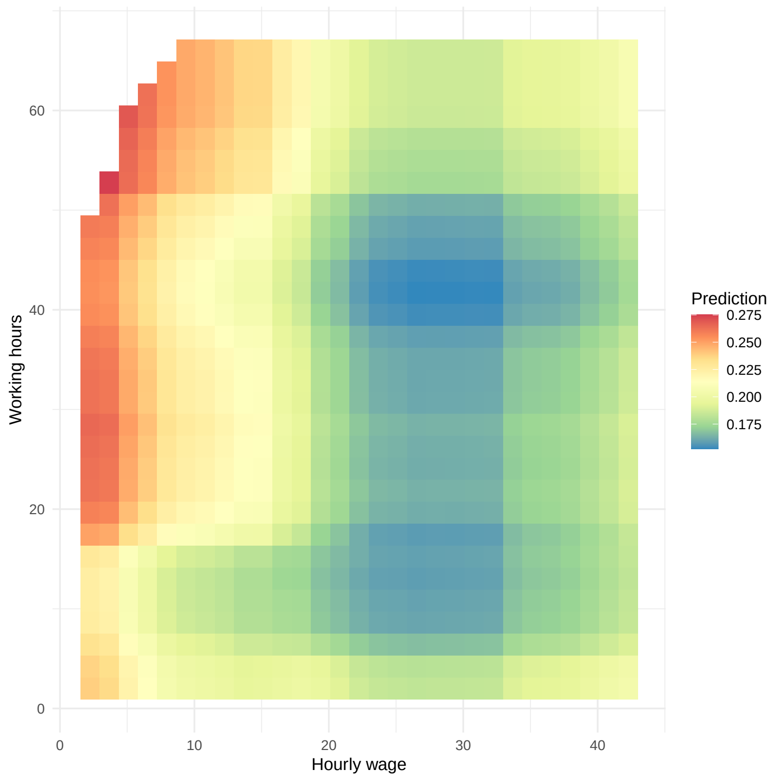
FIGURE 5. Strength of interactions between the model features related to short-time work during the onset of the COVID-19 pandemic.



Note: Figure 5 displays Friedman's H-statistic for pairwise combinations of features based on the empirical correspondence of Equation 3.

To gain deeper insight into the strength of interaction between hourly wage and weekly working hours with respect to short-time work, we use the corresponding bivariate PD function. Figure 6 plots the related result. Here, we see that, among individuals who have hourly wages up to 15 euro, those with particular high weekly working hours are more likely to have been working in short-time during the COVID-19 pandemic.

FIGURE 6. Strength of interaction of weekly working hours with hourly wage related to short-time work during the onset of the Corona pandemic.



Note: Figure 6 the empirical realizations of the bi-variate PD function associated with working hours and hourly wage.

Approximating the Complex Model Structure. The surrogate model, as described in Section 3.4, is visualized in Figure 7. In line with our earlier results, the primary splitting feature in the decision tree is the sector of employment, which divides the sample into two groups: one with a low incidence and another with a high incidence of short-time work. Specifically, Node 2 represents individuals employed in a large number of sectors identified at the first split, accounting for 66% of the sample, where only 10% are predicted to be in short-time work. In contrast, Node 3, comprising the remaining four sectors and 34% of the sample, has a significantly higher short-time work incidence of 32%. This initial split maximizes the separation between these groups by minimizing impurity, as described in the theory section.

Within Node 3, a second split occurs, again based on the respondent's sector of employment in 2019. This yields two distinct subgroups: Node 6, consisting of individuals working in sectors 18 "Arts, Entertainment, and Recreation" and 7 "Wholesale and Retail Trade", where 24% are in short-time work; and Node 7, covering the remaining sectors of "Manufacturing" and "Accommodation and Food Services", where the incidence is even higher, at 35%. Notably, sectors with too few individuals, such as 2 "Mining and Quarrying" and 21 "Activities of Extraterritorial Organizations", are excluded from this analysis for robustness.

For individuals in Node 2, hourly wage becomes the critical splitting feature. Respondents earning less than €17 per hour are grouped into Node 5, where 15% are predicted to be in short-time work. Conversely, those earning €17 or more are in Node 4, with a lower predicted incidence of 7.8%. This highlights the role of wage disparities in shaping the risk of short-time work.

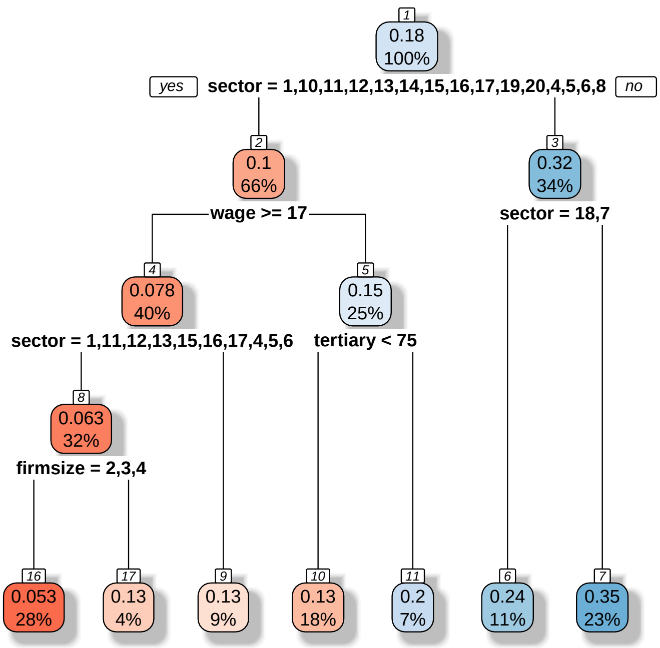
Further splits reveal additional layers of complexity. Among individuals in Node 5, the share of the tertiary sector in the local economy is a decisive factor. In areas where the tertiary sector comprises less than 75%, the predicted short-time work rate is 13%, compared to 20% in regions with a higher tertiary sector share. Similarly, for individuals in Node 4, sector affiliation remains a pivotal determinant. For those in Node 8, firm size emerges as a critical feature influencing short-time work probability. This last split yields the leaf that contains the group with the lowest short-time work incidence of only 5.3%. Following the branches from this leaf back to the top of the tree allows to infer who is in this group: Individuals who earn at least 17€ per hour, work in a firm with at least 20 employees, and are employed in the sectors 1, 4, 5, 6, 11, 12, 13, 15, 16, or 17.[2] Similarly, one can apply this procedure to determine the compositions of all other leaves at the bottom of the tree.

Overall, the decision tree provides a simplified yet powerful visualization of the key factors influencing the likelihood of short-time work during the COVID-19 pandemic. The sector of employment emerges as the dominant determinant, consistently shaping the splits and reinforcing its central role in understanding short-time work dynamics. Beyond the sector, individual characteristics such as hourly wage and the economic composition of the local area offer additional explanatory power. By distilling the

[2]See Appendix Table A1 for a detailed description of sector numbers.

complexity of the random forest into an interpretable framework, the decision tree not only confirms our earlier findings but also offers actionable insights for policymakers aiming to address disparities in short-time work risk. When analyzing groups at risk by the means of a random forest, taking a look at the composition of the leaves allows policy-makers to tailor transfers and aid to a specific socio-demographic group that has the highest need. More generally, this model demonstrates the value of interpretable machine learning techniques in bridging the gap between data-driven analysis and policy relevance.

FIGURE 7. Single tree model as global surrogate.



Note: Figure 7 displays the estimate of a decision tree, trained on the predictions for the complete sample. A detailed overview of sectors and corresponding numbers can be found in Appendix Table A1.

5. Conclusion

Powerful tools, such as machine learning, are finding more and more applications in the social sciences. These models are powerful ways to improve the predictive accuracy of models. However, the usage of these tools is often limited by the fact that those models provide only limited information about how the various features are associated with the target of interest.

We present flexible methods that allow for capturing two functional forms realizations, which are typically considered to be captured well with machine learning methods. At the core of these methods are PD functions, which allow for investigating non-linearities and interactions between features. These are relatively easy to understand and, with the appropriate domain-knowledge, also allow for a causal notion (Zhao and Hastie 2021).

We exemplify these methods using a real world example: The COVID-19 pandemic in Germany. We use SOEP and SOEP-CoV data to predict the risk of working in short-time during the onset of the COVID-19 pandemic in Germany. For this, we rely on a random forest. We investigated the implied functional form, approximated by a random forest, with the help of the PD function and the H-statistic, which builds on the PD function.

One important aspect of these methods is that they are model agnostic. That means they can be applied to all statistical models, regardless of their degree of complexity. This makes them unusually flexible. Further, while our application focuses on a prediction case, one could also apply the methods to illustrate how various features are associated with treatment effects in the estimation of heterogeneous treatment effects, as in Wager and Athey (2018).

From a normative perspective, we find that immutable individual level characteristics, such as age gender, and migration background, do not matter for the assignment to short-time work. This suggests that the norm of equality of opportunity was not violated in general. However, this conjecture holds only after marginalizing out all other characteristics, such as job position and sector. This means that such disparities due to individual-level characteristics are still relevant in an unconditional observation. However, since those disparities are potentially mediated by sector affiliations or selection into different job positions, we do not argue that individual level characteristics, such as gender, do not matter. The disproportionate representation of certain groups in certain sectors is an equilibrium outcome that can be explained by forces outside of the COVID-19 pandemic.

References

Adams-Prassl, Abi, Teodora Boneva, Marta Golin, and Christopher Rauh. 2020. "Inequality in the impact of the coronavirus shock: Evidence from real time surveys." *Journal of Public economics* 189: 104245.

Breiman, L, and A Cutler. 2001. "Random forests." *Machine learning* 45: 5–31.

Breiman, Leo. 2001. "Random forests." *Machine learning* 45: 5–32.

Chetty, Raj, John N Friedman, Michael Stepner et al. 2020. "The economic impacts of COVID-19: Evidence from a new public database built using private sector data."Technical report, national Bureau of economic research.

Clark, Andrew E, Conchita D'Ambrosio, Niccolo Gentile, and Alexandre Tkatchenko. 2022. "What makes a satisfying life? Prediction and interpretation with machine-learning algorithms." *CEP Discussion Papers*.

Clark, Andrew E, Conchita d'Ambrosio, and Anthony Lepinteur. 2021. "The fall in income inequality during COVID-19 in four European countries." *The Journal of Economic Inequality* 19: 489–507.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of statistics* 29 (5): 1189–1232.

Friedman, Jerome H, and Bogdan E Popescu. 2008. "Predictive Learning via Rule Ensembles." *The annals of applied statistics* 2 (3): 916–954.

Goebel, J, M M Grabka, S Liebig, M Kroh, D Richter, C Schröder, and J Schupp. 2019. "The German socio-economic panel (SOEP)." *Journal of Economics and Statistics* 239 (2): 345–360.

Haddouchi, Maissae, and Abdelaziz Berrado. 2019. "A survey of methods and tools used for interpreting random forest." In *2019 1st International Conference on Smart Systems and Data Science (ICSSD),*: 1–6, IEEE.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning.*: Springer New York.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human decisions and machine predictions." *The Quarterly Journal of Economics* 133 (1): 237–293.

Kühne, Simon, Martin Kroh, Stefan Liebig, and Sabine Zinn. 2020. "The need for household panel surveys in times of crisis: The case of SOEP-CoV." *Survey Research Methods* 14 (2): 195–203.

Li, Jinjing, Yogi Vidyattama, Hai Anh La, Riyana Miranti, and Denisa M Sologon. 2020. "The impact of COVID-19 and policy responses on Australian income distribution and poverty." *arXiv preprint arXiv:2009.04037*.

Molnar, Christoph. 2022. *Interpretable Machine Learning: A Guide For Making Black Box Models Explainable.*: Independently published.

Mühlbauer, Sabrina, and Enzo Weber. 2022. "Machine Learning for Labour Market Matching." *IAB-Discussion Paper* 03/2022.

Mullainathan, Sendhil, and Ziad Obermeyer. 2022. "Diagnosing physician error: A machine learning approach to low-value health care." *The Quarterly Journal of Economics* 137 (2): 679–727.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine learning: an applied econometric approach." *Journal of Economic Perspectives* 31 (2): 87–106.

Nembrini, Stefano, Inke R König, and Marvin N Wright. 2018. "The revival of the Gini importance?." *Bioinformatics (Oxford, England)* 34 (21): 3711–3718.

O'Donoghue, Cathal, Denisa M Sologon, Iryna Kyzyma, and John McHale. 2020. "Modelling the distributional impact of the COVID-19 crisis." *Fiscal Studies* 41 (2): 321–336.

Oparina, Ekaterina, Caspar Kaiser, Niccolò Gentile, Alexandre Tkatchenko, Andrew E Clark, Jan-Emmanuel De Neve, and Conchita D'ambrosio. "Machine Learning in the Prediction of Human Wellbeing."

Sandri, Marco, and Paola Zuccolotto. 2008. "A Bias Correction Algorithm for the Gini Variable Importance Measure in Classification Trees." *Journal of computational and graphical statistics: a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America* 17 (3): 611–628.

Stantcheva, Stefanie. 2022. "Inequalities in the Times of a Pandemic." *Economic Policy* 37 (109): 5–41.

Wager, S, and S Athey. 2018. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association* 113 (523): 1228–1242.

Zhao, Qingyuan, and Trevor Hastie. 2021. "Causal Interpretations of Black-Box Models." *Journal of business & economic statistics: a publication of the American Statistical Association* 39 (1): 272–281.

Appendix A. Additional tables.

TABLE A1. Description of NACE Industry Sectors.

Number	Sector description
1	Agriculture, forestry and fishing
2	Mining and quarrying*
3	Manufacturing
4	Electricity, gas, and water supply
5	Repair services
6	Construction
7	Wholesale and retail
8	Transportation and storage
9	Accommodation and food service activities
10	Information and communication
11	Financial and insurance activities
12	Real estate activities
13	Professional, scientific, techn. activities
14	Administrative and support service activities
15	Public administration and defence
16	Education
17	Human health and social work activities
18	Arts, entertainment and recreation
19	Other service activities
20	Domestic work
21	Extraterrestrial organizations*

* This sector has been omitted from the analysis because there are too few observations.

TABLE A2. Confusion matrices

	Opt. cut-point	True positive	False negative	False positive	True negative
Random forest	0.223	74	36	157	417
Logit	0.200	67	43	155	419

Note: Table A2 displays the optimal cut-point to classify individuals in the test data based on the predictions. The optimal cut-point is such, that it maximizes the sum of sensitivity and specificity. True positives are individuals which are predicted to be in short-time work and worked in short-time. False negatives are individuals which are predicted not to work in short-time work but worked in short-time. False positives are individuals that are predicted to work in short time but did not work in short-time. True negatives are individuals that are predicted not to work in short-time and did not work in short-time.

TABLE A3. Optimal cut-point

	Opt. cut-point	Sensitivity	Specificity
Random forest	0.223	0.673	0.727
Logit	0.200	0.609	0.730

Note: Table A3 displays the optimal cut-point, which maximizes the sum of sensitivity and specificity, to classify the predictions based on the random forest and Logit model as well as the associated sensitivity and specificity.