Understanding the Heterogeneity of Earnings Losses After Job Displacement: A Machine-Learning Approach*

Andreas Gulyas† Krzysztof Pytka‡

Current version:
April 2025
First version:
October 2019

Abstract

Using generalized random forests in a difference-in-differences framework, we analyze heterogeneity in earnings losses following job displacement. We find substantial variation in short-term losses ranging 20-70%. While all workers face long-term losses, employment rather than wage changes primarily drives heterogeneity. The most important predictor for employment losses is worker's age, while the firm wage premium is the most important one for wage losses. We predict similar loss patterns for the population excluded by the typical sample restrictions in the literature. We further show how to create simple decision rules to target active labor market programs to high-loss individuals.

Keywords: Job displacement, Earnings losses, Causal machine learning

*A previous version was circulated under the name of "Understanding the Sources of Earnings Losses After Job Displacement: A Machine-Learning Approach." Funding by the German Research Foundation (DFG) through CRC TR 224 (Project A3) and support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1134-1 FUGG is gratefully acknowledged. Johannes Gönsch, Michaela Paffenholz, Tim Bartelsen, and László Kavalecz provided excellent research assistance. We thank the editor, two anonymous referees, Clara Santamaría, Jan Stuhler, Stefan Wager and seminar participants at various conferences and universities for helpful comments. All remaining errors are ours.

†University of Mannheim; andreas.gulyas@uni-mannheim.de

‡University of Mannheim; pytka@uni-mannheim.de

I. Introduction

A sizable literature documents that, on average, workers displaced during a mass layoff experience significant losses in annual income lasting over 15 to 20 years.[1] Many labor market policies are designed to avoid job losses in general. Prominent examples include firm bailouts, employment protection, and employment subsidies such as short time work schemes, which are often applied in a non-discriminatory fashion. Critics argue that not all jobs are worth saving, and such programs should be, if anything, targeted. To design effective policy responses, we not only need to identify the consequences of job loss for the average worker, but also understand how earnings losses differ across individuals, and which pre-displacement characteristics lead to higher losses. While many papers show that earnings losses differ by certain worker or firm characteristics, this paper systematically documents the heterogeneity of short- and long-term cost of job loss across the population and identifies the most important predictors for losses.

To this end, we draw on recent advancements in machine learning to estimate heterogeneous treatment effects. To document the whole distribution of earnings, employment, and wage losses across different workers, we implement a generalized random forest by Athey, Tibshirani, and Wager (2019) to a difference-in-difference (hereafter DiD) setup. Furthermore, we study how post-displacement employment histories differ across high- and low-loss individuals and discuss which pre-displacement characteristics are the most important predictors for the cost of job loss. We then show how our results could be used to design simple decision rules for targeting displaced workers predicted to face especially severe employment and wage losses.

Our approach estimates the causal cost of job loss nonparametrically as a function of observables. This is achieved by exploring possible data splits and choosing those ones for which between-group differences in losses are the highest. By recursively splitting the dataset into smaller and smaller subsamples, the algorithm builds a tree which detects heterogeneity in the difference-in-difference estimator of the cost of job loss. Thus, we need to assume that the outcome variables for the treatment group absence of job loss would have evolved in parallel to the control group for every subsample. We verify that this assumption is plausible with several diagnostic checks. To avoid detection of artificial heterogeneity, instead of

[1] Jacobson et al. (1993), Neal (1995), Couch and Placzek (2010), Davis and Von Wachter (2011), Farber (2011), Farber (2017), among many others. Furthermore, job displacements have been shown to have detrimental effects on health (Schaller and Stevens, 2015), longevity (Sullivan and Von Wachter, 2009), and on children of displaced workers (Lindo, 2011; Rege et al., 2011).

growing a single tree, we build a random forest consisting of many tree-based models. Every tree is trained using a random subset of observations and variables and then their estimates are combined.[2] Another appealing property of random forests is that we are able to quantify standard errors arising from two sources, the machine-learning procedure and the estimation procedure.

We use the universe of Austrian social security records over three decades to examine the heterogeneity in short- and long-term earnings, employment, and wage losses. We apply the typical identification strategy and focus on mass-layoffs. Displaced Austrian workers suffer on average losses of similar magnitude compared to those displaced in Germany (Schmieder et al., 2023) and the United States (Davis and Von Wachter, 2011).

Our algorithm documents losses conditional on workers' and job characteristics. For this reason, we construct 58 variables and feed them into our learning procedure. We seek to understand how earnings losses vary with pre-displacement worker, firm, time, and local labor market characteristics and to identify which of these are the most important predictors for losses.

To comprehensively study the heterogeneity in the consequence of job losses, we build separate forests for short- (one year after job loss) and long-term (10 years after job loss) losses in relative earnings, employment, and log-wages. We document that the consequences of job losses are far from uniform. While almost no worker is predicted to have losses lower than 20% in the year after the job loss, the worst affected workers face losses of more than 70%. The 25th percentile of losses is 42%, while the 75th is 29%. This heterogeneity is not driven by noisy estimates of the individual treatment effect. At the 95% confidence level, we estimate that 67% of workers face significantly different short-run earnings losses than the median worker with 35%. 10 years after job displacement, the whole distribution of long-run losses shifts towards lower losses, and its dispersion contracts markedly. 25% of workers face losses of more than 19% in the long-run and the interquartile range decreases to 6 percentage points (pp). In our sample, 99% of all workers are predicted to experience losses in the long-run. Thus, in line with recent evidence from the United States (Rose and Shem-Tov, 2023), job loss is costly even for low wage workers with disadvantaged worker and firm characteristics. Although there is some convergence of losses across workers, the correlation of short and long-term losses at the individual level is fairly high with 0.4.

[2] This statement is for expositional reasons and might not be precise enough. Strictly speaking, for an individual characterized by observables $\mathbf{z}$, the random forest provides weights $\alpha_i(\mathbf{z})$ measuring similarity of all other observations indexed by i. Those weights are used in weighted linear regression to estimate the causal effect at $\mathbf{z}$. The whole algorithm is presented in Subsection IV.A in greater detail.

Earnings losses are a result of both forgone income due to non-employment and declines in wages after reemployment. Overall earnings losses are more strongly correlated with employment than with wage losses, indicating that the dispersion in employment losses is a somewhat more important driver of the heterogeneity in overall losses. The 25th and 75th percentile of predictable short-run employment losses are 125 and 85 days. In the long-run, only 25% of workers are predicted to lose more than 37 employment days per year, the interquartile-range narrows significantly to 16 days. Wage losses in the short run range from 35% to even wage gains. Interestingly, the distribution of wage losses in the short and long-run are very similar, suggesting that wage losses are very persistent. This is confirmed by the high correlation of close to 0.8 for short and long-term wage losses.

How do labor market histories after job loss differ across high- and low-earnings-loss individuals? Those individuals identified to face the highest losses have the lowest propensity to remain in their original industry, and move to more unstable firms, and subsequently have more job changes. They also suffer the largest declines in employer wage premia. Interestingly, we find that across the earnings-loss distribution, declines in employer wage premia explain a roughly constant 60-70% of wage losses.

What pre-displacement worker and job characteristics are the most important predictors for the severity of losses? We compute how predicted earnings, employment, and wage losses change from the lowest to the highest realization of the covariates, holding every other observable variable constant at the median. The variable that predicts the highest change in short-term losses is age, the oldest workers face losses of 50%, whereas the youngest workers with the same worker and firm characteristics face losses of 30%. Firm size and the firm wage premia are the second and third most important predictors of short-term earnings losses. For long-run earnings losses, the firm's labor-market share and firm separation rate prior to the mass-layoff event are the first and second most important variables, followed by workers' age. *Ceteris paribus*, workers from firms with the largest labor market share or most stable employers face 7-8pp higher earnings losses 10 years after job loss, highlighting the special role of job losses at regional flag-ship firms (Gathmann et al., 2020) and particularly stable jobs (Jarosch, 2023). As employment losses are strongly correlated with overall earnings losses, the variable importance is similar for employment losses.

Recently, the role of firm wage premia for the cost of job loss moved in the center of attention in the literature. Fackler, Mueller, and Stegmaier (2021), Schmieder, Von Wachter, and Heining (2023), and Bertheau, Acabbi, Barceló, Gulyas, Lombardi, and Saggio (2023) find that losses in firm wage premia explain a large fraction of wage losses, whereas Lachowska, Mas, and Woodbury (2020) find limited role of firm wage premia. We can shed light on this discussion from a new angle: Is the pre-displacement firm wage premium an important predictor for subsequent earnings, employment, and wage losses? Although the firm wage premium is not among the most important predictors for employment losses, it is the most important variable for short and long-run wage losses. Workers being displaced from the highest paying decile of firms face short-term losses of 18pp and long-run losses of 9pp more than the employees of the lowest paying firms. Thus, for each log point higher pre-displacement firm fixed effects, long-run wage losses increase by roughly 0.15 log points.

Interestingly, many variables that the prior literature has shown to matter for earnings losses such as business cycles (Davis and Von Wachter, 2011; Schmieder et al., 2023), manufacturing (Helm et al., 2023), job tenure (Jacobson et al., 1993; Neal, 1995) do not show up as important predictors of losses in our setting.

In principal our results could be directly used to target high-loss individuals with active labor-market programs. A disadvantage is that we use variables that are constructed using statistical measures (*e.g.*, the fixed effects) which may not be readily understandable by the general public and conveniently observable by policy makers. To address this, we use comprehensible variables to build a transparent and clearly communicable decision tree to target high employment- and wage-loss individuals. They correctly identify approximately 75-77% of individuals experiencing above-median losses.

We compare our generalized random forest approach with alternatives. LASSO-penalized DiD regression shows greater variation in treatment effects but identifies fewer individuals with statistically significant differences from the median (21% compared to our forest's 67%). Quantile-based approaches struggle with individual-level effects due to poor correlation between pre- and post-displacement earnings ranks and a simple comparison of control and treatment groups does not give a sense of statistical accuracy and does not identify the most important predictors. The comparison highlights that our forest-based approach is the best for analyzing predictable individual losses.

In addition to previously mentioned papers, we also contribute to the small but growing number of papers using machine learning to study heterogeneous treatment effects in labor economics.[3] Gregory, Menzio, and Wiczer (2025) use a k-means algorithm to classify workers into three clusters based on labor market histories and document the heterogeneity in earnings losses for these three groups. We document the whole distribution of losses and identify pre-displacement worker and job characteristics that are most important predictors

[3] *E.g.,* Davis and Heller (2020); Knaus et al. (2022).

of earnings losses. In a recent working paper, Athey et al. (2023) use a similar machine-learning algorithm to ours to study heterogeneity in earnings losses in Sweden. They also find that worker's age is the most important predictor for short-term earnings losses. We go beyond their study and examine both short and long-term earnings losses as well as log-wage losses. We show that the most important predictors for long-term losses shift from worker's age towards firm characteristics. Firm wage premium, not present in their study, is by far the most important predictor for wage losses. In addition, we leverage our forest to address a long-standing concern in the literature: Do the strict sample restrictions for a clean identification of the displacement event select individuals that are bound to experience large earnings losses? We document that the distributions of predicted long-term earnings losses for the general population and for displaced workers in our sample are surprisingly similar.

The rest of the paper is organized as follows. The next section describes the empirical setting in Austria, as well as the sample selection. Section III presents the average cost of job displacement. Section IV describes our machine-learning algorithm, and Section V documents the heterogeneous scarring effects of job displacement and discusses their most important predictors. Section VI demonstrates how our results can be translated into practical targeting strategies using transparent decision rules, while Section VII juxtaposes our generalized random forest approach with alternative methods for analyzing treatment effect heterogeneity. The last section concludes.

II. Empirical Setting

We use the Austrian social security records from 1984 through 2020, which comprise day-to-day information on all jobs and unemployment spells covered by social security (Zweimüller et al., 2009). It contains all private sector jobs and a large fraction of public sector employment, but excludes self-employed and public servants who are not covered by social security. It contains information on yearly earnings for each worker-establishment pair, and basic socio-demographic information at the worker level such as age, gender, a flag for a blue collar occupation, and citizenship.[4] Each establishment (we use firm and establishment exchangeably from here on) has a unique identifier, we have information on its geographic location, age, and 4-digit industry classifier.

[4] We deflate all earnings to 2017 level using the CPI index provided by the Austrian Statistical Agency.

A. Definition of Job Displacement and Mass Layoff

To ensure comparability with the previous literature on displaced workers, we follow the typically applied definitions and sample restrictions as much as possible. Workers are considered displaced if they separated from their primary employer that experienced a mass layoff in the given year. We define a mass layoff event at the firm level in year t if the size of the firm declined by more than 30% during year t. To avoid selecting volatile firms, we exclude firms that grew by more than 30% in either $t-1$, or $t-2$, as well as firms that are larger 3 years after the event than before. To have a meaningful measure of firm growth, we only consider establishments with at least 30 employees. To avoid misspecifying mergers, outsourcing, or firm restructures as mass-layoffs, we compute a worker cross flow matrix for all firms in each year. We exclude all firms where more than 20% of their workforce ends up working for the same employer in $t+1$.[5] Thereby we exclude mass layoff firms with large worker flows to other firms. Not correcting for these potential measurement errors might lead to a significant underestimation of earnings losses. We also disregard mass layoff events from the public administration (Nace, Level 1, Code O).

B. Sample Construction

We depart from most of the earnings loss literature and do *not* restrict our sample to males only. Specifically, we are interested in how earnings losses vary across different individuals based on their characteristics, including gender. We proceed by selecting everybody who is employed on the reference day of January 1st each year between 1989 and 2010. We follow the literature and restrict our sample to workers aged 24-50, employed at a firm larger than 30 employees, and with job tenure longer than 3 years on the reference day.[6] We further restrict our sample to individuals who have annual earnings of at least €7,500 for two consecutive years to exclude workers working only a few hours per week in so called mini-jobs. This results in 11,804,786 person-year observations. A common critique is that these sample restrictions are selecting workers that are bound to face high earnings losses. We will use our machine-learning algorithm to make out-of-sample predictions on earnings losses for the general population to study whether these restrictions indeed select worker and

[5] For an in-depth discussion, see Hethey-Maier and Schmieder (2013).

[6] We only consider workers aged below 50 because of the low legal and effective retirement age in Austria. Since we follow workers for 10 years after their job loss, and workers retire on average slightly before the age of 60, we have chosen age 50 to minimize confounding earnings losses with early retirement of both the control and treatment group.

job characteristics that are associated with higher losses.[7]

Out of these remaining observations, we define a person to be displaced in t if a worker separates from a firm experiencing a mass layoff, and the worker is not reemployed at the same firm at any point in the next 10 years. If a worker suffers multiple mass layoffs, we only consider the first one. We identify 50,207 displaced worker events between 1989 and 2010.

Some workers disappear from our dataset over time. This happens on the one hand because workers might not find employment subject to social security insurance anymore and drop out of the labor force. On the other hand, this could also happen if workers move into self-employment or move abroad. For those years without any social security spell, we impute zero earnings.[8]

Note that we do not restrict the control group to have stayed at their employer after t. The potential comparison group consists of non-displaced workers subject to the same sample restrictions. This includes workers employed at firms without any mass layoff event during year t or workers in mass layoff firms who did not separate.

Non-displaced workers may differ in many characteristics from the displaced workers, as can be seen in Table 1. Following many papers in the literature, e.g. Schmieder et al. (2023); Bertheau et al. (2023); Halla et al. (2020), we use propensity score matching in order to obtain a control group that is as similar as possible to displaced workers. In each year, for all workers satisfying the sample restrictions, we estimate the propensity to experience a displacement event as a function of the following worker and firm characteristics: worker's log-wage in year $t-1$ to $t-3$, tenure, age, establishment size in year $t-1$, a dummy for working in the production sector, a female indicator, a dummy for having a full year of non-employment or no social security spells in the last 5 years, and the number of non-employment days in the last 5 years.[9] For each displaced worker in a given year, we select the non-displaced worker with the nearest propensity score without replacement. Table 1 shows that our matched control group is very similar to displaced workers in observable characteristics. The two groups are virtually indistinguishable in terms of pre-displacement

[7]Surprisingly, the distribution of losses for mass-layoff workers and the general population is comparable, as discussed in more detail later.

[8]In an earlier working paper version, instead of imputing zero earnings, we dropped any person-year observation without any employment and unemployment spell from our study following Schmieder et al. (2023). With the imputation of zeros, employment losses play a bigger role, and variable importance shifts towards variables more important for wage rather than employment losses. See Gulyas and Pytka (2022) for the detailed results without imputation.

[9]We also experimented with different sets of matching variables, all of which lead to similar results.

	Displaced	Selected Control Group	Not Selected
Age	38.4	38.32	38.46
$\log(\mathrm{w}_t - 1)$	4.49	4.5	4.61
$\log(\mathrm{w}_t - 2)$	4.47	4.48	4.59
$\log(\mathrm{w}_t - 3)$	4.45	4.46	4.56
Job Tenure (in days)	2770.23	2749.75	2916.08
Firm Size	222.4	213.64	825.47
Manufacturing	0.56	0.56	0.46
Female	0.41	0.41	0.37
Whole Yr. Non-Emp.	0.02	0.02	0.02
Days Non-Emp.	44.36	43.51	39.06
No SS Spell	0.01	0.01	0.02
Obs	50207	50207	11704372

Table 1: Sample characteristics of displaced workers compared to the selected control group via propensity score matching and the universe of worker/year observations satisfying sample restrictions

evolution of earnings, days employed, and log-wages, as can be seen in Figure 9 in the appendix, which plots raw averages across these groups over time.

C. Outcome Variables

We study the cost of job loss in terms of three dimensions. First, we estimate how yearly earnings change relative to the average earnings in the two years before job displacement. Second, we measure employment losses as the change in days employed covered by social security. Third, we study how workers' log daily wage at their dominant employer evolves after job loss conditional on re-employment.[10] The details for the estimation are found in Appendix A.

III. The Average Cost of Job Displacement

Throughout the paper, we are interested in identifying how the cost of job loss differs across individuals. We nevertheless start by discussing the estimation strategy for the homogeneous treatment case, which will be extended to heterogeneous treatment effects in the next section.

[10] We compute daily wages by dividing yearly income from the employer with the highest earnings in a given year and divide it by the number of employment days at this employer.

Table 2: DiD Regression

	Earnings		Employment		Log wage	
	(1)	(2)	(3)	(4)	(5)	(6)
τ	−0.361	−0.361	−106.054	−106.054	−0.140	−0.136
	(0.002)	(0.002)	(0.743)	(0.743)	(0.002)	(0.002)
Worker FE	No	Yes	No	Yes	No	Yes
Num. obs.	602484	602484	602484	602484	587727	587727
R^2	0.212	0.385	0.235	0.396	0.012	0.917
Adj. R^2	0.212	0.262	0.235	0.275	0.012	0.900

Table shows estimation results using equation (1) for horizon h=1 with using different controls. See text for details.

The average cost of job loss can be compactly estimated using the following Event-study:

$$\begin{aligned} y_{it} \quad &= \quad \tau_h D_i \times \mathbb{1}(t = t^* + h) + \nu_h \mathbb{1}(t = t^* + h) + \theta D_i + \gamma \\ &\quad + \sum_{j=-5}^{-2} \nu_j \mathbb{1}(t = t^* + j) + \sum_{j=-5}^{-2} \delta_j \mathbb{1}(t = t^* + j) \times D_i + \epsilon_{it}, \end{aligned} \tag{1}$$

where D_i is an indicator equal to one for a displaced person, t^* is the displacement year and h is the horizon of interest. The coefficients ν_j measure the evolution of the left-hand side variable over time for the control group, whereas θ absorbs initial differences in labor market outcomes between the control and treatment group. δ_j measures how the left-hand side variable evolved for displaced workers before the displacement event relative to the control group, whereas τ_h measures the cost of job loss h years after job displacement, relative to $t-1$. Throughout our study, we will focus on short-term ($h = 1$) and long-term losses ($h = 10$). The regression does not feature a year fixed effect, nor a control for age as is typical in the literature, because our machine-learning algorithm will flexibly estimate earnings losses by displacement year and by the age of the workers. Table 2 shows the estimates from this regression for different specifications. For each outcome variable the table reports the cost of job loss using Equation (1) with $h = 1$ without any controls and with worker fixed effects.

Because of the computational burden of the machine learning algorithm, we will not be able to include worker fixed effects as controls as is often done in the literature. But the results from Table 2 show that because our matching procedure selected very similar workers as a control group, adding worker fixed effects does not significantly change the estimated costs of job displacement.

IV. Heterogeneity in the Cost of Job Loss – Machine Learning Approach

The goal of our exercise is to identify the heterogeneity in earnings losses and its most important predictors. To this end, we employ a machine learning procedure built on the methodology of generalized random forests, recently developed by Athey et al. (2019). The average earnings losses after job displacement can be estimated directly from Equation (1) using standard statistical methods. Yet, this approach has some serious limitations. For instance, this type of estimate provides the average treatment effect and does not provide information on its underlying heterogeneity. If the earnings losses are not uniform across individuals, then the individual scarring effects may be a far cry from the estimated average.

Conceptually, our approach consists in estimating the version of Equation (1) with heterogeneous scarring effects:

$$\begin{aligned} y_{it} &= \tau_h(\mathbf{z}_i) D_i \times \mathbb{1}(t = t^* + h) + \nu_h(\mathbf{z}_i)\mathbb{1}(t = t^* + h) + \theta(\mathbf{z}_i) D_i + \gamma(\mathbf{z}_i) \\ &\quad + \sum_{j=-5}^{-2} \nu_j(\mathbf{z}_i)\mathbb{1}(t = t^* + j) + \sum_{j=-5}^{-2} \delta_j(\mathbf{z}_i)\mathbb{1}(t = t^* + j) \times D_i + \epsilon_{it}, \end{aligned} \tag{2}$$

where $\mathbf{z}_i$ are the values of observable variables (henceforth called *partitioning variables*) for individual i. $\tau_h(\mathbf{z}_i)$ is the function that describes how the cost of job loss changes with individual worker and job characteristics $\mathbf{z}_i$. The functional specification of $\tau(\mathbf{z}_i)$ is assumed to be unknown. To uncover the true form of $\tau(\mathbf{z}_i)$, we employ a generalized random forest by Athey et al. (2019), adapted to a DiD setting.

We begin by describing the machine learning algorithm used to detect heterogeneity in treatment effects. We then discuss the partitioning variables that condition the cost of displacement.

A. Bird's-Eye View of Machine Learning Algorithm

We now describe the implemented method of a generalized random forest. We seek to estimate the cost of job loss locally at worker and job characteristics $\mathbf{z}$ from equation (2), which is characterized by following local moment conditions:

$$\mathbb{E}(\mathbf{x}'_{it}\varepsilon_{it}|\mathbf{z}) = \mathbf{0}_{12}, \tag{3}$$

where $\mathbf{x}'_{it} = [\mathbb{1}_{\{t=t^*+h\}}D_i,\ \mathbb{1}_{\{t=t^*+h\}},\ D_i,\ 1,\ \mathbb{1}_{\{t=t^*-5\}}, \ldots, \mathbb{1}_{\{t=t^*+h\}},\ \mathbb{1}_{\{t=t^*-5\}}D_i, \ldots, \mathbb{1}_{\{t=t^*+h\}}D_i]$, ε_{it} is the error term from (2), and $\mathbf{0}_{12}$ is a row vector with zeros of length 12. Our approach consists of defining similarity weights $\alpha_{it}(\mathbf{z})$, which measure the relevance of observation it to estimating the cost of job loss at $\mathbf{z}$, and estimating equation

$$\begin{aligned}
(\tau(\mathbf{z}), \theta(\mathbf{z}), \boldsymbol{\gamma}(\mathbf{z})) \quad = \quad & \operatorname{argmin}_{\{\tau,\theta,\gamma\}} \left(\frac{1}{NT}\sum_{i}^{N}\sum_{t}^{T}\alpha_{it}(\mathbf{z})\mathbf{x}'_{it}u_{it}\right)\left(\frac{1}{NT}\sum_{i}^{N}\sum_{t}^{T}\alpha_{it}(\mathbf{z})\mathbf{x}'_{it}u_{it}\right)', \\
\text{s.t.} \quad & \forall_{i,t} u_{it} = y_{it} - \Bigg(\tau_h(\mathbf{z}_i)D_i \times \mathbb{1}(t = t^* + h) \\
& +\nu_h(\mathbf{z}_i)\mathbb{1}(t = t^* + h) + \theta(\mathbf{z}_i)D_i + \gamma(\mathbf{z}_i) \\
& + \sum_{j=-5}^{-2}\nu_j(\mathbf{z}_i)\mathbb{1}(t = t^* + j) + \sum_{j=-5}^{-2}\delta_j\mathbb{1}(t = t^* + j) \times D_i\Bigg)
\end{aligned} \tag{4}$$

Notice that the problem (4) takes weights $\alpha_{it}(\mathbf{z})$ as given and is solved for each value of partitioning variables $\mathbf{z}$ separately. For constructing these weights, a generalized random forest is used. For exposition purposes, the algorithm is presented in three steps. First, we show how a single tree in the spirit of Breiman et al. (1984) with a modified splitting criterion borrowed from Athey et al. (2019) is grown. Next, the approach is augmented to generalized random forests. Finally, we present the detailed numerical implementation and how the weights can be recovered from our random forest.

A.1. Tree Construction

The tree-based procedure consists in partitioning the dataset into smaller subsamples in which individuals exhibit similar earnings losses and at the same time the differences in earnings losses between subsamples are maximized. The data fragmentation is carried out using a sequence of complementary restrictions on partitioning variables. Due to the computational complexity, a top-down, greedy approach is traditionally used. The procedure of building a tree can be characterized in a recursive way by Algorithm (1). In each data partition (called also a node or a leaf) the treatment effect is estimated from Equation (2) separately.[11]

Algorithm 1 Tree Algorithm of Recursive Partitioning

i. Start with the whole dataset and consider it as one large data partition, $\mathcal{P}$.

ii. For each partitioning variable z_k and its every occurring value $\overline{z}$, split partition $\mathcal{P}$ into two complementary sets of individuals i such that $\mathcal{P}_l = \{i \in \mathcal{P} : z_{ki} \leq \overline{z}\}$ and $\mathcal{P}_r = \mathcal{P} \setminus \mathcal{P}_l$ and estimate cumulative earnings losses τ_l and τ_r for both partitions by running two separate regressions of form (2) on $\mathcal{P}_l$ and $\mathcal{P}_r$.

iii. Choose the variable z_k and value $\overline{z}$ that maximizes:

$$(\tau_l - \tau_r)^2 \frac{n_l \cdot n_r}{N^2}, \tag{5}$$

where n_l and n_r are sizes of $\mathcal{P}_l$ and $\mathcal{P}_r$ and N is the sample size of $\mathcal{P}$.

iv. If (5) is smaller than a tolerance improvement threshold, then stop. Otherwise, go to step (ii) and repeat the splitting procedure for $\mathcal{P}_l$ and $\mathcal{P}_r$ separately, where $\mathcal{P}_l$ and $\mathcal{P}_r$ are new partitions subject to the splitting procedure, $\mathcal{P}$.

The main difference of our procedure from the textbook CART algorithm Breiman et al. (1984), is the splitting criterion. In the original approach, the algorithm aims at building a tree minimizing the squared sum of residuals.[12] In our setup, we are interested in growing a tree that explores the underlying heterogeneity of earnings losses between partitions of individuals with different characteristics. For this reason, we adapt the criterion (5) proposed by Athey et al. (2019). This criterion maximizes between-group differences of earnings losses, $(\tau_l - \tau_r)^2$, with an adjustment for more balanced splits, $\frac{n_l \cdot n_r}{N^2}$.

In applied economics, one alternative to splitting the dataset is to assume that the data-generating process is known and given and to estimate according to that process. In our application this would mean that we make an arbitrary decision upon the specification of $\tau(\mathbf{z}_i)$. However, in our strategy we are upfront about our agnosticism on $\tau(\mathbf{z}_i)$ and employ Algorithm (1) to learn the true specification. As a result, the learning procedure does three things: (i) chooses which variables are important and contribute to accounting for the heterogeneous scarring effects and which do not; (ii) detects non-linear relationships between τ and $\mathbf{z}_i$; (iii) detects interactions (including interactions of higher orders) between partitioning variables.

While the details on the chosen partitioning variables are presented more thoroughly later

[11] It is noteworthy that while all parameters from (2) are estimated, only the parameter of our interest, the scarring effect, is used in the splitting criterion (5).

[12] In regression trees, the squared sum of residuals is defined as $\sum_j \sum_{i \in \mathcal{D}_j} (y_i - \overline{y}_{\mathcal{D}_j})^2$, where $\mathcal{D}_j$ is a data subset obtained through sample partitioning procedure and $\overline{y}_{\mathcal{D}_j} = \frac{1}{|\mathcal{D}_j|} \sum_{i \in \mathcal{D}_j} y_i$ is the mean of the response variable in the specific set of data.

in Subsection B, we provide here the outcome of our tree-growing algorithm to illustrate its mechanics and build intuition about how heterogeneous treatment effects are identified. A more thorough discussion of how different variables correlate with different levels of losses, using predictions from our random forests, will be presented in Section V. Figure 1 illustrates a decision tree capturing heterogeneity in earnings losses, expressed as proportions of pre-displacement income. Each node displays the average earnings loss and the percentage of observations it contains. The tree starts with a root node encompassing the entire sample and terminates in leaf nodes representing distinct groups. In the overall sample, the average earnings loss is 36.1% of prior income which is exactly the same as reported in Table 2. The primary split occurs based on market share, separating firms with lower labor market share from those with higher market share. The tree subsequently branches using various characteristics including worker age, market share, manufacturing status, firm wage premia, and firm size. The final tree structure contains 10 leaf nodes, with each subgroup defined by 2 to 5 binary conditions. There is substantial heterogeneity in earnings losses across these groups. The most severely impacted workers, found among employees displaced from large firms with high labor market share, experience losses of 54.4% of their pre-displacement income. In contrast, the least affected group faces losses of only 18.4%, comprising workers younger than 41 displaced from firms with the lowest labor market share. Overall, the tree highlights substantial heterogeneity in displacement effects, systematically partitioning the sample to maximize the between-group variation in treatment effects while maintaining sufficient observations within each leaf.

A.2. Generalization to Random Forests

One important advantage of tree-based models is their easy and very intuitive graphical illustration. Unfortunately, it is well-known that estimates can be non-robust and it is difficult to properly estimate their standard errors. Random forests proposed by Breiman (2001) are a refinement of the baseline method that addresses the typical concerns of tree-based models. The general idea behind random forests is quite straightforward: build many trees through bootstrapping data observations. Moreover, in each split decision a subsample of considered variables is drawn. Consequently, an ensemble of decorrelated trees is grown, which means that the trees differ from each other and are built with different variables. This way only relationships that consistently show up in different bootstrapped samples are identified. In addition to this, each tree has been built using a so-called “honest” approach. This means that half of the bootstrapped sample was used to determine conditions which

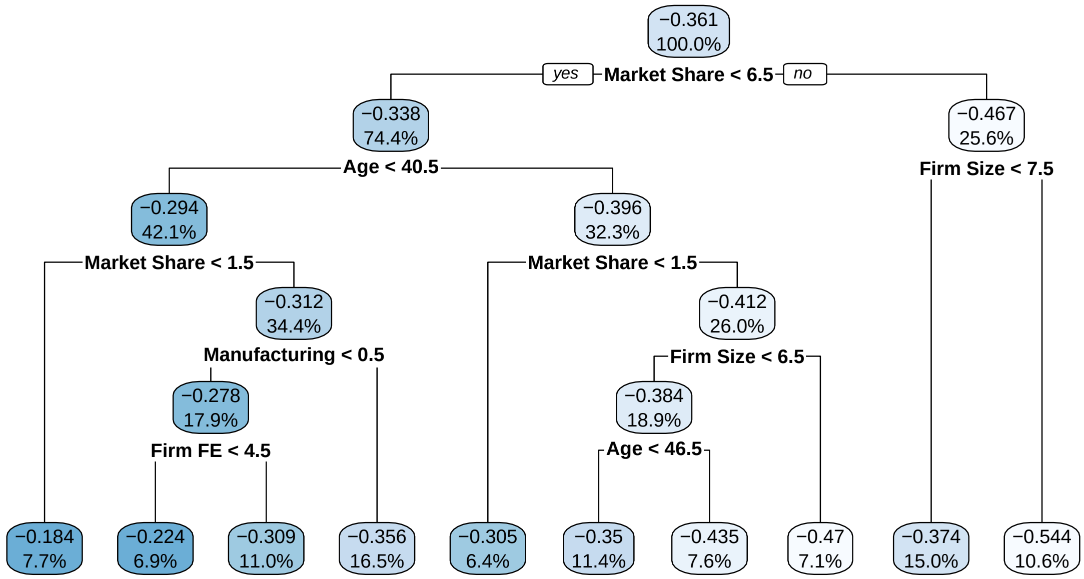


Figure 1: Heterogeneous treatment effect of job displacement. Tree was built with a CART algorithm maximizing between-group heterogeneity as described in the text. Minimum node size of 35,000 person-year observations.

constitute data partitions, while the other half was used to estimate the cost of job loss.[13] With the forest at hand, we can proceed with the construction of weights. Intuitively, the weights measure how often individuals fall into the same partitions across all trees. Then, those weights are used to solve (4) for each value of partitioning variables $\mathbf{z}$ separately.

Forests' specification. To analyze heterogeneity in post-displacement outcomes, we study three separate variables: relative earnings losses, log-wage losses, and employment losses. Each outcome is examined from both a short-term (one year after displacement) and long-term (ten years after displacement) perspective, resulting in six distinct dimensions of heterogeneity. For each case, we grow a forest consisting of 10,000 trees, with a minimum leaf size of 800 person-year observations. In each split decision, 7 partitioning variables are randomly selected.

An advantage of random forests over single trees is that standard errors can be computed relatively easily. We follow the algorithm proposed by Athey et al. (2019) and use the

[13] Thanks to this procedure we make sure we do not document spurious heterogeneity. If by any chance some splits are made due to some outliers, the estimated treatment effects are not affected by this.

bootstrap of little bags method introduced by Sexton and Laake (2009). In this approach, trees are randomly assigned to different "bags," allowing them to be reused multiple times to estimate variability without the need to grow new forests from scratch. This method is significantly more efficient than traditional bootstrap in our context, where growing even a single forest requires substantial computational time.[14] Our bootstrap procedure samples workers; thus, standard errors are clustered at the worker level.

The structure of our random forest allows us to estimate individual cost of job loss conditional on observed characteristics, with a key advantage over standard subgroup analysis: we can examine how losses vary with respect to one variable while holding others constant. In estimating these conditional earnings losses, we focus on variables exhibiting systematic differences, as our forest is regularized to filter out spurious variation driven by noise rather than true effects. While our method identifies variation only along observable characteristics—meaning we cannot detect heterogeneity driven by unobserved factors, which limits our ability to speak to job loss heterogeneity more broadly—the analysis remains highly valuable, particularly for informing policy targeting where eligibility criteria typically depend on observable characteristics.

A potential concern when combining a generalized random forest with difference-in-differences is that the method requires parallel trends not just to hold on average, but within each forest split. If certain covariates correlate with systematic differences in pre-treatment outcome trends, the algorithm might split on these covariates and generate spurious heterogeneity, misinterpreting pre-existing trajectory differences as heterogeneous treatment effects. To address this concern, we conducted several diagnostic checks.

First, we examined the relationship between pre-treatment trends and estimated losses. Using the year immediately prior to displacement ($t = -1$) as our reference, we correlated the group-time fixed effect for the treatment group five years before displacement (δ_{-5} from Equation 2) with estimated post-treatment losses (τ_h) for $h \in \{1, 10\}$. A strong negative correlation between δ_{-5} and τ_h would suggest that estimated losses might be driven by pre-existing trends rather than causal effects. Figure 11 in Appendix C displays the relationships, showing no systematic correlation for earnings, employment, wage losses. If anything, we find a *positive* correlation between pretrend coefficients and losses for small intervals of long-

[14] For simplicity, we present only the general idea of the bootstrap of little bags here. More formally, standard errors are estimated using influence functions and a variance decomposition that captures variability across groups of trees built on different subsamples. This estimated variance is then combined with a problem-specific curvature term to construct valid confidence intervals. A detailed explanation of the method is provided by Athey et al. (2019, pg. 18), while our implementation is described in Section B of the Appendix.

term employment and short-term wage losses. This correlation has the opposite sign of what we would expect if pretrends were driving our heterogeneity results, suggesting that any mild violation would bias our estimates toward understating heterogeneity rather than overstating it. The following additional checks put these into perspective. Using the single regression tree presented in Figure 1, we plotted event study coefficients separately for each final node, revealing flat pre-treatment trends across all nodes (Appendix C). This confirms that even after numerous splits, the parallel trends assumption appears realistic. Finally, we grouped observations by predicted loss deciles and estimated separate event studies (Figure 13 in Appendix C). In general, the pretrends are very small, and we find no systematic association between pre-treatment fixed effects and estimated loss magnitude. Overall, these findings support our assumption of parallel trends conditional on partitioning variables, and thus our interpretation that the heterogeneity uncovered by the forest approach reflects true variation in treatment effects rather than artifacts from parallel trends violations.

An equally important question concerns the accuracy of the specific loss magnitudes identified through the random forest. Measuring this accuracy presents a methodological challenge, as we are not interested in accurate predictions of an observed dependent variable on the left-hand side in Equation 2 but rather precise identification of an unknown treatment effect, one of the estimated regression coefficients on the right hand side. We address this concern by proposing an alternative evaluation procedure. In our exercise we create 50 distinctive groups.[15] For each group, we estimate the average cost of job loss using Equation 1. Then we compare these treatment effects with forest-implied treatment effects. As we show, both measures are very highly correlated, which suggests that our method provides accurate estimates. We relegate a more detailed discussion to Section D and Figures 14 and 15 in the Appendix.

B. Partitioning Variables

Before we present our estimation results in the next section, we discuss and motivate our choices for partitioning variables below. What are the reasons that earnings losses differ across individuals? On the one hand, workers in different environments face different re-employment probabilities after job loss, on the other hand, displaced workers might face different wage declines after re-employment. The workhorse framework in empirical research typically posits that wages consist of multiple components: $\ln(w_{it}) = \alpha_i + \psi_{J(i,t)} + \beta *$

[15] Our groups are created based on the forest-based treatment effects as we find it the most natural. That said, other grouping criteria could be potentially used as well.

$tenure_{it} + \varepsilon_{it}$. α_i captures the compensation a worker receives for their skills, $\psi_{J(i,t)}$ the firm specific wage component, $\beta * jobtenure_{it}$ the remuneration for the worker's job specific human capital accumulated over their job tenure, and ε_{it} represents an idiosyncratic match component.

In on-the-job search models such as Burdett (1978), laid-off workers fall down the job ladder in terms of firm wages and match quality and have to climb back through job-to-job transitions. Thus, earnings losses can vary first, because workers with higher firm wage components or better previous match quality face a steeper fall off the job ladder. In addition, workers with high skills α_i before displacement have more room to face depreciation of these skills. Moreover, workers with higher fixed effects might have accumulated more savings and higher claims to unemployment benefits, which would enable them to be pickier in their job search. This would lead to higher employment losses, but lower wage losses after job loss (Nekoei and Weber, 2017). Motivated by this we estimate the following wage equation (Abowd et al., 1999) on a rolling window of all social security spells in the five years before the event:

$$\ln(w_{it}) = \psi_{J(i,t)} + \alpha_i + \theta_t + x_{it}\beta + \epsilon_{it}, \tag{6}$$

where $\ln(w_{it})$ is the log daily wage of the dominant employer at period t, $\psi_{J(i,t)}$ represents the establishment fixed effect of the employer of worker i at period t, α_i the worker fixed effect, θ_t the year fixed effect, and x_{it} are time varying observables, comprising of a cubic polynomial of age. We then construct deciles of the estimated firm fixed effect $\hat{\psi}_{J(i,t)}$ and worker fixed effect $\hat{\alpha}_i$, and the wage residual $\hat{\varepsilon}_{it}$, which are then used as partitioning variables.

We further include job tenure and the firm specific separation rate in the five years leading up to the mass layoff event as partitioning variables.

In addition, earnings losses might vary because workers face different job finding probabilities or job offer distributions. These might change for example with the business cycle. To capture all time-varying factors we include a dummy for each displacement year as partitioning variables. Further we proxy the average job offer distribution displaced workers face by computing the average percentile of the firm wage effect in the local labor market.[16] In addition we study how the average unemployment rate over our study period in the local

[16]Specifically, we compute the average firm wage premia of all jobs in a given region leaving out all jobs of the worker's current employer. Formally for every worker i employed at firm $J(i,t)$ we compute $\sum_{k \notin J(i,t) \wedge k \in r(i)} \hat{\psi}_{J(k,t)} / \#(k \notin J(i,t) \wedge k \in r(i))$, where $r(i)$ is the region of the worker i.

labor market correlate with earnings losses.[17] Workers in concentrated labor markets or separating from a monopsonist employer might face more difficulty to quickly find a similarly good job (Gathmann et al., 2020). Thus, we include the overall local labor market concentration as measured by the Herfindahl-Hirschman index, the firm size, and the labor market share of the previous employer as partitioning variables. We also include a dummy variable for each NACE-1 industry to capture any time-invariant differences in the cost of job loss across industries.

Workers with higher job mobility in the past might find it easier to climb back the job ladder. The number of previously held jobs is therefore included as a partitioning variable, as well as a number of socio-demographic factors such as worker age, a dummy for Austrian citizenship, gender, a dummy for working in a blue collar occupation, and firm age.

To enhance interpretability, we categorize all continuous variables into deciles according to the overall distribution of Austrian workers on our reference day, and not only the selected displaced worker sample. This way the heterogeneity is easily interpretable in terms of the overall employment distribution in Austria. Table 4 in the appendix summarizes all the definitions of the partitioning variables, and Figure 10 reports a correlogram of all partitioning variables.

V. Heterogeneous Cost of Job Loss

In this section we use the machine learning algorithm to document the unequal consequences of job loss across workers. We focus on three labor market aspects: losses in relative earnings, employment days, and log wages. In our analysis, we examine both short-term and long-term losses.

A. The Distribution of the Cost of Job Loss

The generalized random forest provides an estimated treatment effect for each individual worker based on their characteristics. We are thus in a position to document how the scarring effect of job displacement differs across workers. We start by plotting the distributions of τ_h for relative earnings changes, employment, wage, for horizon $h \in \{1, 10\}$. Figure 2 shows the histogram of the cost of job loss for the three variables for the short-run (left panels) and long-run (right panels). The figure clearly shows that the short and long-term consequences of job losses are far from uniform across individuals. While the median worker in our sample

[17] We use the NUTS-3 district (35 categories) for the local labor market

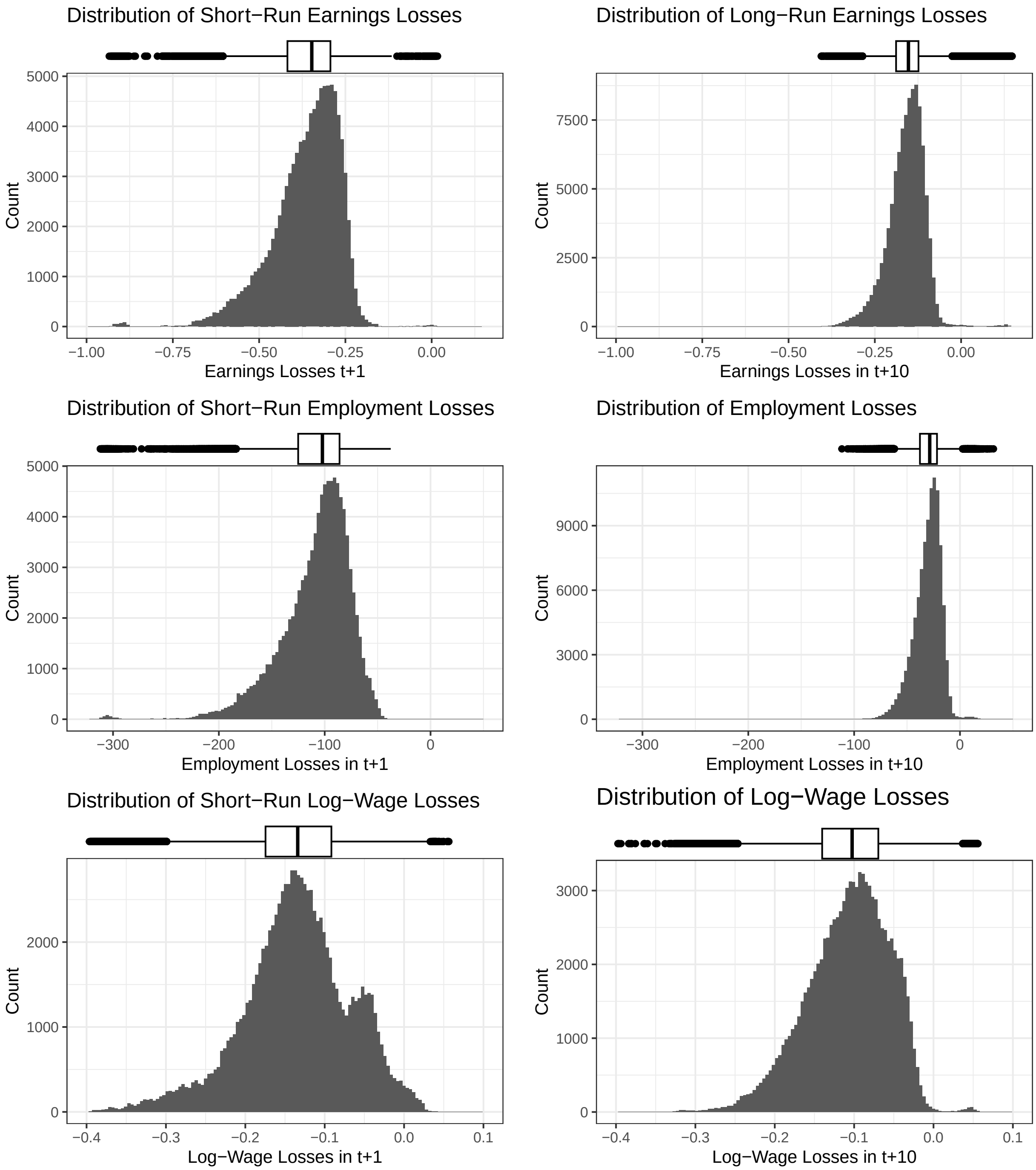


Figure 2: Histogram of short and long-run earnings, employment, and log-wage losses. Estimates from a generalized random forest.

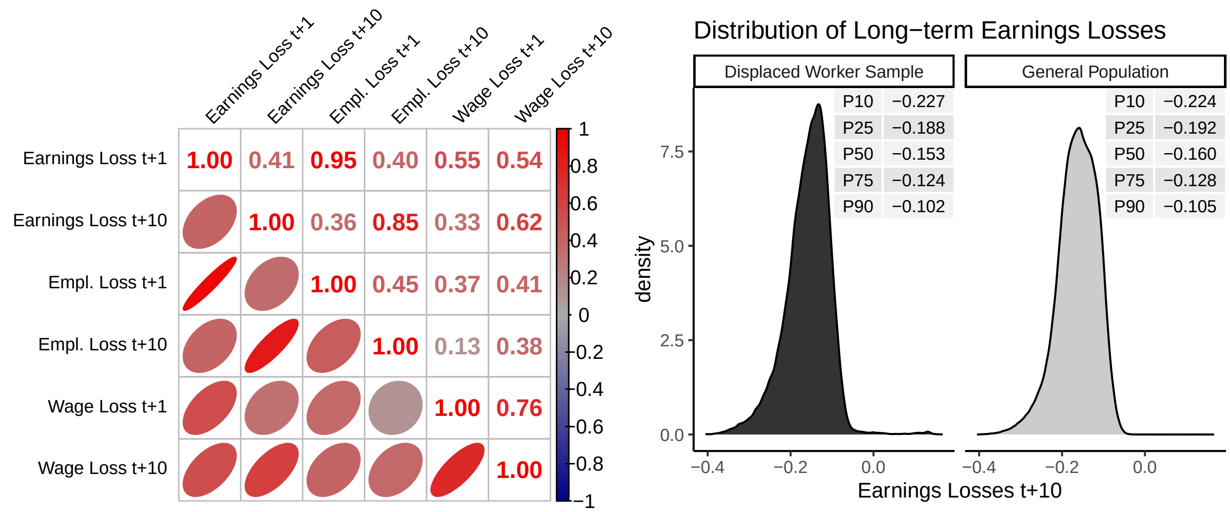
is facing income losses of 35 percent one year after the job loss, some workers lose less than 20%, while the hardest hit are predicted to lose over 70%. The 25th percentile of losses is 42%, while the 75th percentile is 29%. In the long-run, many workers recoup some of their lost earnings. The whole distribution of long-run losses shifts towards lower losses, and its dispersion shrinks markedly. While few workers face long-term losses of more than 25 percent, the interquartile range narrows from close to 15pp in the short run to 6pp in the long-run.

This heterogeneity is not driven by noisy estimates of the individual treatment effect. At the 95% confidence level, we estimate that 67% of workers face significantly different short-run earnings losses compared to the median losses of 35%. Because long-run losses are less dispersed and also somewhat less precisely estimated, close to 40% of workers have significantly different losses compared to the median loss of 15%.

Earnings losses are the result of both forgone income during non-employment and declines in wages after reemployment. In the second and third row of Figure 2, we plot the distributions of employment losses and log-wage losses in the short and long term. Again, there are striking differences in displacement costs across individuals. Predictable employment losses one year after job displacement range from more than 150 days for the 90th percentile to 73 days for the 10th percentile. Similar to overall earnings losses, the dispersion of employment losses is significantly lower 10 years after the mass layoff event; employment losses are concentrated between 50 and 16 days per year, 80 percent of workers face long-term employment losses in this range.

The last row shows that wage losses in the short run range from 35% to even wage gains for a handful of individuals. Interestingly, the distribution of wage losses in the short and long run are very similar, suggesting that wage losses are very persistent.

This is also illustrated quantitatively in Figure 3a, where we present the correlation matrix between short- and long-run earnings, employment, and wage losses. Individuals who face the highest cost of job loss in the short run also face higher losses in the long run. Earnings and employment losses one and ten years after the job loss have a correlation of between 0.40 and 0.45. Wage losses are much more persistent; short- and long-run losses are highly correlated with 0.75. The correlation pattern also reveals whether employment or wage losses are the main drives of overall earnings losses. In the short run, earnings losses are more correlated with employment losses (0.94) than with wage losses (0.52). Although the correlation of wage and earnings losses increases in the long run somewhat, long-run earnings losses are still more related to employment than to wage losses, with the correlation



(a) Correlations

(b) Distribution of Losses

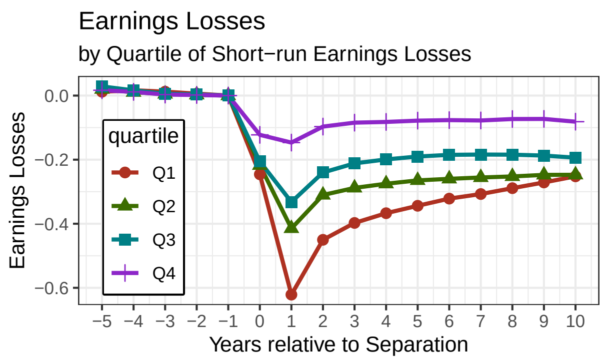
Figure 3: Left panel: Correlation between relative earnings losses, employment losses, and wage losses. Right panel: Distribution of earnings losses for displaced worker sample and the general population. Estimates from generalized random forests.

coefficient being 0.85 versus 0.55.

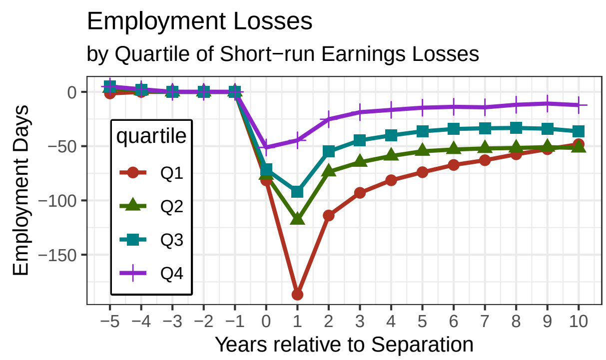
Remarkably, even 10 years after the initial job displacement almost all workers are predicted to experience earnings, employment, and even wage losses conditional on re-employment. One would perhaps expect that workers previously employed in low-wage jobs with high turnover would manage to improve their outcomes relative to the control group, but there is essentially no combination of our partitioning variables that predict this to happen. Thus, in line with recent evidence from the United States (Rose and Shem-Tov, 2023), job loss is costly even for low wage workers with disadvantaged worker and job characteristics.

A long-standing concern in the job displacement literature is that the strict sample selection criteria typically applied select workers that are bound to experience high earnings losses. This raises concerns about the generalizability of the documented earnings losses to the overall population. To tackle this, we use our grown forests to predict earnings losses for each employed worker that did not satisfy our sample criteria given their worker and job characteristics.[18] Figure 3b compares the distribution of long-term earnings losses in our displaced worker sample and the general population. The overall distributions are very similar. The differences in the 10th, 25th, 50th, 75th, and 90th percentiles of earnings losses

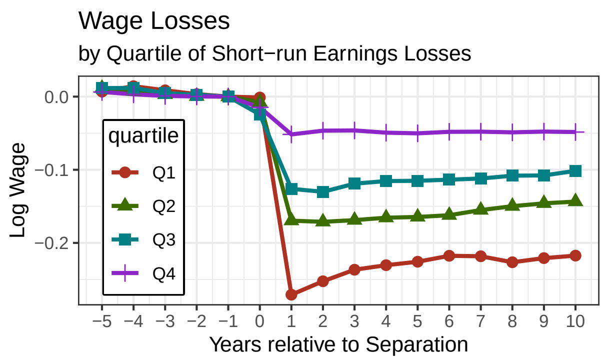
[18]For workers who have lower job tenure or are employed at smaller firms than our sample restrictions, we assume the lowest possible values.



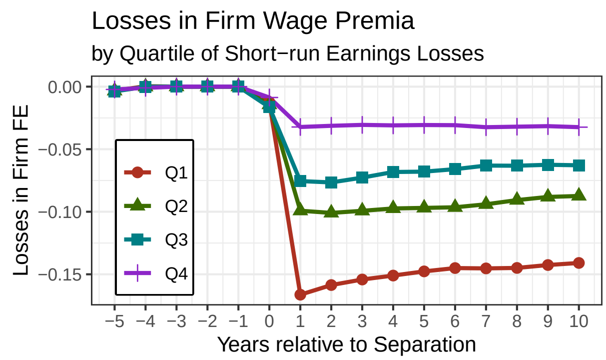
(a) Earnings Losses



(b) Employment Losses



(c) Wage Losses



(d) Firm Wage Premia Losses

Figure 4: Event-study regression estimates of earnings, employment, wage, and firm wage premia losses, by quartile of predicted earnings losses

are all within 1pp. At least in our sample, the sample restrictions do not select worker and job characteristics that are related to significantly higher long-run earnings losses.

B. Post-Displacement Evolution

Why do workers face such different consequences from mass layoffs? Before we turn to which partitioning variables are the most important predictors of the cost of job loss in the next section, we study how labor market trajectories after the job loss differ for low versus high earnings loss individuals. We bin workers into quartiles according to their estimated short-run earnings losses and estimate the full event-study specification of Equation (1) for various left-hand side variables separately for every earnings loss quartile.

Figure 4 plots the dynamic evolution of earnings losses, employment losses, log-wage losses, and firm wage premia losses over time. The figure reconfirms that displaced workers experience vastly different labor market outcomes after job losses and that the heterogeneity in the cost of job loss is persistent. Although we group workers by their earnings losses in $t+1$, only the quartile with the highest losses (Q1) experiences any catch-up in earnings and employment. Wage losses are more persistent; there is barely any convergence noticeable between the groups.

Panel (c) and Panel (d) show the cost of job loss in terms of log-wages and firm wage premia measured in log points. Recently, a number of papers have studied the contribution of losses in firm wage premia to wage losses.[19] Changes in the firm wage component explain between 60-70 percent of the overall change in wage losses for each group. Changes in the firm wage premia also explain about 60-70 percent of the between-group differences in the evolution of wages after job loss. Thus, in our setting, the contribution of firm wage losses is rather uniform over the earnings-loss distribution.

Panel (a) of Figure 5 shows that differences in match quality, as measured by the residual in the AKM Mincer wage regression Equation (6), cannot explain the large wage loss differences.

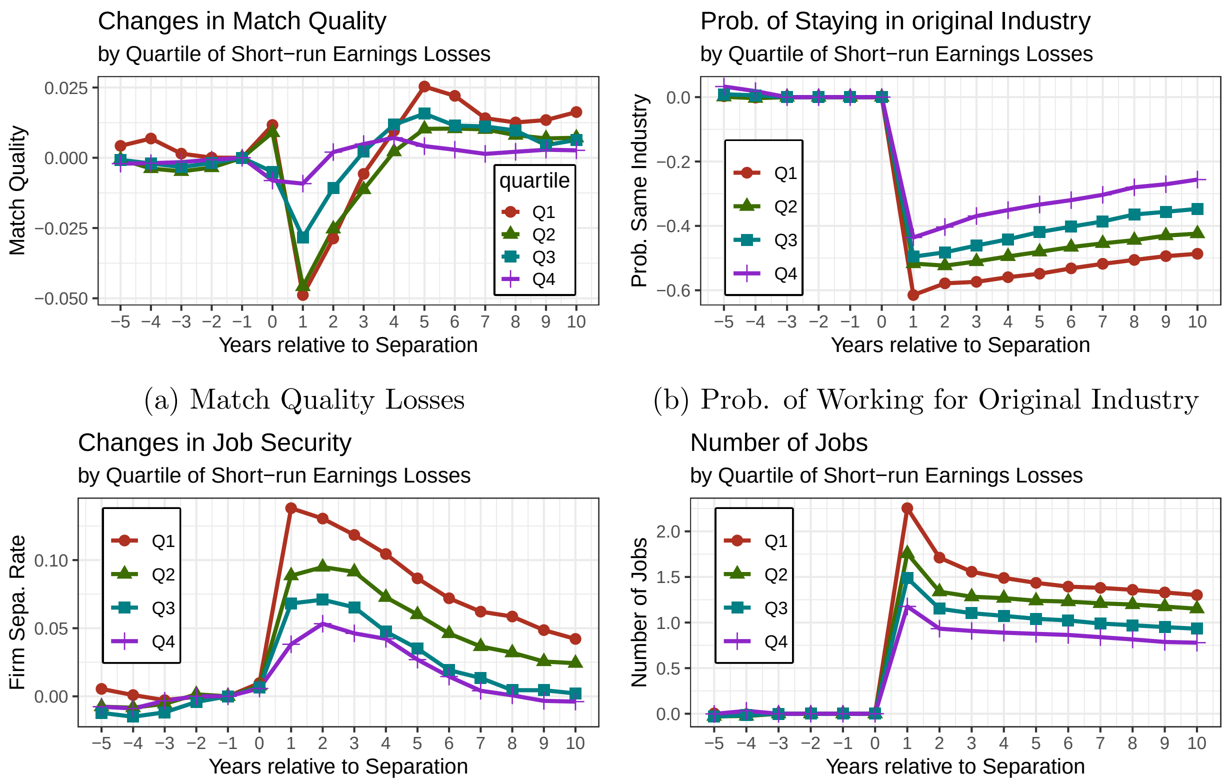
One explanation for wage losses is that workers lose firm- and industry-specific human capital.[20] Indeed, Panel (b) shows that the higher the losses, the less likely it is that workers stay in their original NACE-1 industry. Any job- and industry-specific human capital losses might be compounded if workers in addition cannot find stable employers after job losses (Jarosch, 2023). Panels (c) and (d) show that workers with the highest short-run earnings losses move to firms with less job security and have more job changes consequently.

C. Correlates of Earnings Losses

What are the most important pre-displacement worker and job characteristics that predict the cost of job loss? These variables could also be used to target policy interventions, something we tackle in the subsequent section.

An intuitive way to judge the importance of individual factors is to compute the elasticity of losses with respect to individual variables, holding all other factors constant. We compute how earnings, employment, and wage losses change with different values of one variable at a time, while holding all other variables fixed at their median. This way we are able to control for observable confounding factors. Figure 6 provides an overview by how much losses change, when moving individual partitioning variables from the lowest (min) to its

[19] *E.g.*, Fackler et al. (2021), Schmieder et al. (2023), Bertheau et al. (2023), and Lachowska et al. (2020).
[20] See *e.g.*, Jacobson et al. (1993), Neal (1995), and Jarosch (2023),



(a) Match Quality Losses

(b) Prob. of Working for Original Industry

(c) Job Security Losses

(d) Number of Jobs

Figure 5: Event-study regression estimates of match quality changes, probability of staying in the same industry, changes in firm separation rates, and number of jobs, by quartile of predicted short-run earnings losses.

highest (max) value, while holding other variables constant at their median. We present the point estimates of our forests, alongside the 95 percent confidence intervals that take into account both the uncertainty arising from the estimation and the building of the forests. In appendix section E we present the estimates for the cost of job loss for each realization of the partitioning variables, but as losses typically exhibit monotone and near-linear relationships with all variables, thus Figure 6 provides an accurate overall picture. For more readability, we only show the ten most important variables. As can be seen in the figures, the elasticity of losses with respect to less important variables is close to zero.

The first row shows the most important predictor of earnings losses in the short (left panel) and long run (right panel), while the second and third row present the results for employment and log-wage losses. Worker's age is the most important correlate for short-term losses and the third most important for long-term earnings losses. The oldest workers

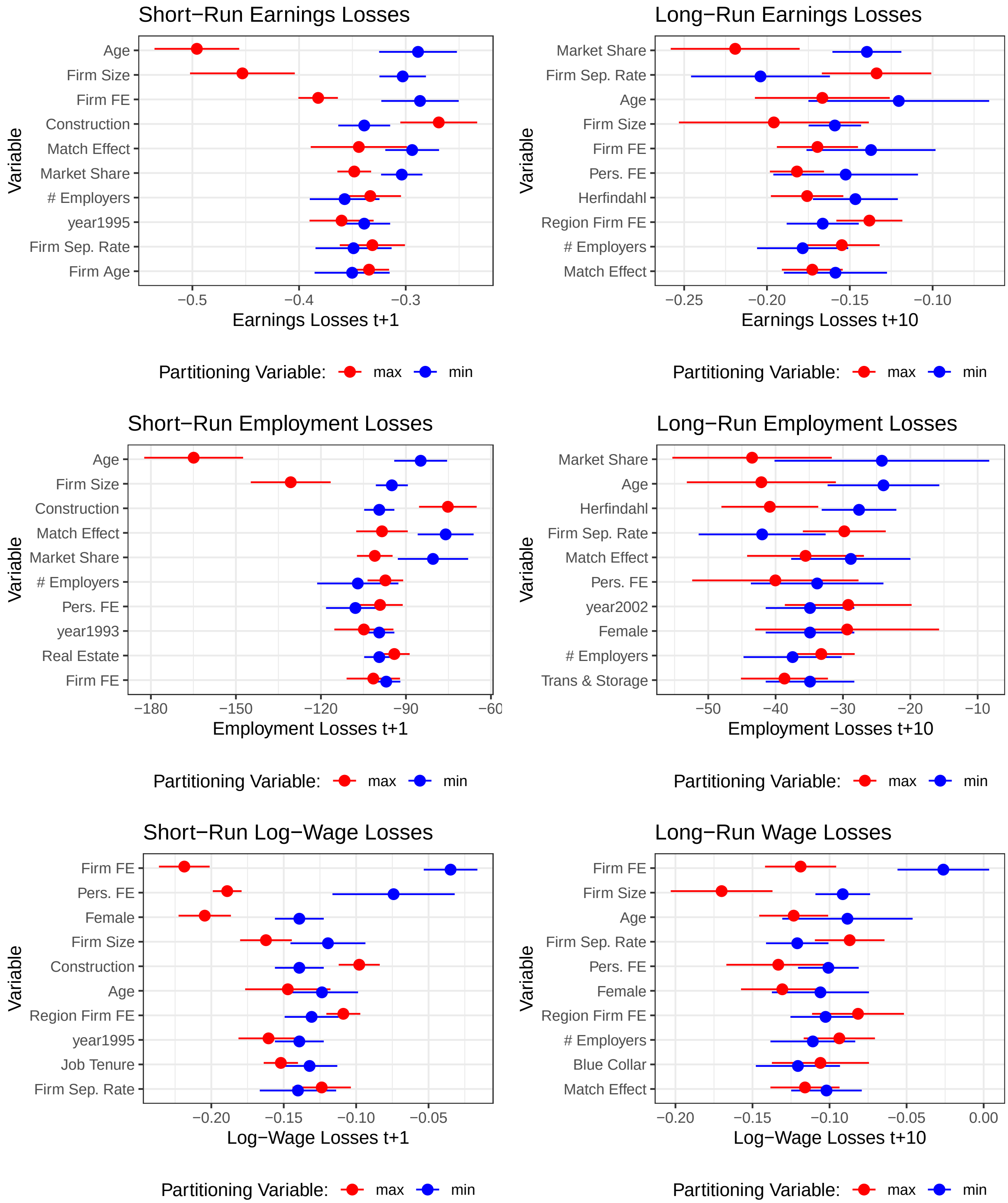


Figure 6: 10 most important correlates of earnings, employment, and wage losses. Estimates from a generalized random forest. All other variables are held at the median value.

in our sample face short-term earnings declines of 50 percent, almost twice as high as the youngest workers, even when we hold all other partitioning variables constant. In the long run the oldest workers face 4pp higher losses.[21] This finding is consistent with Guvenen et al. (2017), who also find that old workers have much higher earnings losses. Do older workers face higher earnings losses because of worse employment or wage losses? Although age is always among the 10 most important variables for employment and wage losses, age plays a much bigger role for employment losses. The oldest workers face employment declines of 165 days in the year following the job loss, while young workers experience close to 90 lost employment days. In the long run a differential of 18 employment days persists. The differential for wage losses is much smaller, with 2.5-3.5pp.

Furthermore, the displacing firms' characteristics play an important role in shaping the cost of job loss. Firm size, the labor market share, the firm's separation rate, and the firm wage premium all are among the most important variables for earnings, employment, and wage losses.

Firm size is the second most important variable for short-term and market share the most important variable for long-term earnings losses. This is in line with Fackler et al. (2021), who find elevated losses in larger firms in Germany. In our setting, this holds even after controlling for many observable worker and firm characteristics. Workers from the largest firms face earnings declines that are 15pp larger in the short run and 8pp larger in the long run than the losses in the smallest firms.

Gathmann et al. (2020) document that mass layoffs at regional flagship firms can have negative spillover effects on other firms in the region, and thus depress labor market outcomes for the whole region. Our results confirm this: separations from firms with a high labor market share in a given region depresses long-run earnings, mostly through more persistent employment losses. Workers separating from firms with the highest employment share face close to 20 more lost employment days 10 years after the job loss.

Recently, the role of firm wage premia for the cost of job loss moved in the center of attention in the literature. Fackler et al. (2021), Schmieder et al. (2023), and Bertheau et al. (2023) find losses in firm wage premia explaining a large fraction of wage losses, whereas Lachowska et al. (2020) find limited role of firm wage premia. We can shed light on this discussion from a new angle: Is the pre-displacement firm wage premium an important

[21] Labor force attachment of old workers has risen over our sample period. Did this higher attachment lead to a change in the earnings loss profile over age? To answer this question, we also predicted the age profile for different displacement years. For short-term losses there is no noticeable change over the years, for long-term losses the age profile flattens somewhat in the 2000s, although this change is very small.

predictor for subsequent earnings, employment, and wage losses? For wage losses, the answer is clearly yes. The firm fixed effect is the most important variable for short and long-run wage losses. Workers being displaced from the highest paying decile of firms face short-term losses of 18pp and long-run losses of 9pp more than the employees of the lowest paying firms. Thus, for each log point higher pre-displacement firm fixed effects, long-run wage losses increase by roughly 0.15 log-points. Although the firm fixed effect is not associated with higher employment losses, due to its strong correlation with wage losses, it features as the third and fifth most important variable for overall earning losses, with loss differentials of 10pp and 3pp for the short and long run, respectively.

Consistent with the theory in Jarosch (2023), workers who lose jobs at particularly stable firms face 8pp higher earnings losses. This is the second highest elasticity for long-run earnings losses. These higher earnings losses are a combination of higher employment (12 days) and wage losses (3.4pp).

In contrast to employment losses, wage losses vary strongly with the pre-displacement person fixed effects and gender. Even when comparing workers with the same characteristics, females are predicted to face 6.5pp higher short and 2.5pp higher long-run wage losses than men, similar to findings in Illing et al. (2024). The person fixed effect is even more important. Workers in the highest decile of worker fixed effects experience 11pp higher and 3pp higher wage losses compared to the lowest.

Few of the remaining variables show notable associations with losses. This is surprising, as some of these variables are discussed prominently by the prior literature. One example is business cycles, Schmieder et al. (2023), Davis and Von Wachter (2011) among others point to large differences in losses over the business cycles. Over our study period, Austria faced several recessions: 1995, 2003, 2009 all featured a decline in GDP together with rising unemployment rates. Despite this, year dummies do not emerge as important predictors for losses. Our results suggest that at least in our setting, it is more that the characteristics of displaced workers change in a recession, rather than the recession having a large impact by itself.[22]

The literature has often documented elevated losses of high-tenured workers and has taken this as evidence for the important role of human capital depreciation for explaining losses (Jacobson et al., 1993; Neal, 1995). We in contrast do not find job tenure to be an important predictor for the cost of job loss.

[22] We study this in more detail in an accompanying paper for the great recession and the Covid recession (Gulyas and Pytka, 2020).

Helm et al. (2023) brought into attention the special case of job displacement in manufacturing and document that manufacturing workers face elevated losses. In our setting, once we control for different firm and worker characteristics of displaced workers across manufacturing and non-manufacturing workers, manufacturing is no longer special. Generally speaking, none of the industry dummies with the exception of construction shows up among the most important correlates.

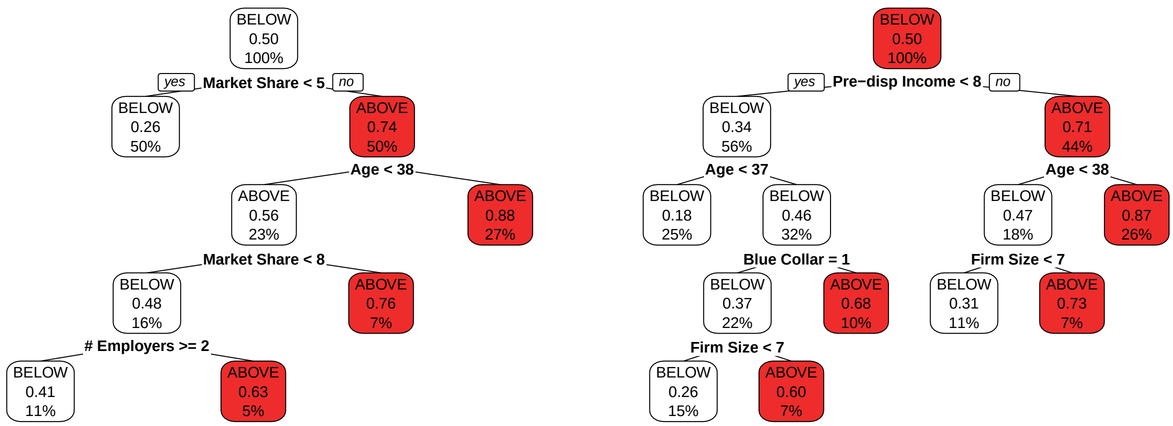
Perhaps surprisingly, regional characteristics such the regional unemployment rate, the Herfindahl index of labor market concentration, and average firm pay premia in the region do not feature very often among the most important variables. The only exception is the Herfindahl index, very concentrated labor markets are associated with 15 days of additional long-run employment losses. Summarizing, losses are more strongly correlated with firm and worker characteristics than with environmental factors.

One potential criticism of the partial effects presented so far is that an individual with median characteristics might not be representative for the whole population, and presented effects might be very different for characteristics other than the median. To tackle this critique, we use partial dependence plots proposed by Friedman (2001) to better understand how a single variable affects *on average* the earnings losses in the sample. This approach consists in estimating the earnings losses for each individual by changing the value of one variable, while holding all other characteristics constant at their empirical values, and then averaging over individuals. The appendix shows that this yields very similar results.

An alternative and popular way in the machine learning literature to assess which factors are the most important is the occurrence frequency of variables in the splitting criteria. Variables chosen more frequently and earlier in the trees have a higher contribution in explaining the heterogeneity of scarring effects. Figure 16 in the appendix shows that the overall picture for the most important predictors is very similar using this method.

VI. Policy Targeting

Given that the cost of job loss differs substantially across individuals, targeted labor market interventions may improve efficiency by prioritizing individuals predicted to experience prolonged non-employment spells or substantial wage losses. In principle, our forests could directly target high-loss individuals but many variables are not easily understood by the general public and policymakers. Therefore, we construct classification trees using only straightforward criteria like age, pre-displacement income, and firm size and we exclude complex variables (firm wage premia, regional concentration metrics, match effects) and year dummies. This way we create transparent, time-invariant decision rules that policymakers can readily understand and implement. We use the costs identified by our random forest and project those onto simpler classification tree structures due to Breiman et al. (1984). This approach ensures "the right to explanation" for those affected by policies while enabling effective targeting of different programs to workers facing employment or wage losses.[23]



(a) Long-term employment losses

(b) Long-term log wage losses

Figure 7: Policy trees for different long-term losses. Each leaf in the tree is characterized by three elements: (i) whether individuals in that group are predominantly above or below the median loss (ABOVE or BELOW), (ii) the share of individuals in the leaf who experience above-median losses, and (iii) the share of the total sample represented by that leaf. The root (top) node includes the entire sample, so it always shows 100%. The shares across all terminal nodes sum to 100%. At each split, the left branch groups individuals who satisfy the displayed condition.

Figure 7 presents classification trees that identify individuals with above median long-term losses. These correspond to workers with at least a 10% drop in wages and employment losses of at least 28 days. The red color indicates leaves where at least 50% of the workers experience above-median losses.

As shown in Panel 7a, labor market program especially effective at mitigating long non-

[23]In EU countries, policymakers must provide grounds for their decisions. The General Data Protection Regulation (EU) 2016/679 introduced by the European Parliament and Council of European Union (2016) established "the right to explanation" for individuals subject to automated decision-making. Article 12 requires *"concise, transparent, intelligible and easily accessible form, using clear and plain language"* for information processing, while Article 13 guarantees the right to *"meaningful information about the logic involved."*

employment spells should target the following groups: (i) employees older than 38 displaced from firms with above-median market share, (ii) workers younger than 38 displaced from firms in the top decile of market share or who had relatively few previous employers before displacement. Using these relatively simple targeting criteria, policymakers could correctly identify 75.6% of individuals who will experience above-median employment losses.

Regarding policies addressing severe long-term wage losses, Panel 7b indicates that resources should primarily target the following groups: (i) workers with pre-displacement income in the top quintile who are either older than 38 or displaced from large firms, (ii) workers outside the top income quintile who are older than 37 and are blue-collar or displaced from large firms. This targeting approach would correctly identify 77.3% of workers who will face above-median wage losses.

The exercise in this section illustrates a potential application for guiding optimal policies. In principle, alternative trees can be easily built using different loss thresholds and variables.

VII. Heterogeneity Detection: Comparison with Other Methods

In this section, we compare the short-term losses in relative earnings identified through our generalized random forest with results from several alternative methods commonly used to study heterogeneity in treatment effects. We distinguish between two complementary perspectives. First, we examine heterogeneity that can be related to observed covariates — the same variables used for growing our forests — by estimating a difference-in-differences regression in which the regression coefficients are interacted with covariates selected via a LASSO penalty, an increasingly popular approach in empirical work. Second, we consider overall heterogeneity in displacement costs without restricting attention to variation explained by observables. To this end, we employ two quantile-based DiD estimators—Changes-in-Changes and quantile DiD—introduced by Athey and Imbens (2006), which allow us to identify the full distribution of treatment effects. Additionally, we use the statistical "twins" method from Schmieder et al. (2023), which compares outcomes of treated and control individuals matched using propensity score matching.

One strategy for studying heterogeneity in earnings losses is to interact all coefficients from Equation (1) with observable individual characteristics. However, if we were to interact all values of all partitioning variables, we would end up estimating a model with 64,795 parameters. This approach risks overfitting and potentially yields very noisy estimates. To address this problem, we implement LASSO (Tibshirani, 1996) penalization on all coefficients interacted with dummies associated with each realization of the partitioning variables.[24] Overall, the regularized model reduces the number of parameters to 5,402.

Table 3 shows the distribution of identified losses and their accuracy. The interquartile range for treatment effects estimated using LASSO is approximately 30pp, which is larger than the roughly 15pp observed with our generalized random forest approach. Despite this wider dispersion in earnings losses, LASSO estimates are less precisely determined: only 21% of individuals exhibit losses that are statistically different from the median level. This is more than three times lower than our random forest approach, where 67% of individuals have losses statistically different from the median. Another concern with LASSO is that the estimated median treatment effect (-0.100) differs substantially from the homogeneous effect (-0.361) estimated in previous sections.

One potential explanation for these differences lies in the objective function of regularization methods like LASSO. This method focuses on accurately estimating the *dependent* variable rather than specifically identifying treatment effects. Since earnings losses constitute only a small proportion of overall earnings levels, the regularization algorithm may prioritize variables that contribute to earnings levels generally, rather than those most important for identifying treatment heterogeneity specifically.

	Treatment Percentiles			Fraction of individuals with losses *statistically significant*:		
	25%	50%	75%	below the median	above the median	below or above the median
GRF	-0.42	-0.35	-0.29	0.34	0.33	0.67
LASSO	-0.26	-0.10	0.05	0.11	0.10	0.21
Statistical "twins"	-0.81	-0.31	-0.01		–	

Table 3: Heterogeneity in earnings losses and estimation precision. The left panel shows the 25th, 50th, and 75th percentile of earnings losses for LASSO, generalized random forest, and statistical "twins" methods. The right panel presents the fraction of workers with statistically significant losses different from the median at the 95% confidence level. The median treatment effect used to determine statistical significance is calculated separately for each method. Note that the statistical "twins" approach, borrowed from Schmieder et al. (2023), does not provide a straightforward way to estimate standard errors for individual effects, hence the missing values in the right panel.

Another approach to detecting heterogeneity in displacement losses comes from the statistical twins method proposed by Schmieder et al. (2023). It leverages our matching procedure,

[24]The details of the analysis using LASSO are presented in Appendix G.

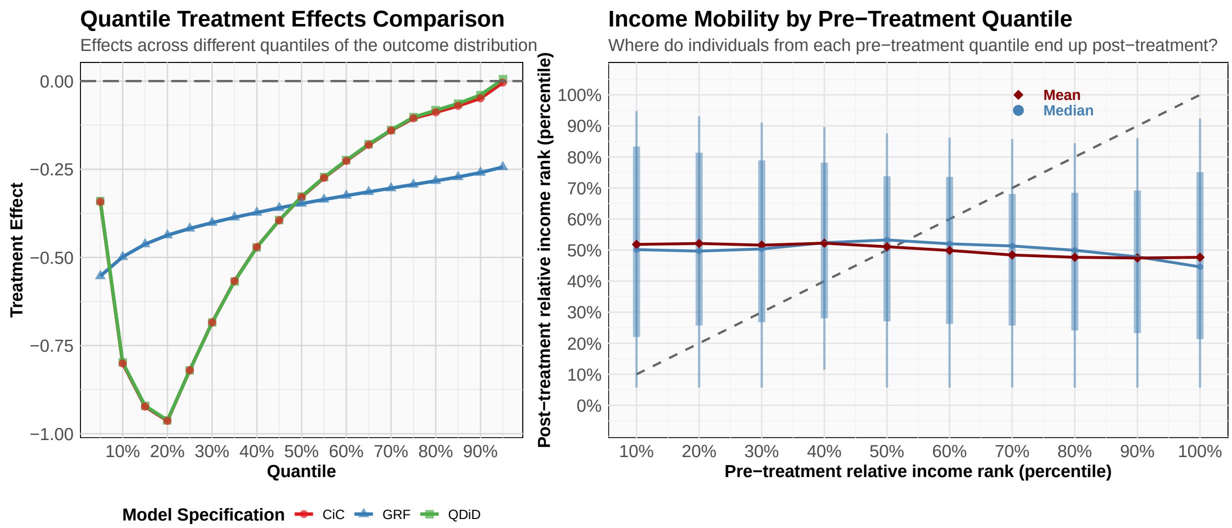


Figure 8: The left panel shows changes in earnings deciles before and after displacement, estimated using QDiD and CiC methods. For GRF, the plot displays deciles of individual treatment effects. The horizontal axis indicates the quantiles of the outcome distribution, and the vertical axis represents the estimated treatment effect. The right panel illustrates the relationship between individuals' pre-treatment relative income rank (5 years before displacement) and their post-treatment relative income rank (one year after displacement). Each vertical line shows the inter-quartile range (25th–75th percentile) and the 10th–90th percentile of post-treatment ranks. The 45-degree dashed line denotes perfect rank persistence.

where each job loser in our sample is paired with a statistical control "twin" who retained employment. We calculate individual-level estimates of wage losses at displacement as the DiD in short-term changes before and after job displacement between the displaced worker and their matched statistical "twin." As shown in Table 3, this method identifies a much wider distribution of losses, with an interquartile range of approximately 80pp in the short term—substantially higher than the ranges identified by GRF and LASSO methods. While this approach provides a comprehensive overview of the distribution of earning losses, it has important limitations. First, extracting systematic patterns by filtering statistical noise is challenging, and there is no clear way to identify standard errors. Furthermore, unlike our method, the twins approach does not allow for predicting losses based on covariates.

The final approach, based on quantile regression, detects variation in earnings losses independently of observed characteristics. Specifically, we use two non-linear extensions of the DiD method: Changes-in-Changes (CiC) and Quantile Difference-in-Difference (QDiD).

Both methods identify changes in the distribution of the dependent variable caused by a treatment.[25] The treatment effect for a specific q-quantile is identified as the difference between the q-quantile of earnings in the treatment group and the corresponding q-quantile in the control group. This method is popular for studying distributional changes of earnings, as exemplified by Jarosch (2023, Section 4.1.4).

The left panel of Figure 8 shows that the distribution of individual earnings losses from our forest is much less dispersed than changes in distributions identified through CiC and QDiD. While CiC and QDiD effectively evaluate overall distributional responses to treatment, they impose stringent requirements for identifying individual-level effects. Quantile regressions can only accurately assign individual earnings losses if the ranks of earnings before and after treatment remain perfectly correlated — a condition rarely met in practice. The right panel of Figure 8 clearly illustrates this limitation in our data and strong mean reversion.[26] Workers' pre-treatment ranking of the dependent variable (relative income), on average falls down after job separation, while the pre-treatment bottom 50% climb up in their ranking. Consequently, these methods capture distributional changes, but they do not reveal the distribution of individual losses.

VIII. Conclusions

We implement a generalized random forest (Athey et al., 2019) to a DiD setting to comprehensively and systematically study the heterogeneity of earnings losses of displaced workers. This methodology allows us to make a number of important empirical contributions to the existing literature.

First, we document substantial predictable heterogeneity in the causal cost of job loss across individuals. While almost no worker is predicted to have losses lower than 20% in the year after the job loss, the worst affected workers face losses of more than 70%. Although there is some convergence of losses across workers, the correlation of short and long-term losses at the individual level is fairly high at 0.4. Essentially all workers are predicted to face earnings losses in the long run, suggesting that job losses are costly even for low paid workers with disadvantaged characteristics. In general, earnings losses are more strongly correlated with employment than with wage losses, suggesting the employment losses are a

[25] A detailed discussion on how to impute the comparison group in the DiD setting can be found in Athey and Imbens (2006), where CiC and QDiD were first proposed.

[26] This violation of rank preservation leads to non-monotone estimates for quantile effects, as shown in the left panel of Figure 8. The effect occurs because the difference between the 20th quantile of the treatment group and the control group is higher than the differences observed at higher quantiles.

more important driver of the heterogeneity in earnings losses. Wage losses in the short and long-run have a correlation of 0.8, suggesting that wage losses are very persistent.

Second, a long-standing concern is whether the strict sample restrictions used in the literature select individuals that are bound to experience large earnings losses. We use the random forest to predict long-term earnings losses for the population that fails to meet the sample selection criteria and show that they are surprisingly similar.

Third, we show that the most important predictors for overall earnings losses are worker's age, and firm characteristics such as size, employment share in the region, its job security before the mass layoff event, and its wage premia.

Last but not least, we show how our results can be used to derive easily understandable decision rules to target specific programs to workers predicted to face long non-employment spells and long-run declines in wages. We cannot speak to the welfare effects of active labor market programs, but our results suggest that policies aimed to mitigate the consequences of job loss such as firm bailouts, employment protection, and employment subsidies such as short time work schemes, which are often applied universally, should be targeted.

References

Abowd, John M, Francis Kramarz, and David N Margolis (1999) "High wage workers and high wage firms," *Econometrica*, 67 (2), 251–333.

Athey, Susan and Guido W. Imbens (2006) "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74 (2), 431–497, http://www.jstor.org/stable/3598807.

_ , Lisa K Simon, Oskar N Skans, Johan Vikstrom, and Yaroslav Yakymovych (2023) "The heterogeneous earnings impact of job loss across workers, establishments, and markets," *arXiv preprint arXiv:2307.06684*.

_ , Julie Tibshirani, and Stefan Wager (2019) "Generalized random forests," *The Annals of Statistics*, 47 (2), 1148–1178.

Bertheau, Antoine, Edoardo Maria Acabbi, Cristina Barceló, Andreas Gulyas, Stefano Lombardi, and Raffaele Saggio (2023) "The unequal consequences of job loss across countries," *American Economic Review: Insights*, 5 (3), 393–408.

Breiman, Leo (2001) "Random Forests," *Machine Learning*, 45 (1), 5–32, 10.1023/A:1010933404324.

_ , J. H. Friedman, R. A. Olshen, and C. J. Stone (1984) *Classification and Regression Trees*, Statistics/Probability Series, Belmont, California, U.S.A.: Wadsworth Publishing Company.

Burdett, K. (1978) "A theory of employee job search and quit rates," *The American Economic Review*, 68 (1), 212–220.

Couch, Kenneth A and Dana W Placzek (2010) "Earnings losses of displaced workers revisited," *The American Economic Review*, 572–589.

Davis, Jonathan MV and Sara B Heller (2020) "Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs," *Review of economics and statistics*, 102 (4), 664–677.

Davis, Steven J and Till Von Wachter (2011) "Recessions and the Costs of Job Loss," *Brookings Papers on Economic Activity*, 43 (2 (Fall)), 1–72.

European Parliament and Council of European Union (2016) "General Data Protection Regulation," *Council (EU) 2016/679*, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679.

Fackler, Daniel, Steffen Mueller, and Jens Stegmaier (2021) "Explaining wage losses after job displacement: Employer size and lost firm wage premiums," *Journal of the European Economic Association*, 19 (5), 2695–2736.

Farber, Henry S (2011) "Job loss in the Great Recession: Historical perspective from the displaced workers survey, 1984-2010,"Technical report, National Bureau of Economic Research.

— (2017) "Employment, hours, and earnings consequences of job loss: US evidence from the displaced workers survey," *Journal of Labor Economics*, 35 (S1), S235–S272.

Friedman, Jerome H. (2001) "Greedy function approximation: A gradient boosting machine.," *The Annals of Statistics*, 29 (5), 1189–1232, 10.1214/aos/1013203451.

Gathmann, Christina, Ines Helm, and Uta Schönberg (2020) "Spillover Effects of Mass Layoffs," *Journal of the European Economic Association*, 18 (1), 427–468, https://EconPapers.repec.org/RePEc:oup:jeurec:v:18:y:2020:i:1:p:427-468.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy (2019) "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 87 (4), 1307–1340, 10.3982/ECTA16566.

Gregory, Victoria, Guido Menzio, and David Wiczer (2025) "The alpha beta gamma of the labor market," *Journal of Monetary Economics*, 150, 103695.

Gulyas, Andreas and Krzysztof Pytka (2020) "The consequences of the COVID-19 job losses: who will suffer most and by how much," *Covid Economics*, 1 (47), 70–107.

— and — (2022) "Understanding the heterogeneity of earnings losses after job displacement: A machine-learning approach."

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song (2017) "Heterogeneous Scarring Effects of Full-Year Nonemployment," *American Economic Review*, 107 (5), 369–73.

Halla, Martin, Julia Schmieder, and Andrea Weber (2020) "Job displacement, family dynamics, and spousal labor supply," *American Economic Journal: Applied Economics*, 12 (4), 253–287.

Helm, Ines, Alice Kügler, and Uta Schönberg (2023) "Displacement Effects in Manufacturing and Structural Change."

Hethey-Maier, Tanja and Johannes F Schmieder (2013) "Does the Use of Worker Flows Improve the Analysis of Establishment Turnover? Evidence from German Administrative Data," *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften*, 133 (4), 477–510.

Illing, Hannah, Johannes Schmieder, and Simon Trenkle (2024) "The gender gap in earnings losses after job displacement," *Journal of the European Economic Association*, jvae019.

Jacobson, L.S., R.J. LaLonde, and D.G. Sullivan (1993) "Earnings losses of displaced workers," *The American Economic Review*, 685–709.

Jarosch, Gregor (2023) "Searching for job security and the consequences of job loss," *Econometrica*, 91 (3), 903–942.

Knaus, Michael C, Michael Lechner, and Anthony Strittmatter (2022) "Heterogeneous employment effects of job search programs: A machine learning approach," *Journal of Human Resources*, 57 (2), 597–636.

Lachowska, Marta, Alexandre Mas, and Stephen A Woodbury (2020) "Sources of displaced workers' long-term earnings losses," *American Economic Review*, 110 (10), 3231–3266.

Lindo, Jason M (2011) "Parental job loss and infant health," *Journal of Health Economics*, 30 (5), 869–879.

Neal, Derek (1995) "Industry-Specific Human Capital: Evidence from Displaced Workers," *Journal of Labor Economics*, 13 (4), pp. 653–677, `http://www.jstor.org/stable/2535197`.

Nekoei, Arash and Andrea Weber (2017) "Does Extending Unemployment Benefits Improve Job Quality?" *American Economic Review*, 107 (2), 527–61, 10.1257/aer.20150528.

Pytka, Krzysztof ⓡ Daniel Runge (2025) "Looking into the Consumption Black Box: Evidence from Scanner Data," Working Paper.

Rege, Mari, Kjetil Telle, and Mark Votruba (2011) "Parental job loss and children's school performance," *The Review of Economic Studies*, 78 (4), 1462–1489.

Rose, Evan K and Yotam Shem-Tov (2023) "How replaceable is a low-wage job?"Technical report, National Bureau of Economic Research.

Schaller, Jessamyn and Ann Huff Stevens (2015) "Short-run effects of job loss on health conditions, health insurance, and health care utilization," *Journal of health economics*, 43, 190–203.

Schmieder, Johannes F, Till Von Wachter, and Jörg Heining (2023) "The costs of job displacement over the business cycle and its sources: evidence from Germany," *American Economic Review*, 113 (5), 1208–1254.

Sexton, Joseph and Petter Laake (2009) "Standard errors for bagged and random forest estimators," *Computational Statistics and Data Analysis*, 53 (3), 801 – 811, https://doi.org/10.1016/j.csda.2008.08.007, Computational Statistics within Clinical Research.

Sullivan, D. and T. Von Wachter (2009) "Job displacement and mortality: An analysis using administrative data," *The Quarterly Journal of Economics*, 124 (3), 1265–1306.

Tibshirani, Robert (1996) "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58 (1), 267–288.

Zweimüller, Josef, Rudolf Winter-Ebmer, Rafael Lalive, Andreas Kuhn, Jean-Philippe Wuellrich, Oliver Ruf, and Simon Büchi (2009) "Austrian social security database," *Available at SSRN 1399350*.

Online Appendix

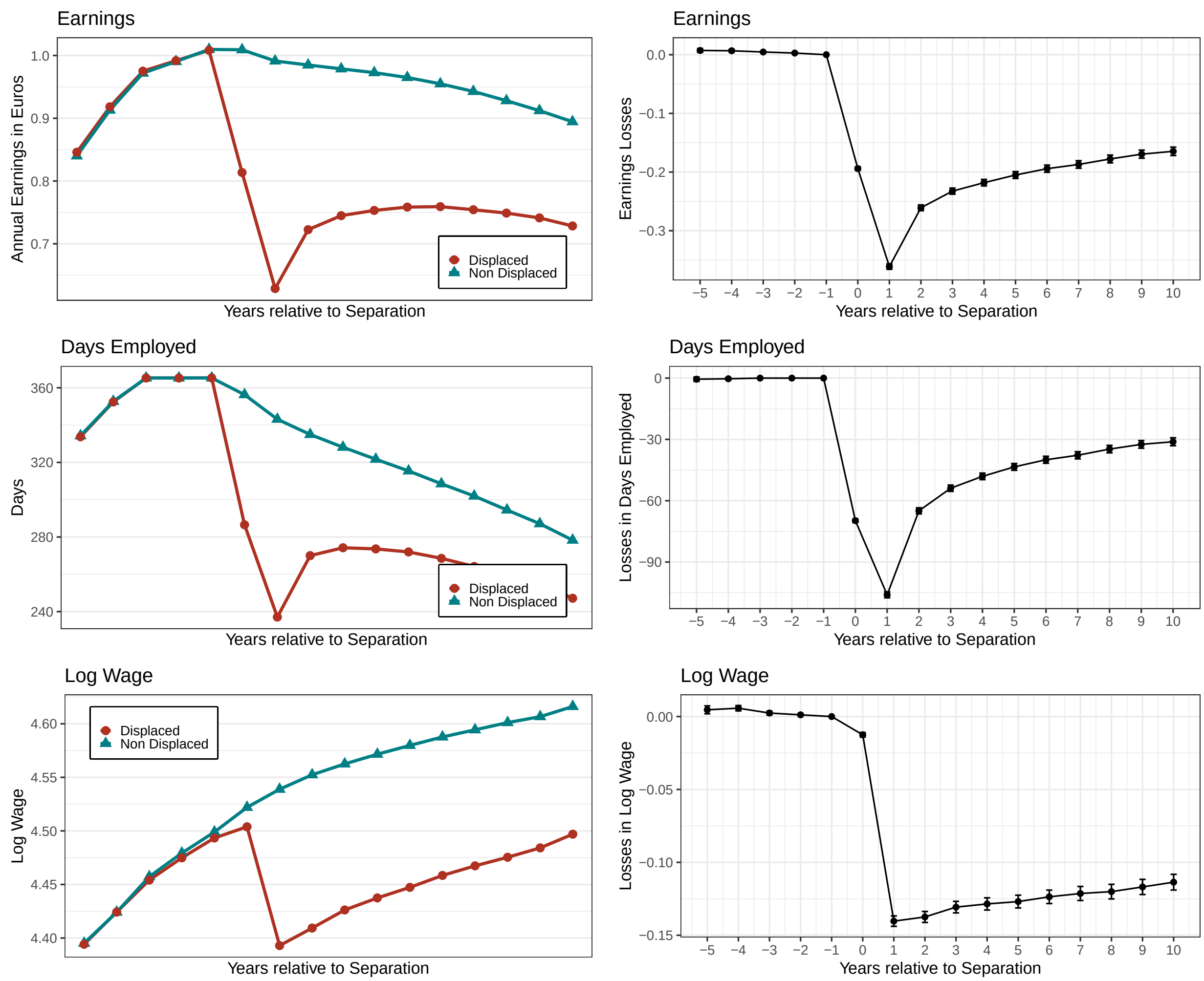


Figure 9: Earnings Losses of displaced workers - Eventstudy regression estimates of equation (1). Period 0 corresponds to the separation year. Earnings and days employed are computed for the whole year, log-wages are computed as the log average daily wage from the employer on 1st January. Control group is selected via propensity score matching.

A. Data Appendix

We use the labor market database provided by the Austrian social security agency. The data comprises all the relevant information to compute all benefits covered under social security in Austria. These include benefits related to old-age, unemployment, sick-leave, and maternity/paternity leave. Thus the dataset contains many overlapping spells that are not necessarily related to the labor market state of a worker. We follow the recommendations in the data manual provided by the data provider to eliminate overlapping spells and thus define unique labor market states for workers. For overlapping employment spells, we select the spell with the higher yearly income to define a unique employer at each point in time for workers.[27]

A. Partitioning Variables

Table 4 summarizes all partitioning variables and their definitions and Figure 10 shows how these variables are correlated with each other in the estimation sample.

[27] Working at multiple employers is very uncommon in Austria, this applies to less then 0.1 percent of spells.

Table 4: Definitions of the Partitioning Variables

Variable	Definition
Pers. FE	Worker Effect from Regression equation (6) binned to deciles
Firm FE	Firm Effect from Regression equation (6) binned to deciles
Match Effect	Residual from Regression equation (6) binned to deciles
Avg. Region Firm FE	Average percentile of the firm effect from regression equation (6), leaving out the previous employer. Binned into deciles.
Austrian	Indicator for Austrian citizenship
Female	Indicator for worker's gender
Blue Collar	Indicator for blue collar employment relationship
Worker's age	Worker's age in years
Firm age	Firm's age binned to deciles
Job Tenure	Job tenure measured at the beginning of the event year, binned into deciles
U-rate Region	Average unemployment rate 1984-2019 in NUTS-3 region of previous employer
Industry Dummies	Dummy for each industry in the NACE-1 classification of the previous employer
Firm Size	Number of employees on 1st of January of the event year
Market Share	Employment share of previous employer in NUTS-3 region and nace-1 industry
Herfindahl	Herfindahl-Hirschman index of labor market concentration in NUTS-3 and nace-1 industry.
Firm Sep. Rate	Average firm separation rate in the 5 years leading up to the event year, excluding recalls.
Year dummies	Dummy for each calender year
# Employers	Number of Employers before event year, observations binned above 4

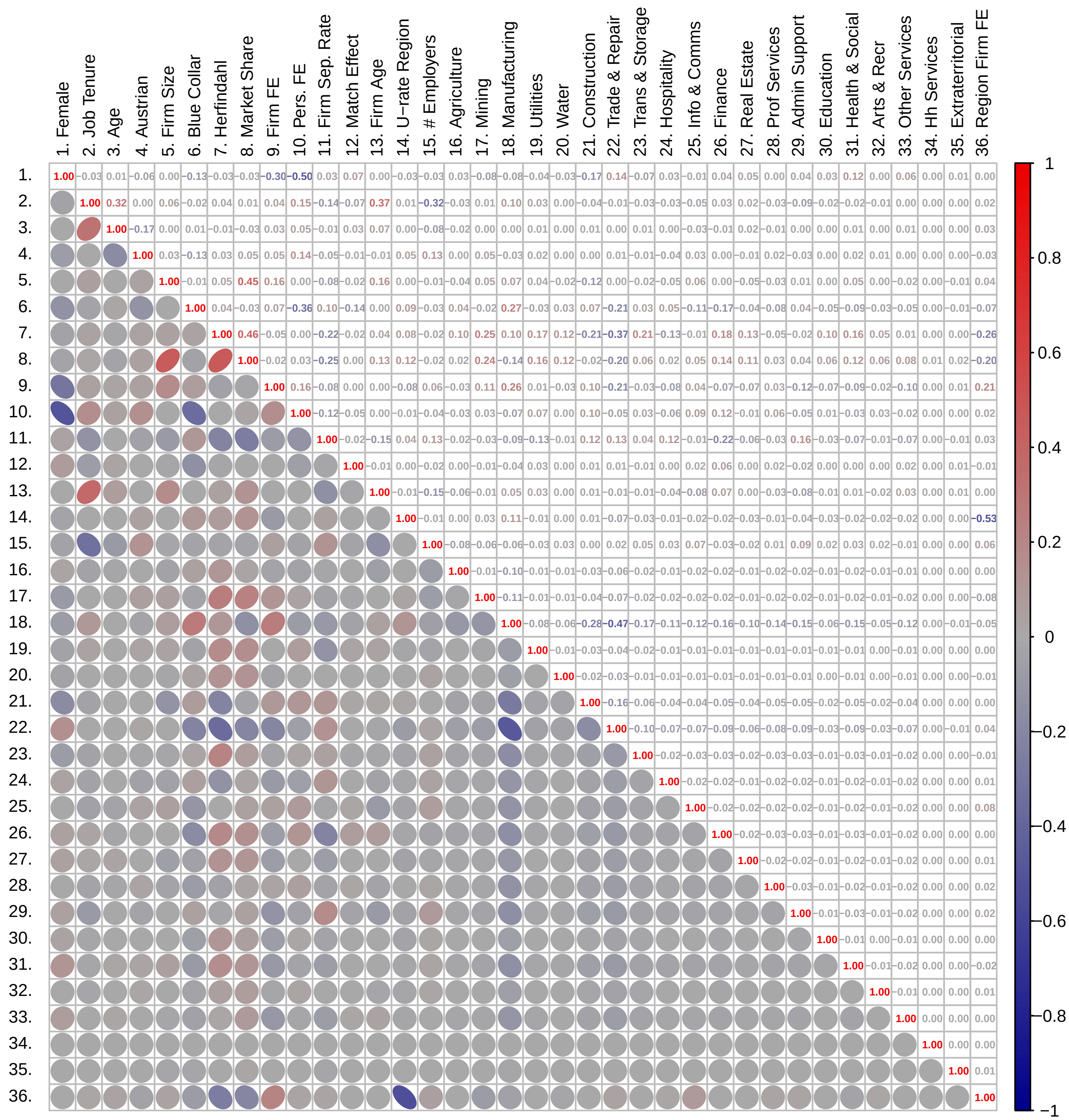


Figure 10: Correlogram of partitioning variables (year dummies not shown)

B. Construction of Weights and Numerical Implementation

With the forest at hand, we can proceed with the construction of weights. Suppose that there is a forest with B trees indexed by b. Then weight $\alpha_{it}^b(\mathbf{z})$ measures the similarity of observation (i,t) with $\mathbf{z}$ and is defined as:

$$\alpha_{it}^b(\mathbf{z}) := \begin{cases} \frac{1}{|L_b(\mathbf{z})|}, & \mathbf{z}_{it} \in L_b(\mathbf{z}) \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where $L_b(\mathbf{z})$ is the set of all observations, which share the same terminal node ("leaf") with an individual with characteristics $\mathbf{z}$ in tree b and $|L_b(\mathbf{z})|$ is the size of this set. The weight $\alpha_i(\mathbf{z})$ used in (4) is the average across all trees: $\alpha_i(\mathbf{z}) := \frac{1}{B}\sum_{b=1}^{B} \alpha_{it}^b(\mathbf{z})$.

As mentioned before, the forest is built to maximize the heterogeneity of treatment effects (with an additional adjustment for balanced subsamples) across splits and this is expressed by (5). That said, because of computational complexity, this criterion is replaced by more numerically efficient approximation in the spirit of gradient boosting due to Friedman (2001). However, before presenting the exact procedure, one remark should be made. In a given data partition $\mathcal{P}$, the OLS estimator trained on $\mathcal{P}$ meets the following condition:

$$\frac{1}{N_{\mathcal{P}}} \sum_{(i,t)\in\mathcal{P}} \mathbf{x}'_{it} u_{it} = \mathbf{0}_{18}. \tag{8}$$

Then the treatment effect τ_{C_k} of any subset $C_k \in \mathcal{P}$ can be approximated by:

$$\tau_{C_k} \approx \tau_{\mathcal{P}} + \xi' \left(\frac{1}{N_{\mathcal{P}}} \sum_{(j,s)\in\mathcal{P}} \mathbf{x}_{js}\mathbf{x}'_{js} \right)^{-1} \cdot \frac{1}{N_{C_k}} \sum_{(i,t)\in C_k} \mathbf{x}_{it} u_{it}, \tag{9}$$

where $\xi' = (1, \mathbf{0}_{11})$ is a vector selecting τ from the vector of all regression coefficients, and u_{it} is the residual term from the model estimated on $\mathcal{P}$.[28]

Then, the impact of an individual observation (i,t) on τ_{C_k} is given by:

$$\rho_{it} = \xi' \left(\frac{1}{N_{\mathcal{P}}} \sum_{(j,s)\in\mathcal{P}} \mathbf{x}_{js}\mathbf{x}'_{js} \right)^{-1} \cdot \mathbf{x}_{it} u_{it}. \tag{10}$$

[28] Notice that this approximation can be interpreted as an improved guess $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ where the function $f()$ is given by (3), x_n is $\tau_{\mathcal{P}}$ and x_{n+1} corresponds to τ_{C_k} in the textbook Newton-Raphson root-finding algorithm.

Using the CART algorithm by Breiman et al. (1984) on transformed outcomes (10), we are able to find such a split into C_1 and C_2 which minimizes the within-group sum of squares of ρ. Using the fact that the grand mean of ρ in the parent node is equal to zero, this implies that the algorithm maximizes the between-group sum of squares, *i.e.*:

$$\frac{1}{N_{C_1}}\left(\sum_{(i,t)\in C_1}\rho_{it}\right)^2+\frac{1}{N_{C_2}}\left(\sum_{(i,t)\in C_2}\rho_{it}\right)^2, \tag{11}$$

which, as Athey et al. (2019) show for a more general case, is consistent with maximizing criterion (5). Thanks to this relabelling strategy, the whole procedure of building a forest gains substantial computational performance.

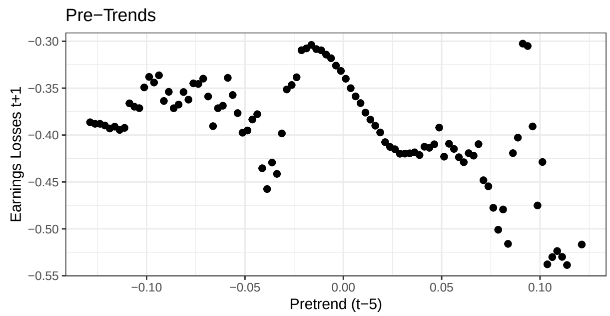
The formula for the variance of estimates can be derived, as in the standard GMM, by applying the delta method to the moment conditions of a weighted least squared regression, $f(\mathbf{z}) := \sum_{it}\alpha_{it}(\mathbf{z})\mathbf{x}_{it}\varepsilon_{it}(\mathbf{z})$. In our case we are interested only in $\hat{\tau}(\mathbf{z})$, so the whole formula is multiplied by ξ, which picks the estimate of our interest. As a result, the variance of $\hat{\tau}(\mathbf{z})$ is given by:

$$\mathrm{Var}(\hat{\tau}(\mathbf{z})) = \xi' V(\mathbf{z})^{-1}H(\mathbf{z})\left(V(\mathbf{z})^{-1}\right)'\xi. \tag{12}$$

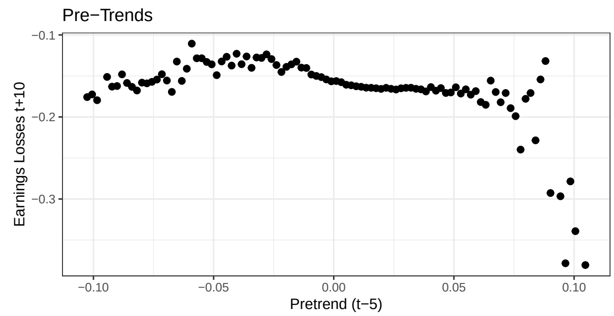
where $H(\mathbf{z}) := f(\mathbf{z})f(\mathbf{z})'$ is the variance of $f(\mathbf{z})$ and $V(\mathbf{z}) := \nabla_{\{\tau,\theta,\gamma\}}f(\mathbf{z}) = -\sum_{it}\alpha_{it}(\mathbf{z})\mathbf{x}_{it}\mathbf{x}'_{it}$ is the Jacobian of $f(\mathbf{z})$. That said, by no means Equation (12) should be estimated by simply using training observations, just like in the traditional GMM. The underlying reason for that is this would ignore the whole model selection step, which give rise to values of $\alpha_{it}(\mathbf{z})$ and $\varepsilon_{it}(\mathbf{z})$ in the presented formula. To circumvent this concern, as suggested by Athey et al. (2019), we employ a so-called bootstrap of little bags in the spirit of Sexton and Laake (2009) to evaluate $H(\mathbf{z})$. This procedure involves computing a between-group variance of $\hat{\tau}(\mathbf{z})$, where trees are pooled into bags and built using the same bootstrap subsample. Using one-way ANOVA it can be shown that this measure is approximately equal to (12). Thanks to this, our standard errors measure estimation accuracy affected by both machine-learning uncertainty and estimation noise.

C. Assessment of the Parallel Trends Assumption

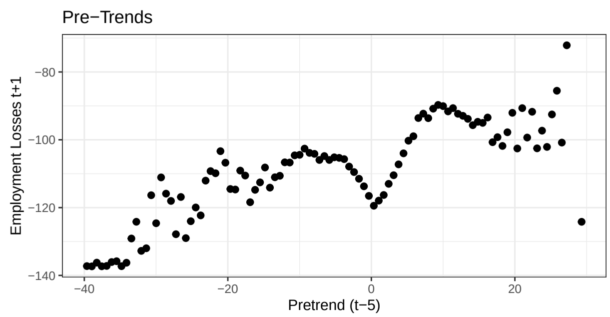
Figure 12 displays event-study coefficients for each terminal node of the decision tree in Figure 1, allowing for an assessment of the parallel trends assumption. Across subgroups,



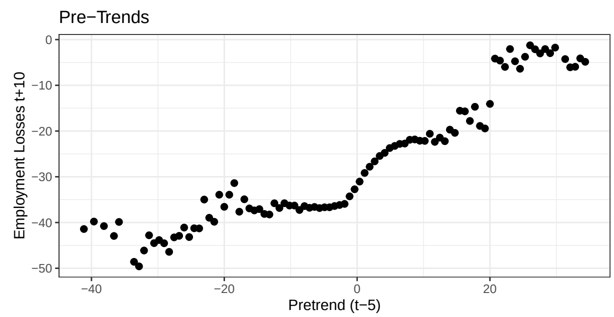
(a) Short-term relative earnings losses



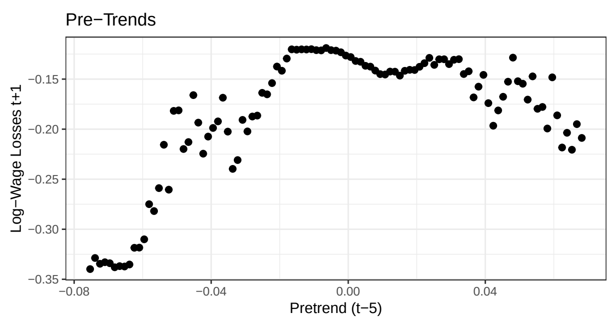
(b) Long-term relative earnings losses



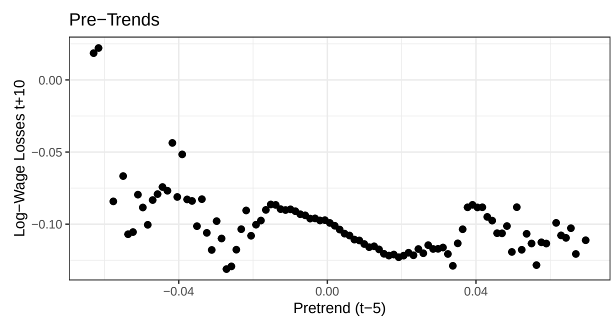
(c) Short-term employment losses



(d) Long-term employment losses



(e) Short-term log-wage losses



(f) Long-term log-wage losses

Figure 11: Correlation between pre-Trends and estimated cost of job loss. Scatter plot of δ_{-5} (pre-trend coefficients) against τ_h (estimated treatment effects) for relative earnings, log wages, and employment.

pre-displacement trends are generally stable. While some panels show slight movements before separation, these differences are either statistically insignificant or small in magnitude. Overall, the visual evidence supports the plausibility of the parallel trends assumption within the identified subgroups.

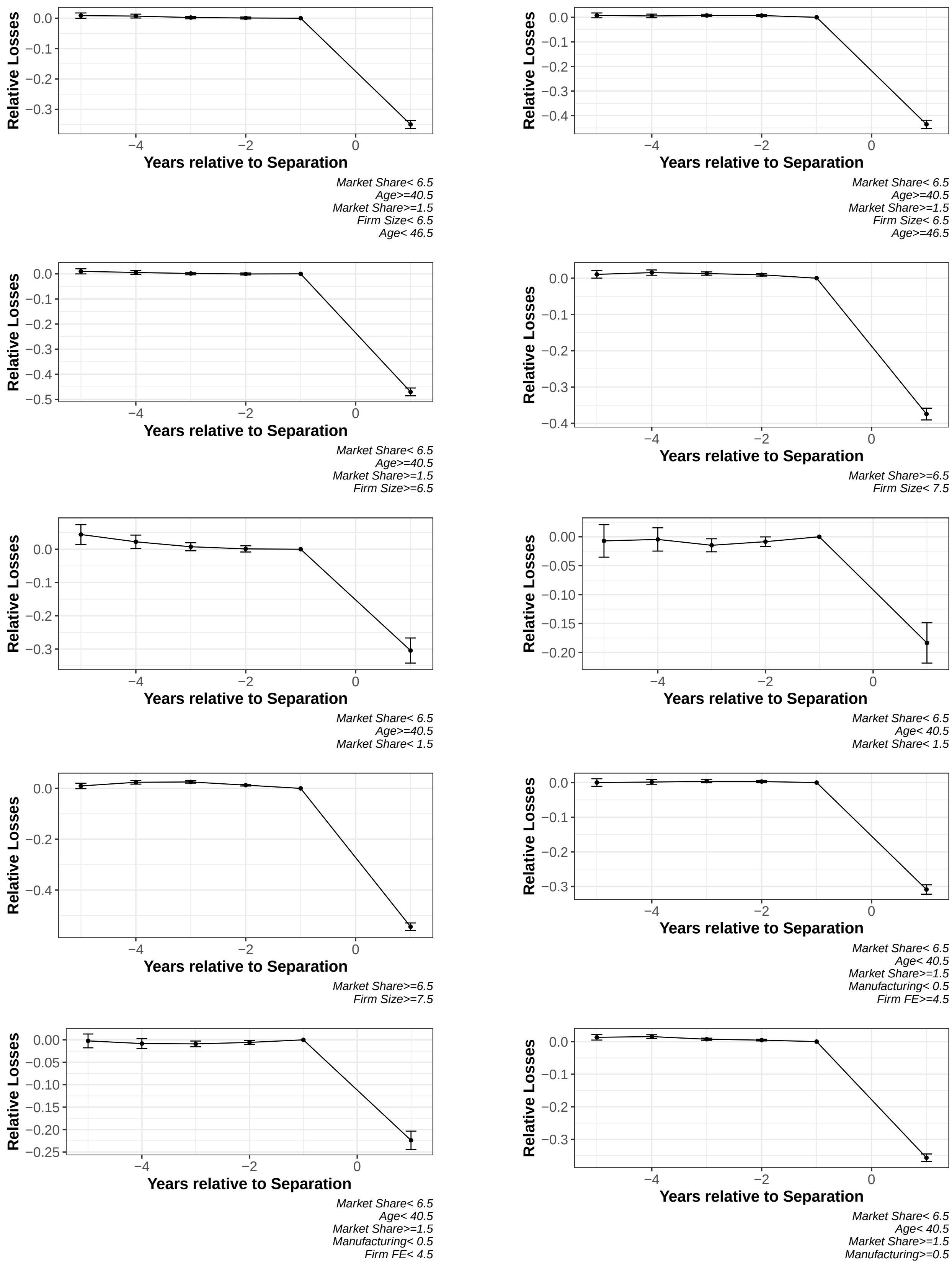


Figure 12: Event-study coefficients for each terminal node of the decision tree in Figure 1. Each panel shows relative earnings losses over time for a distinct subgroup. The selection criteria leading to each node are shown in the bottom right of each panel.

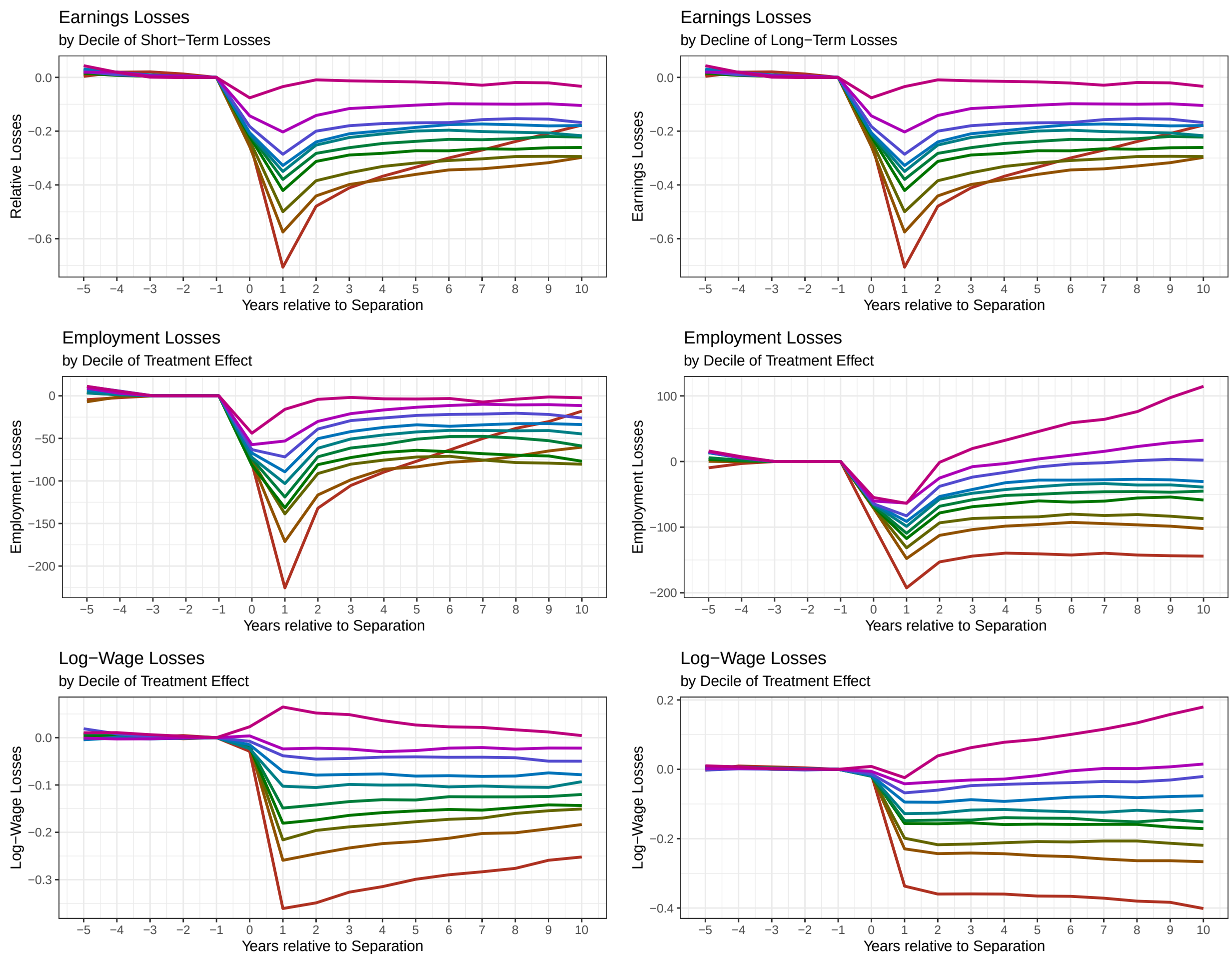


Figure 13: Event studies by predicted loss deciles. Estimated event study coefficients for groups defined by predicted loss deciles. No systematic association is observed between pre-treatment fixed effects and the magnitude of estimated losses.

D. Accuracy of the Random Forest

How accurate are the estimates of the random forest? Evaluating the accuracy is not a straightforward task. We are not estimating an observed outcome, but a treatment effect. Thus, there is no ground truth which we can use to evaluate the estimates. To nevertheless provide a measure of accuracy, we execute the following exercise. First, using the estimates by our random forest, we bin individuals into 50 groups based on their estimated earnings losses. For each of these groups, we separately estimate equation (1) using OLS. We then compare the OLS earnings loss estimates with the results of our random forest. Figure 14 compares the rank correlation between the two approaches. Both, the OLS estimates and the random forest rank the groups almost in the same way, the rank correlation above 0.988 for each forest. Figure 15 plots the OLS estimates against the earnings loss estimates from the random forest. The correlations are with around 0.95 equally high. A closer inspection reveals that the earnings loss estimates from the random forest are somewhat regularized, meaning that the OLS estimates suggest a higher level of heterogeneity. We think of this as a feature, rather than a shortcoming. OLS is going to overfit towards outliers, whereas the bootstrapping estimation procedure of the random forest is only picking up heterogeneity that consistently occurs across the bootstrapped samples (bags). Put differently, the random forest only identifies predictable heterogeneity.

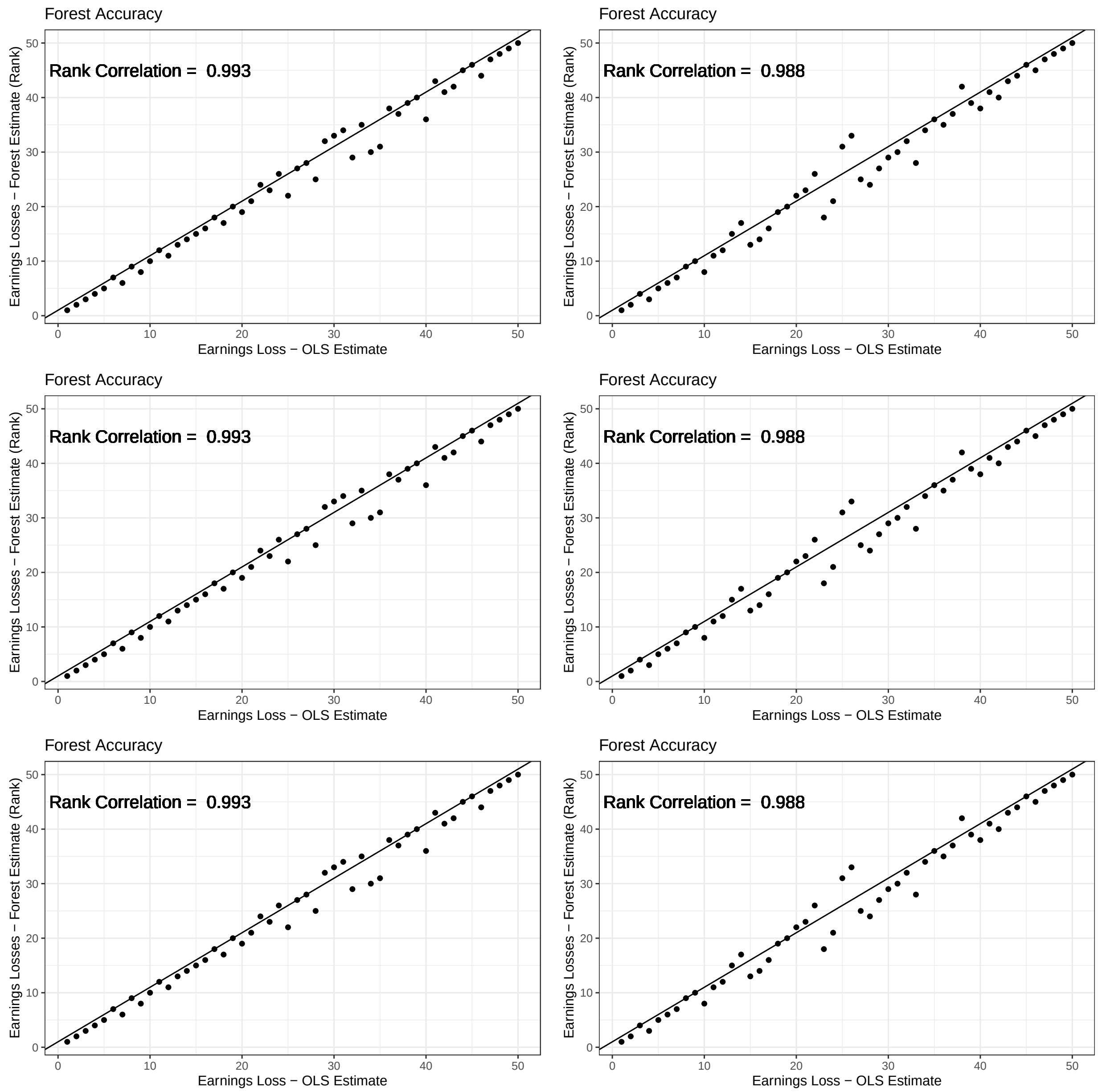


Figure 14: Bin scatter plot of random forest accuracy. We bin all individuals by their estimated treatment effects into 50 bins. For these 50 subgroups, we compute the OLS regression and plot the rank of estimated cost of job displacement against the rank of average forest estimates

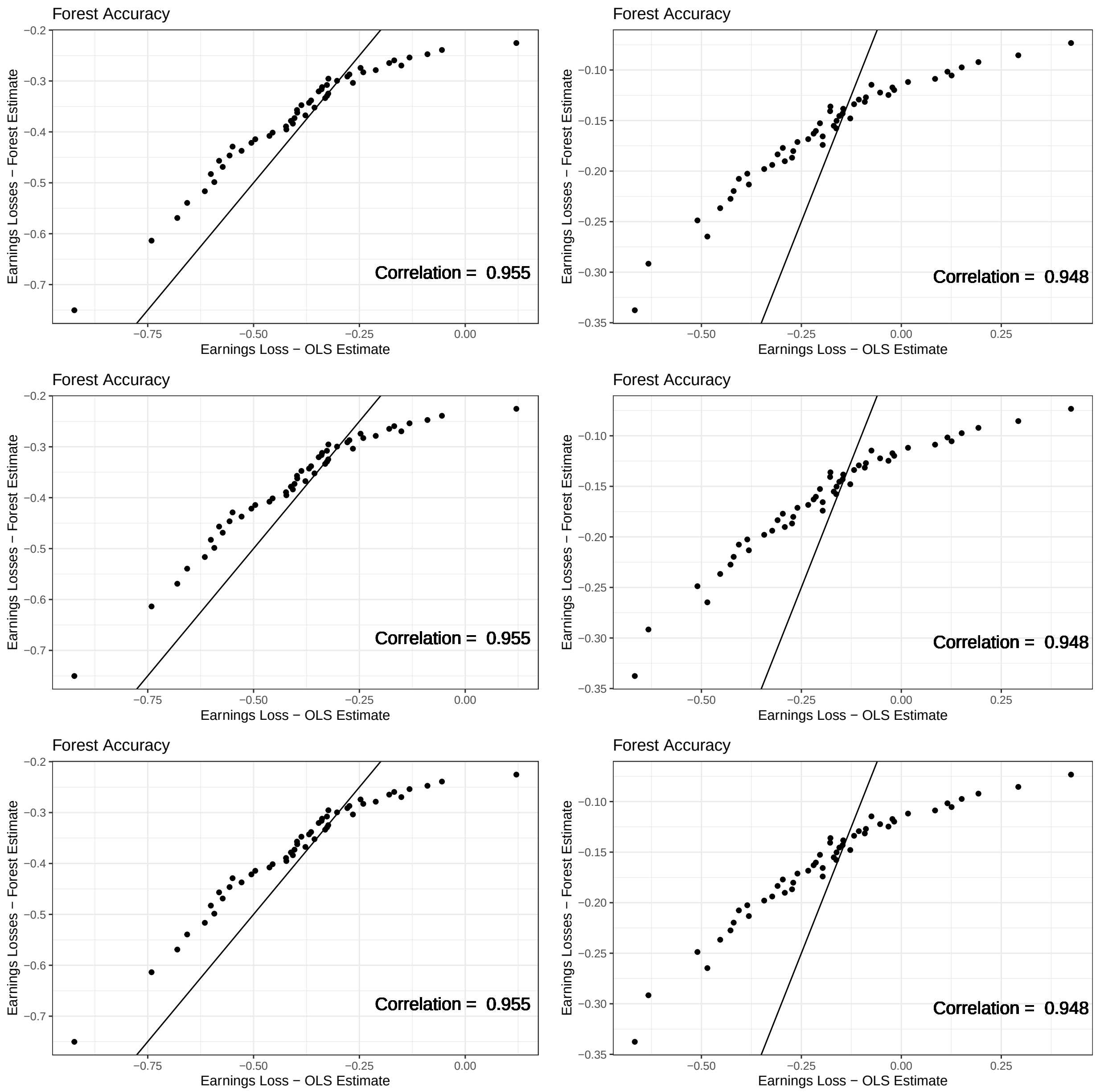


Figure 15: Bin scatter plot of random forest accuracy. We bin all individuals by their estimated treatment effects into 50 bins. For these 50 subgroups, we compute the OLS regression and plot the estimated cost of job displacement against the average forest estimates

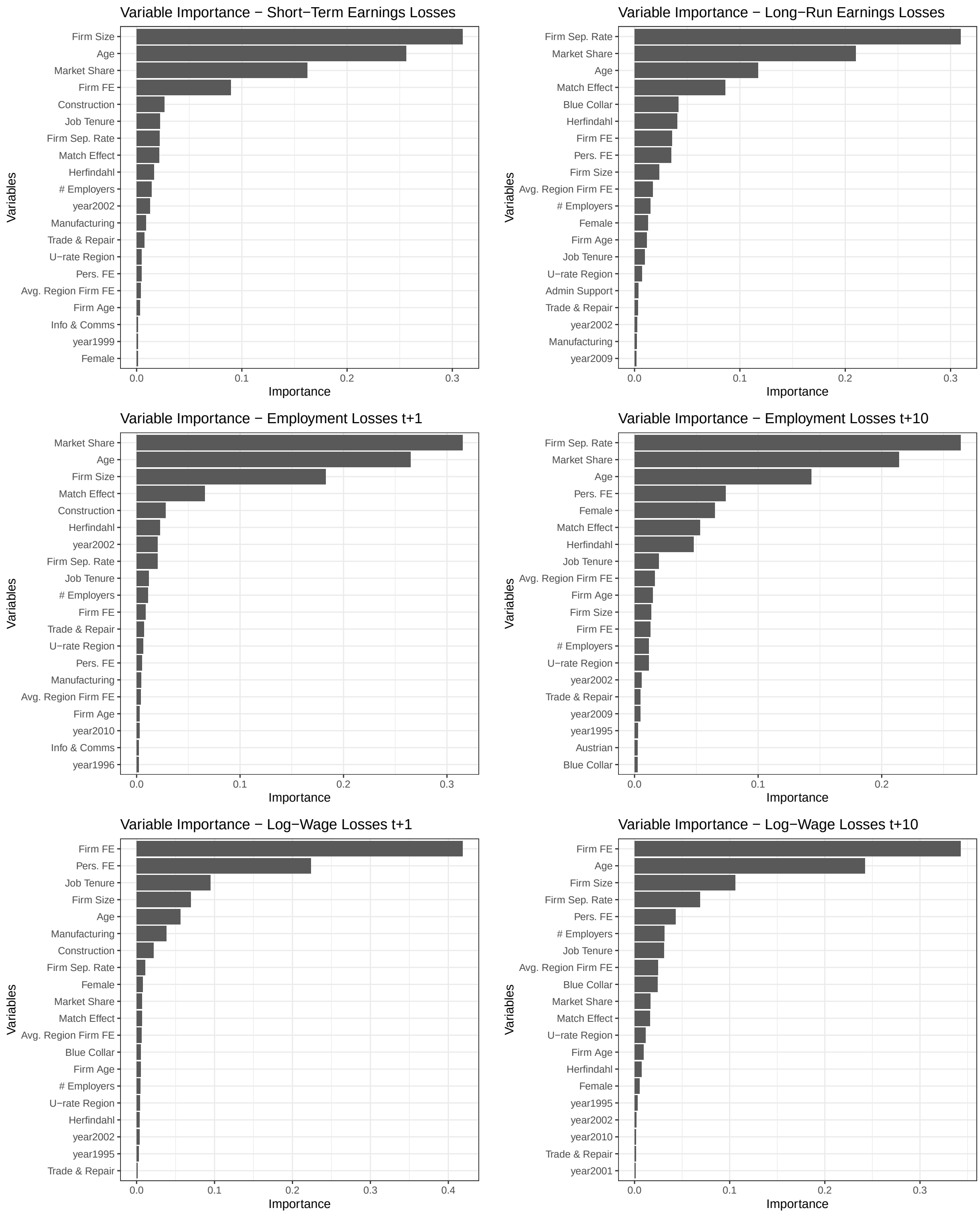


Figure 16: Variable importance from GRF with a decay exponent equal to -2 and the maximum depth level of nodes equal to 4. All values sum to 1.

E. Partial Effects

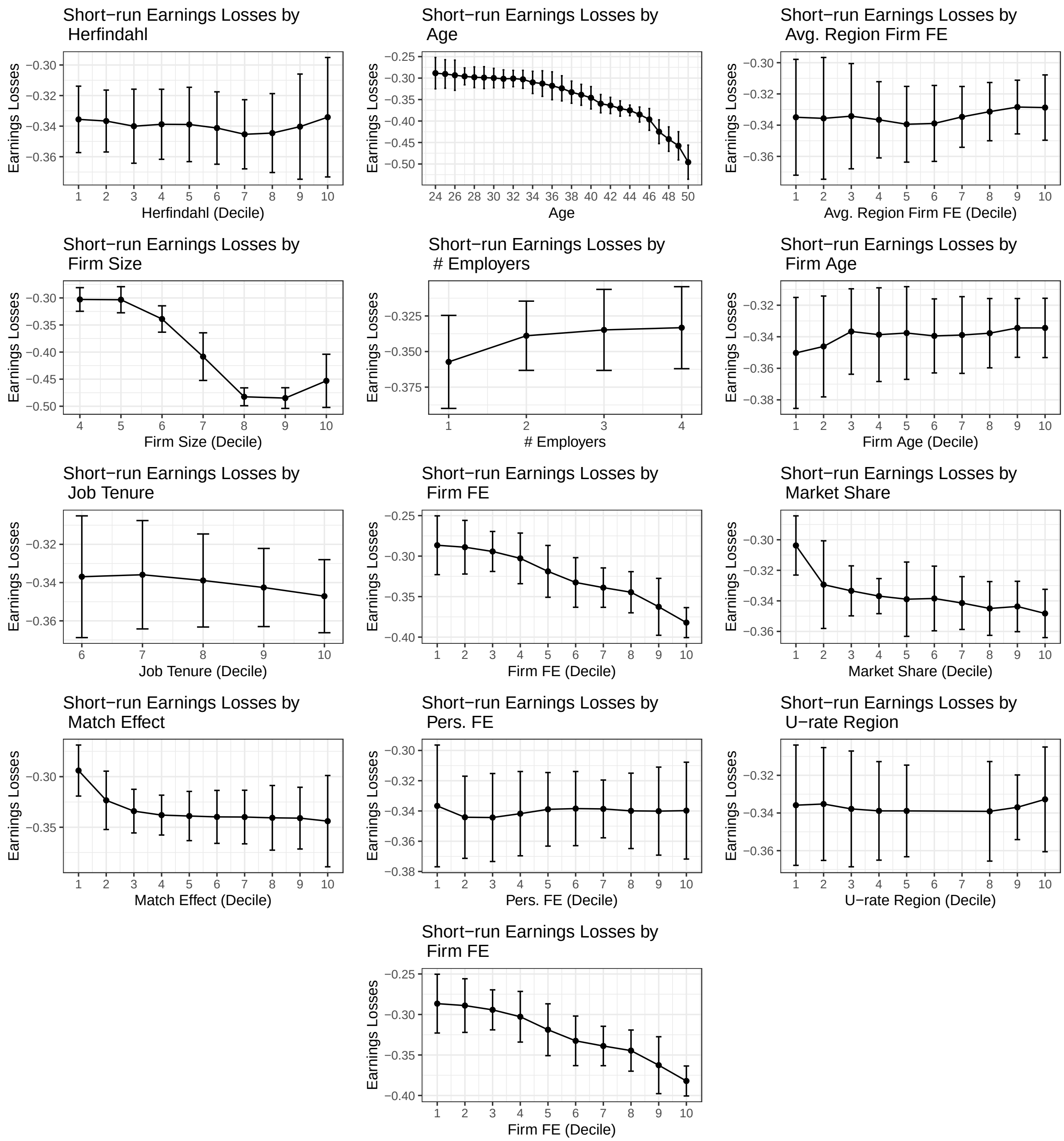


Figure 17: GRF estimates with 95% CI of short-run $(t + 1)$ earnings losses by partitioning variables. All other variables are set to their median values.

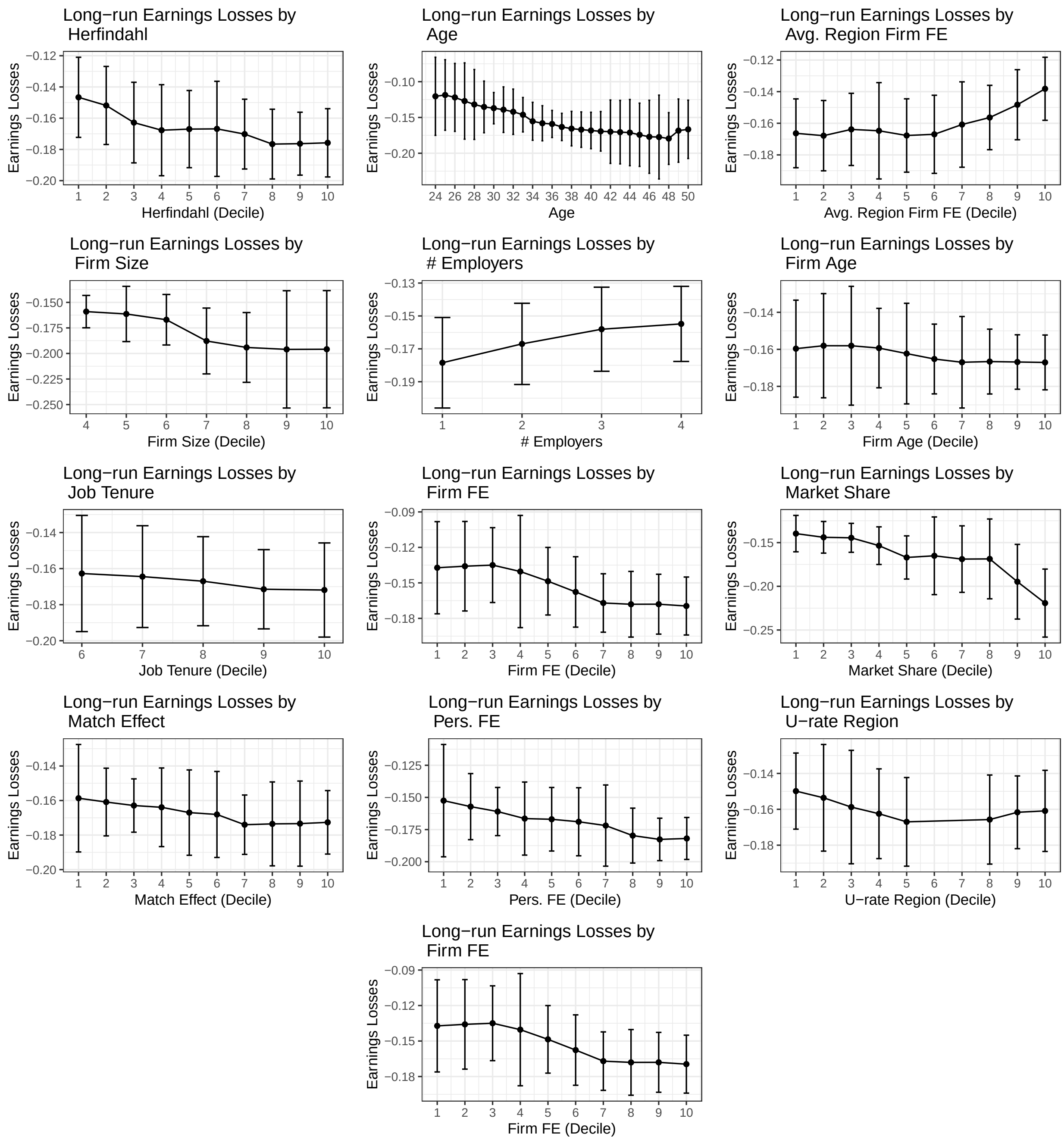


Figure 18: GRF estimates with 95% CI of long-run ($t + 10$) earnings losses by partitioning variables. All other variables are set to their median values.

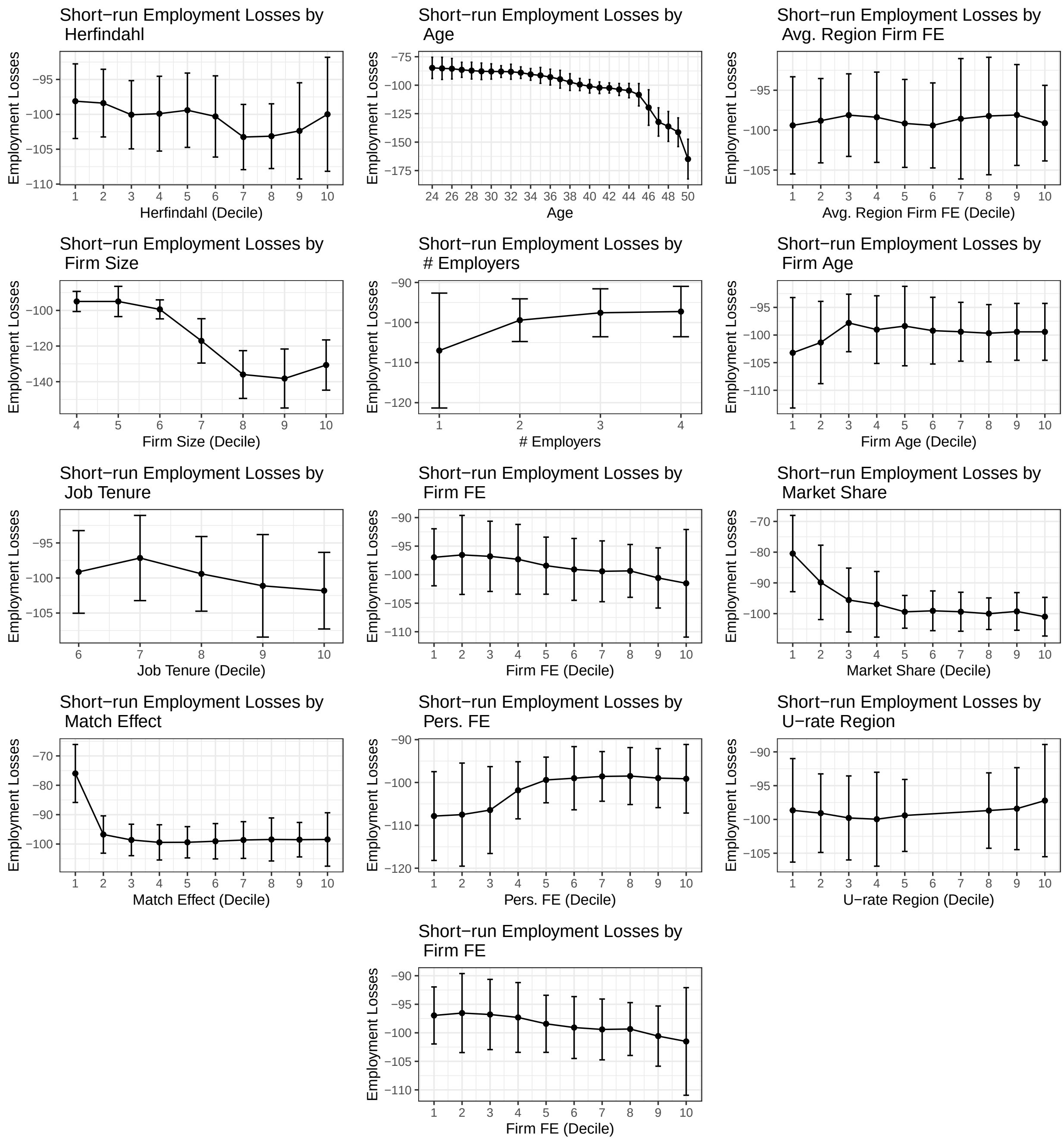


Figure 19: GRF estimates with 95% CI of short-run ($t+1$) employment losses by partitioning variables. All other variables are set to their median values.

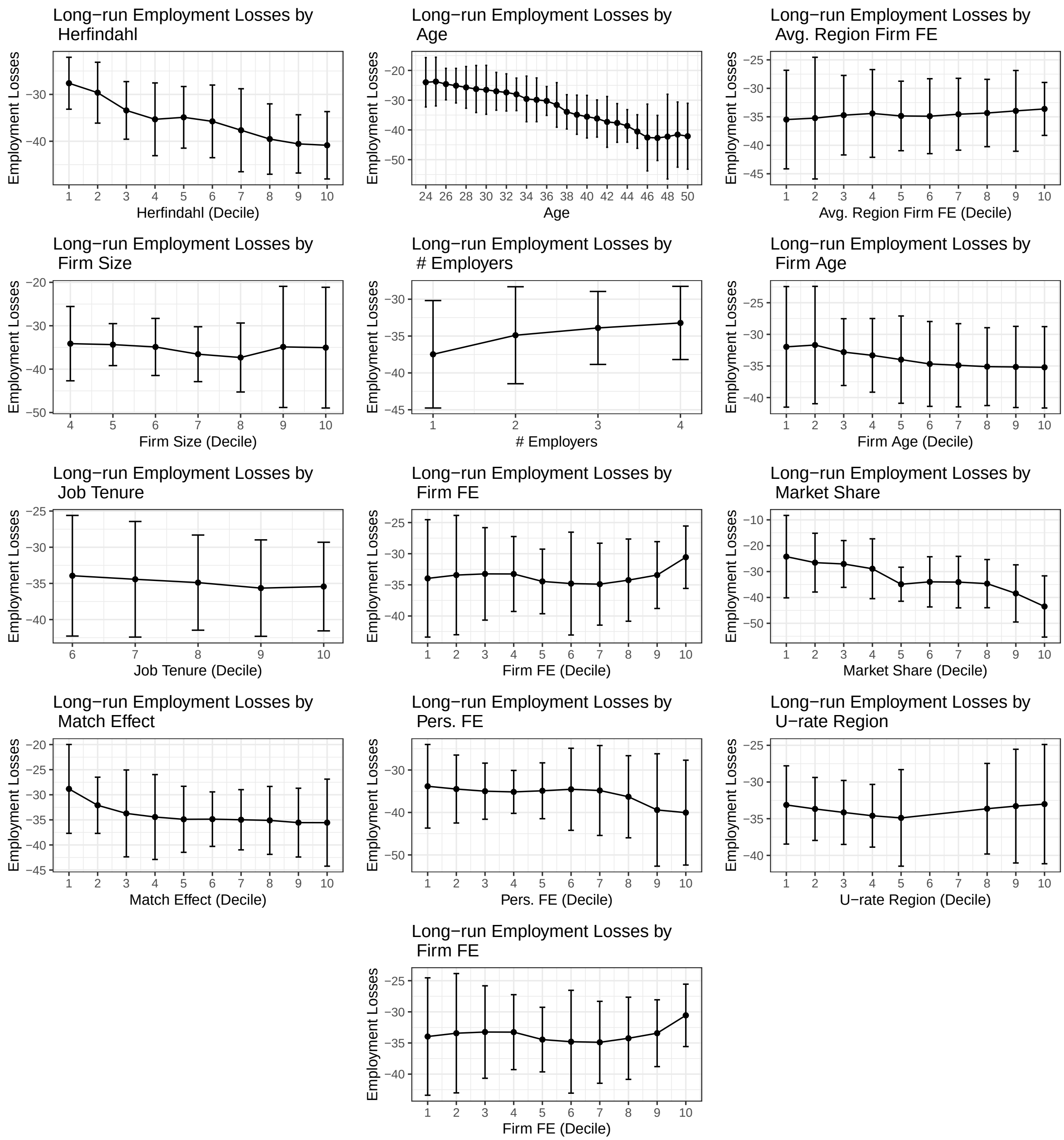


Figure 20: GRF estimates with 95% CI of long-run ($t+10$) employment losses by partitioning variables. All other variables are set to their median values.

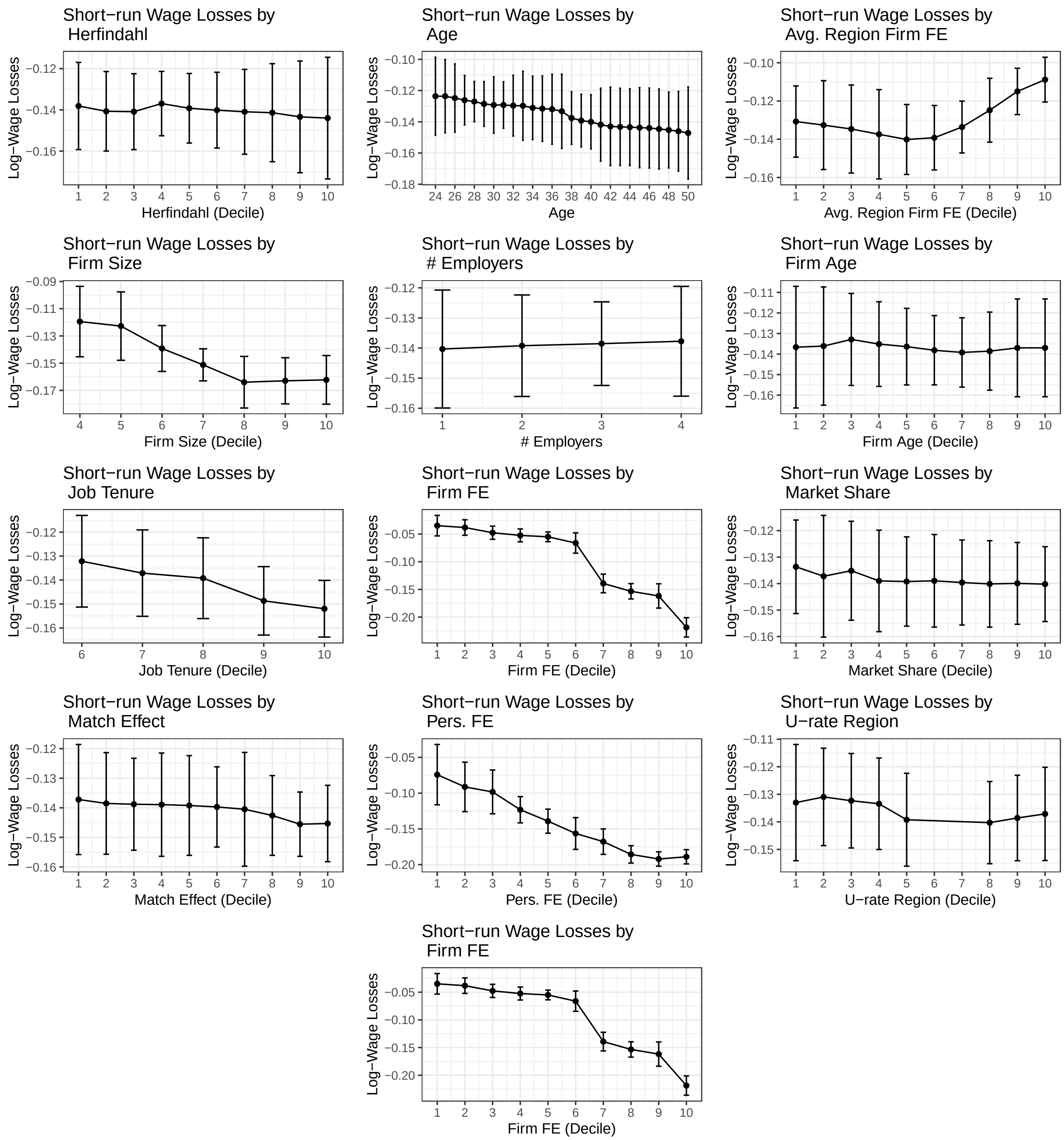


Figure 21: GRF estimates with 95% CI of long-run $(t+1)$ wage losses by partitioning variables. All other variables are set to their median values.

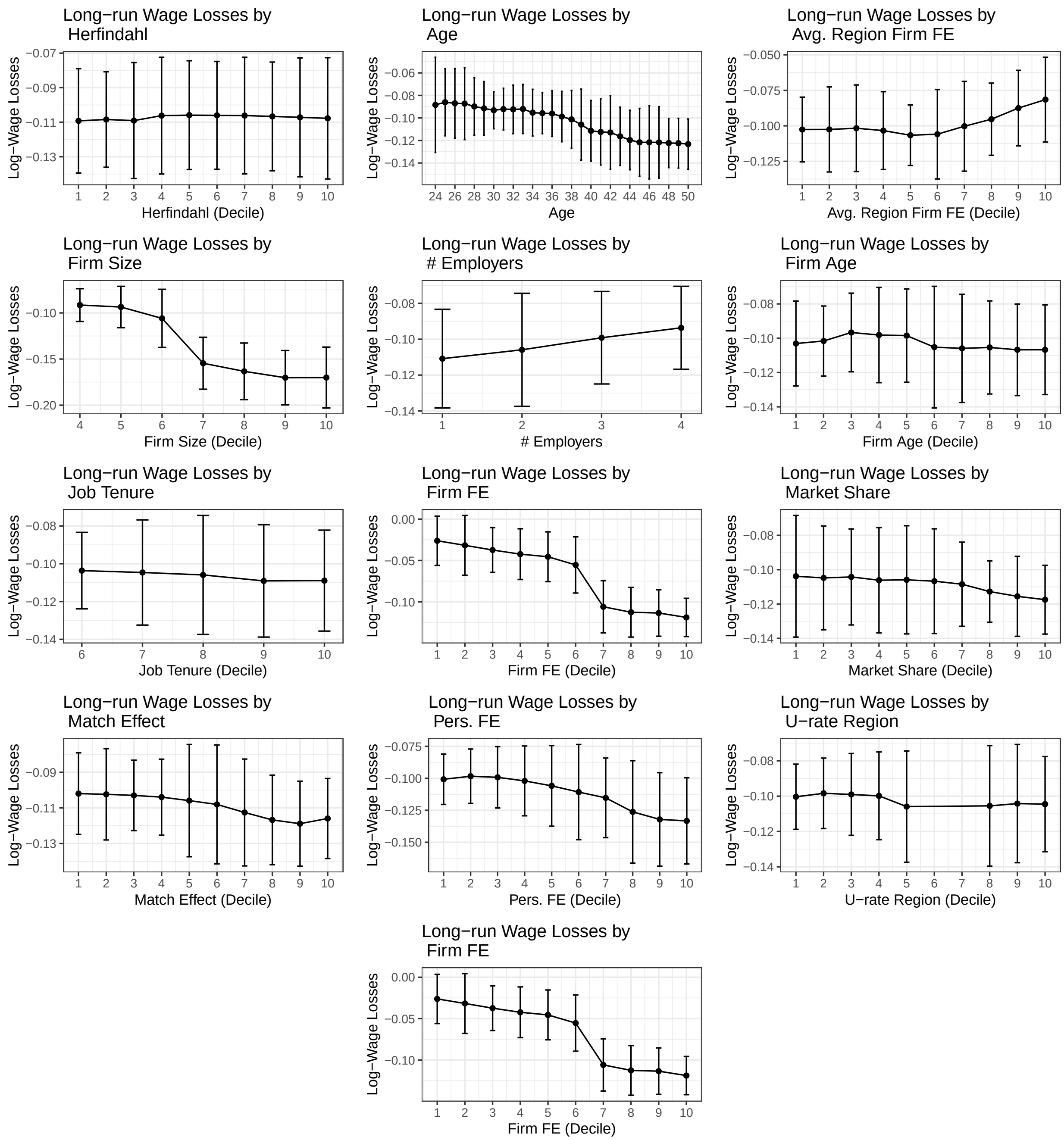


Figure 22: GRF estimates with 95% CI of long-run ($t + 10$) wage losses by partitioning variables. All other variables are set to their median values.

F. Partial Dependence Plots

One potential criticism of the partial effects is that an individual, with median characteristics might not be representative for the whole population, and those effects might be very different from median realizations. To tackle this critique, we use partial dependence plots proposed by Friedman (2001) to better understand how a single variable affects *on average* the earnings losses in the sample. This approach consists in estimating the earnings losses for each individual by changing the value of one variable $z^k = \overline{z}$, while holding all other characteristics constant at their empirical values $\mathbf{z}^{-k}$. The counterfactual outcomes are then obtained by averaging over the sample distribution $F(\mathbf{z}^{-k})$. Formally we compute:

$$\mathbb{E}_{\mathbf{z}_{-k}}\hat{\tau}(z_k = \overline{z}; \mathbf{z}_{-k}) = \int \hat{\tau}(z^k = \overline{z}; \mathbf{z}^{-k})\mathrm{d}F(\mathbf{z}^{-k}), \tag{13}$$

which in our application can be estimated on our training set: $\frac{1}{N}\sum_{i=1}^{N}\hat{\tau}(z^k = \overline{z}; \mathbf{z}_i^{-k})$. Figure 23 depicts partial dependence plots of losses in earnings, employment, and wages for the endpoints of all partitioning variables. All the main findings from the previous exercise preserve.

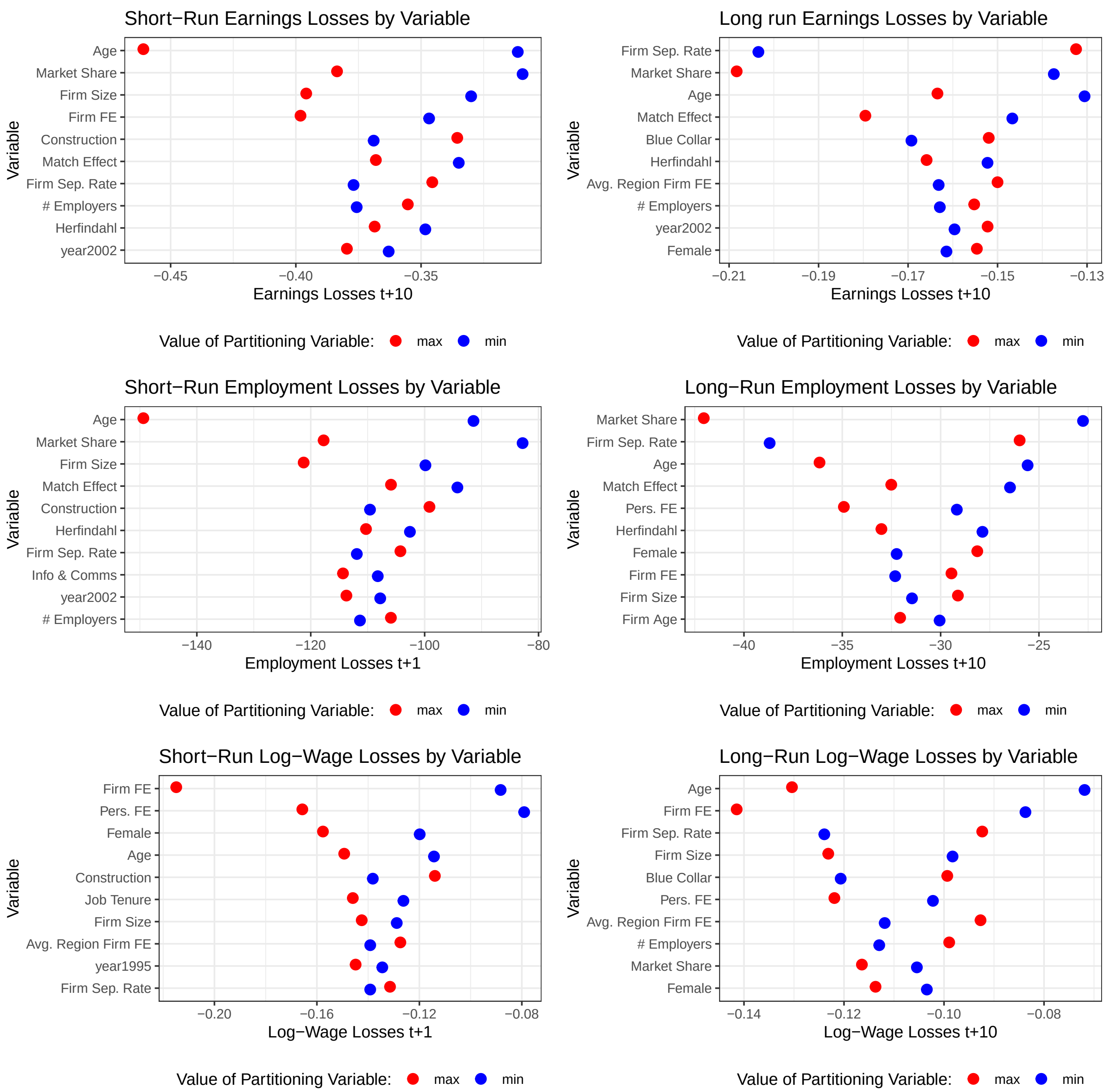


Figure 23: 10 most important correlates of earnings, employment, and wage losses. Estimates from a generalized random forest. Partial dependence plot, see text for details.

G. Heterogeneity Detection: Comparison with Other Methods

One strategy for studying heterogeneity in earnings losses is to interact coefficients with observable individual characteristics. However, if we were to interact all values of all partitioning variables, we would end up estimating a model with 64,795 parameters. This approach risks overfitting and potentially yields very noisy estimates. To address this problem, we implement LASSO (Tibshirani, 1996) penalization on all coefficients interacted with dummies associated with partitioning variables:

$$
\begin{aligned}
y_{it} \;=\; & \tau_h(\mathbf{z}_i)D_i \times \mathbb{1}(t=t^*+h) + \nu_h(\mathbf{z}_i)\mathbb{1}(t=t^*+h) + \theta_1(\mathbf{z}_i)D_i + \theta_0(\mathbf{z}_i) \\
& + \sum_{j=-5}^{-2} \nu_j(\mathbf{z}_i)\mathbb{1}(t=t^*+j) + \sum_{j=-5}^{-2} \delta_j \mathbb{1}(t=t^*+j) \times D_i \\
& + \sum_m \sum_{r\in v(m)} X_{im}^r \Big[\tau_{hm}^r D_i \times \mathbb{1}(t=t^*+h) + \nu_{hm}^r \mathbb{1}(t=t^*+h) + \theta_{1m}^r D_i + \theta_{0m}^r \\
& + \sum_{j=-5}^{-2} \nu_{jm}^r \mathbb{1}(t=t^*+j) + \sum_{j=-5}^{-2} \delta_{jm}^r \mathbb{1}(t=t^*+j) \times D_i\Big] + \epsilon_{it}, \qquad (14)
\end{aligned}
$$

where X_{im}^r denotes the dummy variable for worker i's characteristic m associated with value r, and $v(m)$ denotes the set of possible values for characteristic m. The parameters τ_{hm}^r, ν_{hm}^r, θ_{1m}^r, θ_{0m}^r, ν_{jm}^r, and δ_{jm}^r are estimated with LASSO regularization:

$$
\begin{aligned}
\min_{\boldsymbol{\beta}} \frac{1}{2N}\sum_{i=1}^{n}\sum_t \Big[y_{it} - f(X_{it},\boldsymbol{\beta})\Big]^2 + \lambda \cdot \Big[\sum_m \sum_{r\in v(m)} \Big(|\tau_{hm}^r| + |\nu_{hm}^r| \\
+|\theta_{1m}^r| + |\theta_{0m}^r| + \sum_{j=-5}^{-2}|\nu_{jm}^r| + \sum_{j=-5}^{-2}|\delta_{jm}^r|\Big)\Big]. \qquad (15)
\end{aligned}
$$

Here, $f(X_{it},\boldsymbol{\beta})$ represents the right-hand side of equation (14) excluding the error term, $\boldsymbol{\beta}$ is the vector of all parameters, and $\lambda = 1.4924e-4$ is the regularization parameter that controls the strength of the LASSO penalty. This value was determined to minimize the Akaike information criterion.[29] Overall, the regularized model reduces the number of parameters from 64,795 for unrestricted OLS to 5,402. Using a bootstrap procedure (with 200 repetitions and with sampling at the worker level), we then estimate earnings losses for

[29] While cross-validation may appear as the more conventional choice for regularization, we use information criteria due to computational constraints, an approach widely adopted for tuning LASSO in the literature dealing with large datasets (*e.g.,* Gentzkow, Shapiro, and Taddy, 2019; Pytka ⓡ Runge, 2025).

each individual and calculate 95% confidence intervals.