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

Andreas Gulyas and Krzysztof Pytka (U Mannheim)

June 29, 2021

NBP Summer Workshop 2021

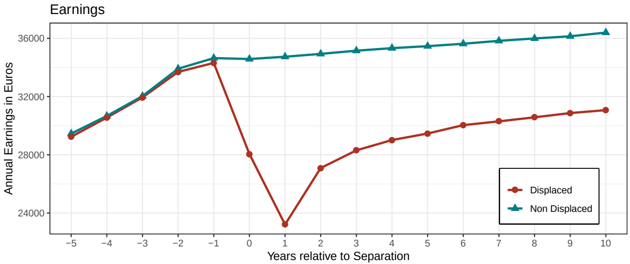
Motivation

- Job loss has long lasting and persistent detrimental effects on workers
- To design effective policy responses, we need to understand:
    1. How do earnings losses differ across individuals?
    2. What are the sources of large and persistent earnings losses?
    3. Why do earnings losses vary over the business cycle?

Motivation

- Job loss has long lasting and persistent detrimental effects on workers
- To design effective policy responses, we need to understand:
    1. How do earnings losses differ across individuals?
    2. What are the sources of large and persistent earnings losses?
    3. Why do earnings losses vary over the business cycle?

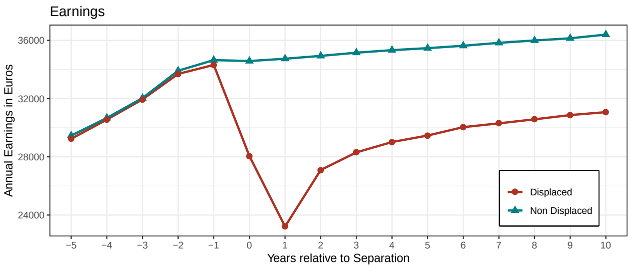
Big Data: Universe of Austrian social security records from 1984-2019



Motivation

- Job loss has long lasting and persistent detrimental effects on workers
- To design effective policy responses, we need to understand:
    1. How do earnings losses differ across individuals?
    2. What are the sources of large and persistent earnings losses?
    3. Why do earnings losses vary over the business cycle?

Big Data: Universe of Austrian social security records from 1984-2019



This paper

We use machine learning to answer all these & more questions

Methodology

Typical approaches for studying heterogeneity:

Structural models	➔	losses driven by *assumed* theories.
Reduced-form models	➔	cut by variables motivated by *assumed* theories.

Methodology

Typical approaches for studying heterogeneity:

Structural models	➔	losses driven by *assumed* theories.
Reduced-form models	➔	cut by variables motivated by *assumed* theories.

We want to be theory-agnostic but at the same time not theory-blind.

Methodology

Typical approaches for studying heterogeneity:

Structural models	➔	losses driven by *assumed* theories.
Reduced-form models	➔	cut by variables motivated by *assumed* theories.

We want to be theory-agnostic but at the same time not theory-blind.

- Adapt a **machine-learning algorithm** (Athey et al. 2019) to **difference-in-difference setting**:

  $\Rightarrow$ Estimate causal cost of job loss as a function of worker and job characteristics + SE

Methodology

- How do earnings/employment/wage/firm wage losses vary across individuals?
- Conceptually, we estimate

$$y_{it} = \tau(z_i)\mathbb{1}(t \geq t^*) \times D_i + \theta(z_i)D_i + \gamma_t(z_i) + \epsilon_{it}.$$

Sample construction

- Goal: Estimate how causal cost of job loss varies with explanatory (partitioning variables) z, i.e. $\tau(z)$
- Explanatory variables: Firm wage premia (AKM), firm separation rate, job tenure, recession indicator, regional factors, firm characteristics, match effect (residual), socio-demographics

Contribution to the Literature

- Estimate large and persistent earnings losses for Austria

  Jacobson et al. (1993), Davis and Von Wachter (2011), Halla et al. (2018)

- Comprehensive study of channels behind losses

  Lachowska et al. (2020), Schmieder et al. (2020)

- Provide guidance for structural models

  Burdett, Carrillo-Tudela, and Coles (2020); Jarosch (2021); Gregory, Menzio, and Wiczer (2021)

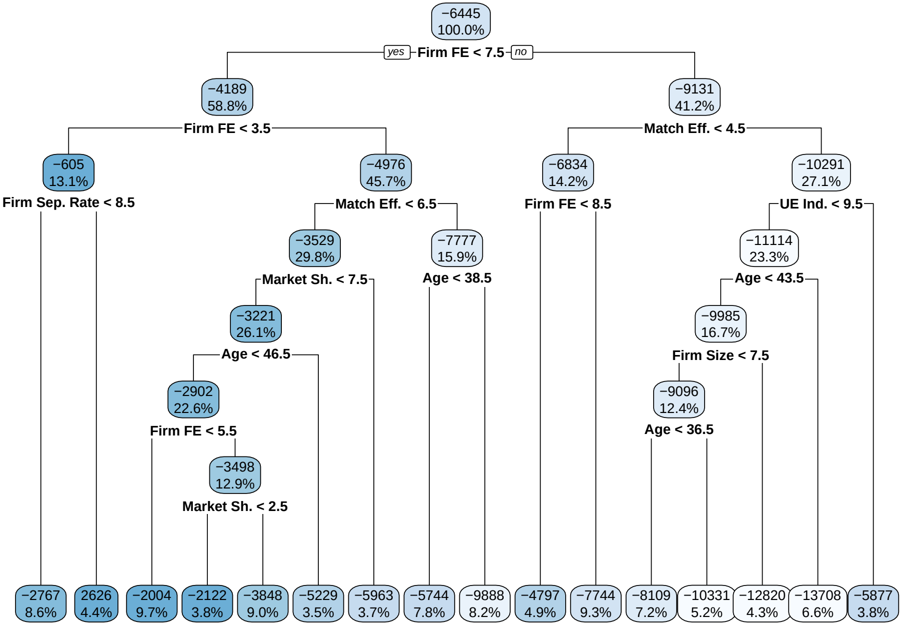
- Machine-Learning in economics:

  Athey et al. (2019); Athey and Imbens (2016)

Preview of the Main Results

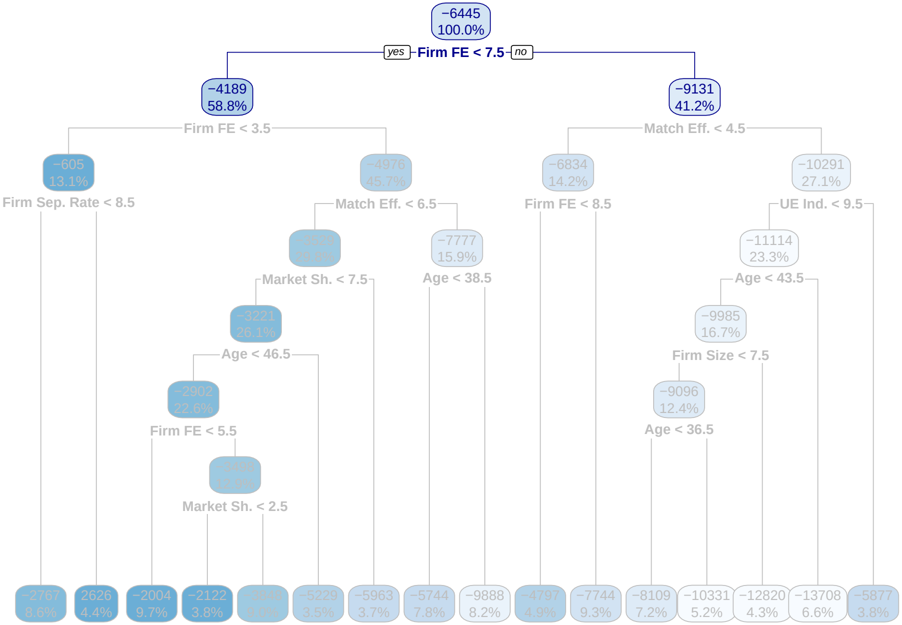
- Large differences in earnings losses across individuals:
    1. A quarter of workers lose 30% on average in terms of wages
    2. A quarter of workers gain on average
- Jointly estimate how losses vary with 18 variables ⇒ **Horse-race** between theories
- Most important factor: **Firm wage premia** (AKM)
    - Explains **42%** of **variation** in log-wage losses
- Earnings losses are cyclical:
    1. 90% explained by **Composition Effect**: Different workers lose jobs in booms/recessions
    2. 10% explained by **Pure Recession Effect**: Identical workers face higher losses in recessions
- Gender differences fully explained by composition
- Implications: Labor market policies aiming to mitigate cost of job loss should be **targeted** and rather **time-invariant**

Causal Tree



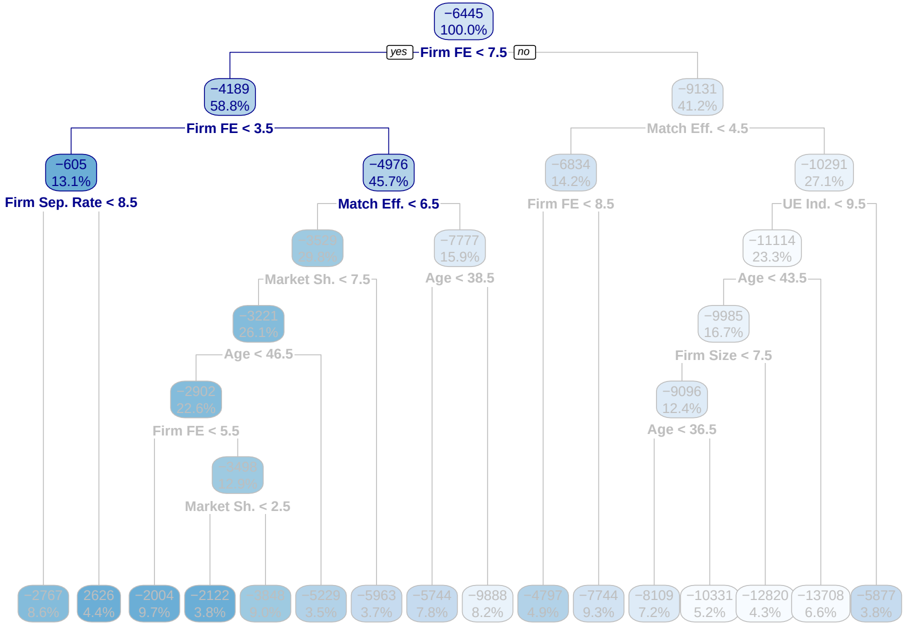
−6445
100.0%
yes
Firm FE < 7.5
no
−4189
58.8%
Firm FE < 3.5
−9131
41.2%
Match Eff. < 4.5
−605
13.1%
Firm Sep. Rate < 8.5
−4976
45.7%
Match Eff. < 6.5
−6834
14.2%
Firm FE < 8.5
−10291
27.1%
UE Ind. < 9.5
−3529
29.8%
Market Sh. < 7.5
−7777
15.9%
Age < 38.5
−11114
23.3%
Age < 43.5
−3221
26.1%
Age < 46.5
−9985
16.7%
Firm Size < 7.5
−2902
22.6%
Firm FE < 5.5
−9096
12.4%
Age < 36.5
−3498
12.9%
Market Sh. < 2.5
−2767
8.6%
2626
4.4%
−2004
9.7%
−2122
3.8%
−3848
9.0%
−5229
3.5%
−5963
3.7%
−5744
7.8%
−9888
8.2%
−4797
4.9%
−7744
9.3%
−8109
7.2%
−10331
5.2%
−12820
4.3%
−13708
6.6%
−5877
3.8%

Causal Tree

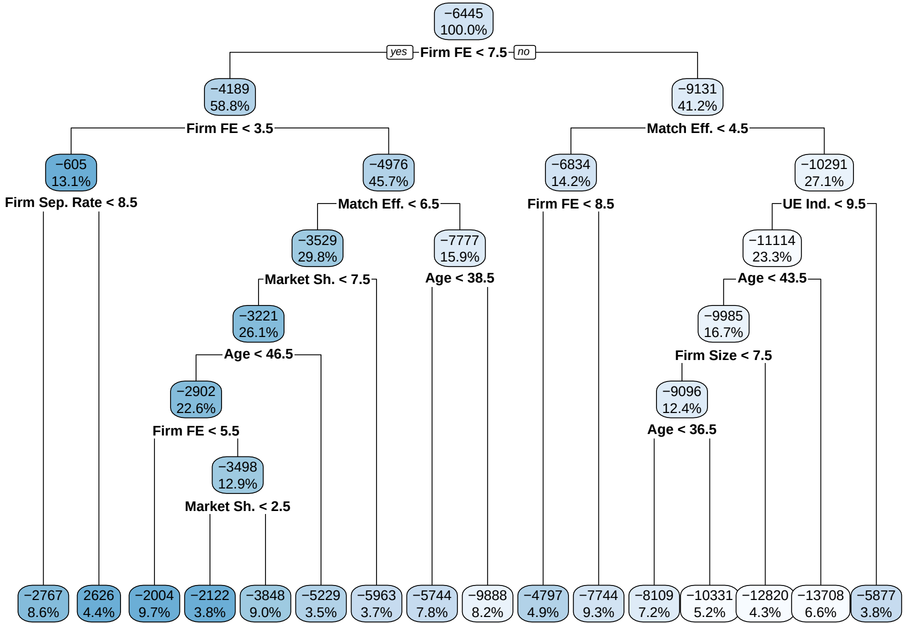


−6445
100.0%
yes
Firm FE < 7.5
no
−4189
58.8%
−9131
41.2%
Firm FE < 3.5
Match Eff. < 4.5
−605
13.1%
−4976
45.7%
−6834
14.2%
−10291
27.1%
Firm Sep. Rate < 8.5
Match Eff. < 6.5
Firm FE < 8.5
UE Ind. < 9.5
−3529
29.8%
−7777
15.9%
−11114
23.3%
Market Sh. < 7.5
Age < 38.5
Age < 43.5
−3221
26.1%
−9985
16.7%
Age < 46.5
Firm Size < 7.5
−2902
22.6%
−9096
12.4%
Firm FE < 5.5
Age < 36.5
−3498
12.9%
Market Sh. < 2.5
−2767
8.6%
2626
4.4%
−2004
9.7%
−2122
3.8%
−3848
9.0%
−5229
3.5%
−5963
3.7%
−5744
7.8%
−9888
8.2%
−4797
4.9%
−7744
9.3%
−8109
7.2%
−10331
5.2%
−12820
4.3%
−13708
6.6%
−5877
3.8%

Causal Tree


−6445
100.0%
yes
Firm FE < 7.5
no
−4189
58.8%
Firm FE < 3.5
−9131
41.2%
Match Eff. < 4.5
−605
13.1%
Firm Sep. Rate < 8.5
−4976
45.7%
Match Eff. < 6.5
−6834
14.2%
Firm FE < 8.5
−10291
27.1%
UE Ind. < 9.5
Market Sh. < 7.5
−7777
15.9%
Age < 38.5
−11114
23.3%
Age < 43.5
−3221
26.1%
Age < 46.5
−9985
16.7%
Firm Size < 7.5
−2902
22.6%
Firm FE < 5.5
−9096
12.4%
Age < 36.5
Market Sh. < 2.5
−2767
8.6%
2626
4.4%
−2004
9.7%
−2122
3.8%
−5229
3.5%
−5963
3.7%
−5744
7.8%
−9888
8.2%
−4797
4.9%
−7744
9.3%
−8109
7.2%
−10331
5.2%
−12820
4.3%
−13708
6.6%
−5877
3.8%

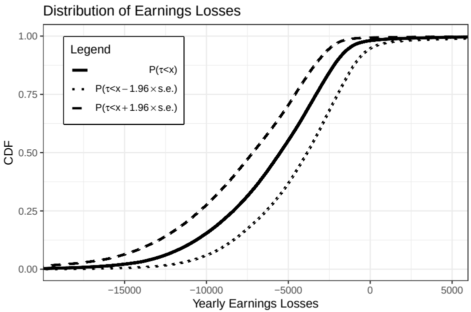
Causal Tree


−6445
100.0%
yes
Firm FE < 7.5
no
−4189
58.8%
Firm FE < 3.5
−9131
41.2%
Match Eff. < 4.5
−605
13.1%
Firm Sep. Rate < 8.5
−4976
45.7%
Match Eff. < 6.5
−6834
14.2%
Firm FE < 8.5
−10291
27.1%
UE Ind. < 9.5
−3529
29.8%
Market Sh. < 7.5
−7777
15.9%
Age < 38.5
−11114
23.3%
Age < 43.5
−3221
26.1%
Age < 46.5
−9985
16.7%
Firm Size < 7.5
−2902
22.6%
Firm FE < 5.5
−9096
12.4%
Age < 36.5
−3498
12.9%
Market Sh. < 2.5
−2767
8.6%
2626
4.4%
−2004
9.7%
−2122
3.8%
−3848
9.0%
−5229
3.5%
−5963
3.7%
−5744
7.8%
−9888
8.2%
−4797
4.9%
−7744
9.3%
−8109
7.2%
−10331
5.2%
−12820
4.3%
−13708
6.6%
−5877
3.8%

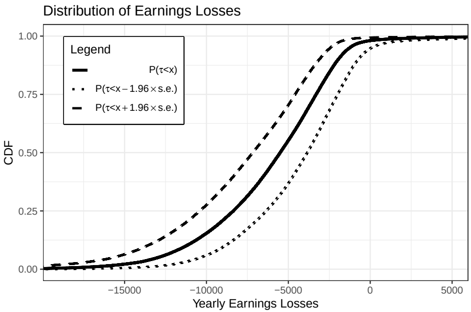
Random Forest

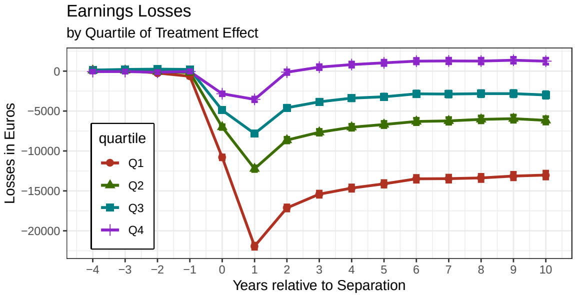
- One single tree has high variance, might be subject to overfitting
- $\Rightarrow$ Random Forest as ensemble of many (10,000) trees
    1. Sample 50 % of observations from dataset
    2. "Honest" estimation (Athey et al., 2019)
    3. Sample $\approx$1/3 of partitioning variables
    4. Grow tree as described before

Heterogeneity in Earnings Losses



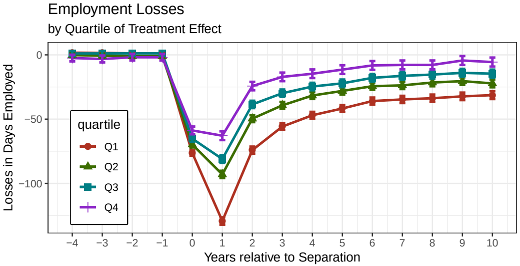
Heterogeneity in Earnings Losses

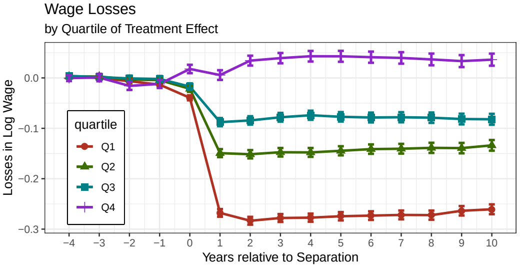




Heterogeneity in Earnings Losses: Wages and Employment

Individuals stacked by *quartiles of identified earnings losses.*





Comparative statics

- Now we know that indeed there is heterogeneity in earnings losses.

  But can we say anything about the drivers of those losses?
- Varying one channel at a time. All other variables fixed at their median.

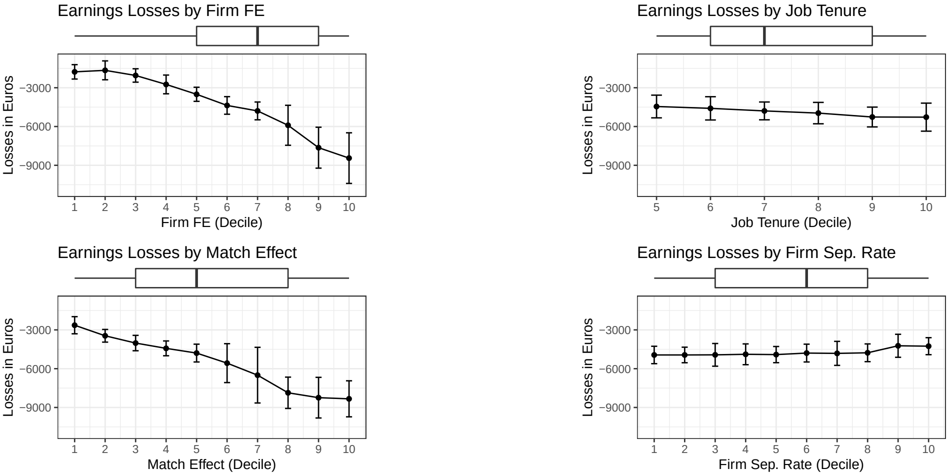
Why is it important?

Comparative statics

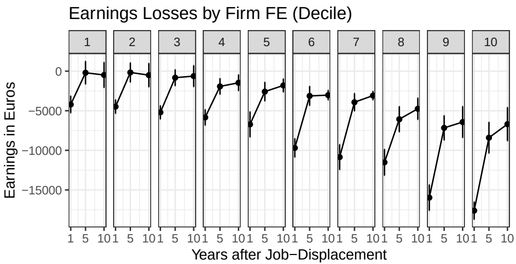
- Now we know that indeed there is heterogeneity in earnings losses.

  But can we say anything about the drivers of those losses?
- Varying one channel at a time. All other variables fixed at their median.

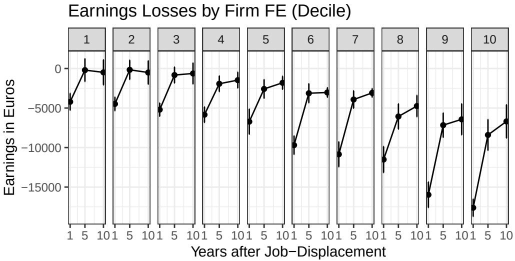
Why is it important?

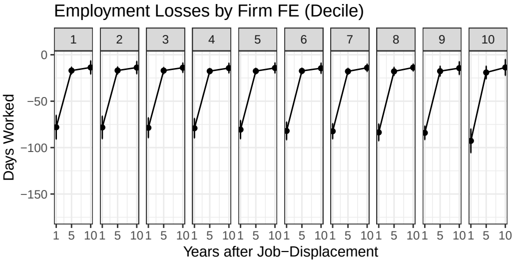


Losses in Earnings & Employment by Firm FE



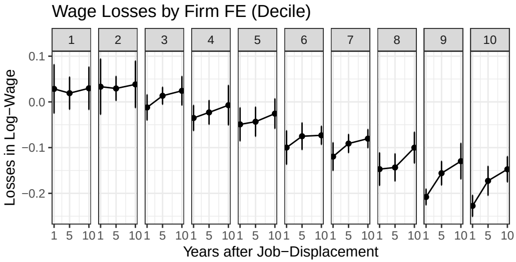
Losses in Earnings & Employment by Firm FE



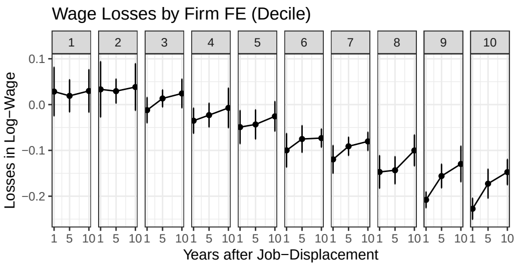


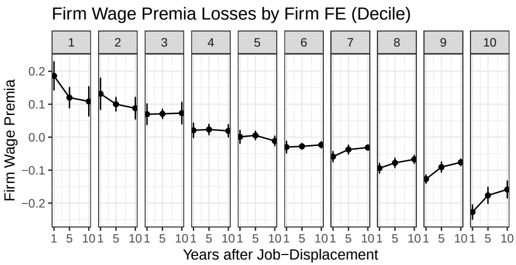
- Steep slope in earnings losses by Firm FE
- No variation in employment losses

Losses in Wages & Firm FE by Firm FE



Losses in Wages & Firm FE by Firm FE

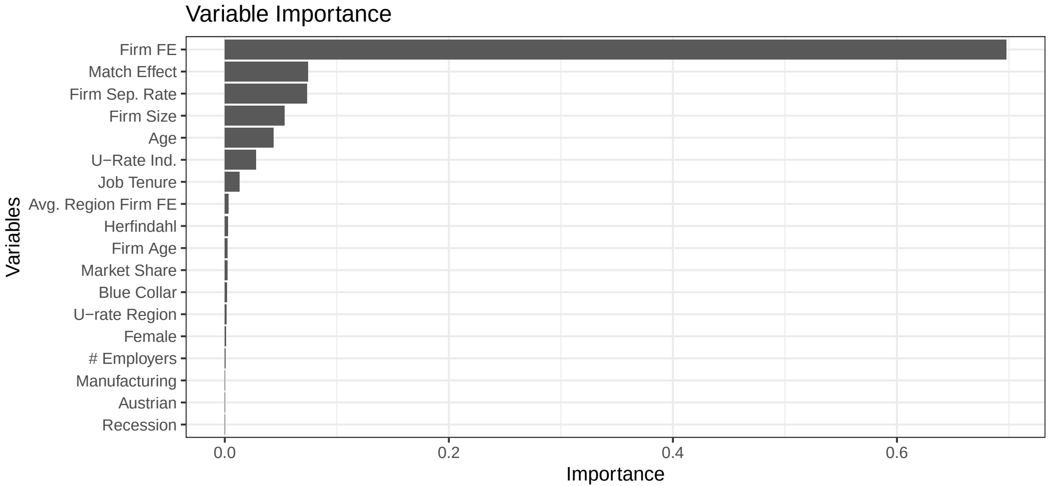




- Long-term earnings losses due to wage losses
- Workers from low paying firms see wage *gains*
- Mean reversion in firm pay

Variable Importance for Earnings Losses

Firm Wage Premia first order



Cyclicality of Losses

- Many labor market policies are enacted/extended in recessions:
  - Extended UI benefits in US
  - Firm bail-outs
  - Short time work subsidies
- Motivation: job loss has been shown to be more costly in recessions

  Davis and Von Wachter (2011), Schmieder et al. (2020)
- Is it the effect of recessions, or is it compositional differences of displaced workers?

Recession effect vs. compositional difference

Cyclicality of losses, $\int \tau(z|rec = 1)\mathrm{d}F(z|rec = 1) - \int \tau(z|rec = 0)\mathrm{d}F(z|rec = 0)$, can be decomposed in two ways:

1. recession distr.:

$$\underbrace{\int [\tau(z|rec = 1) - \tau(z|rec = 0)]\, \mathrm{d}F(z|rec = 1)}_{\text{Recession effect}}$$
$$+$$
$$\underbrace{\int \tau(z|rec = 0)\mathrm{d}F(z|rec = 1) - \int \tau(z|rec = 0)\mathrm{d}F(z|rec = 0)}_{\text{Compositional difference}}$$

2. expansion distr.:

$$\underbrace{\int \tau(z|rec = 1)\mathrm{d}F(z|rec = 1) - \int \tau(z|rec = 1)\mathrm{d}F(z|rec = 0)}_{\text{Compositional difference}}$$
$$+$$
$$\underbrace{\int [\tau(z|rec = 1) - \tau(z|rec = 0)]\, \mathrm{d}F(z|rec = 0)}_{\text{Recession effect}}$$

		Recession Effect		Composition	
	Difference	Level	Share	Level	Share
Recession dist.	-513.35	-60.00	0.12	-453.34	0.88
Expansion dist.	-513.35	-54.66	0.11	-458.69	0.89

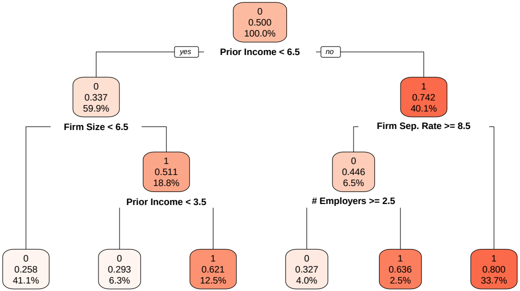
Policy Targeting

- Suppose the govt. wants to detect displaced individuals with relative wage losses above the median level.
- According to GDPR (2016/679), all people have the right to explanation, *i.e.* "meaningful information about the logic involved" in automated decisions. For this reason:
    - we project the identified log-wage losses through our forest on a single policy tree;
    - we exclude four partitioning variables *that may be difficult to understand by most people*:
        i. firm wage premia,
        ii. regional average firm wage premia,
        iii. concentration of the labor market,
        iv. and workers' match effect.

Policy Targeting (cont'd)

The algorithm detects 3 groups of individuals with high relative losses:

- Middle-income individuals displaced from larger firms.
- High-income individuals displaced from not the riskiest firms.
- High-income individuals with low job-mobility displaced from the riskiest firms.



Conclusions

- Use machine learning to understand sources of earnings losses
- Document substantial heterogeneity
  (much more than linear interactions or quantile regression)
- Losses in firm wage premia most important channel
- Cyclicality of losses driven mainly by the compositional effect
- Labor market policies should be targeted, and time-invariant

Estimate your own earnings losses:

Link: gulyas-pytka.app

Conclusions

- Use machine learning to understand sources of earnings losses
- Document substantial heterogeneity
  (much more than linear interactions or quantile regression)
- Losses in firm wage premia most important channel
- Cyclicality of losses driven mainly by the compositional effect
- Labor market policies should be targeted, and time-invariant

Estimate your own earnings losses:

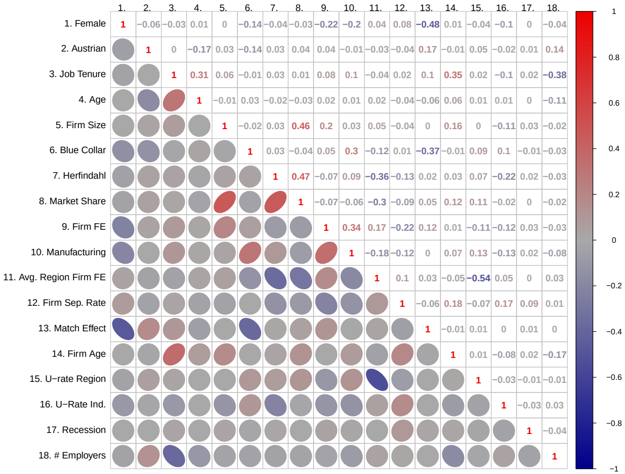
Link: gulyas-pytka.app

Thanks for your attention!

Sample Construction

- Universe of Austrian social security data 1984-2019
- Definition mass layoff event:
  1. Firm size declines by $\geq$ 30 %
  2. Firm size $\geq$ 30 employees
  3. Exclude fast growing firms before mass layoff, and firms that bounce back in size
- Sample Restrictions:
  1. Male & female workers
  2. Worker age: 25-50
  3. Job tenure $\geq$ 2 years
- Control group selected via propensity score matching

Correlogram of pre-displacement characteristics



Why to keep confounding factors fixed? (illustrative example)

- Imagine a simple economy with two sectors: *production* and *services*.
- There are equally numerous groups of workers: *high-school* educated and *college educated*.
- There is sectoral sorting:
  - 20% of the high-school educated work in services and 80% in production.
  - 80% of the college educated work in services and 20% in production.
- *Cognitive Booster Pills* increase cognitive productivy. Higher marginal effect for HS educated and services.

Table 1: Productivity increase after Pills

	Production	Services
High School	\$1.0	\$10
College degree	\$0.7	\$7

Why to keep confounding factors fixed? (illustrative example)

- Imagine a simple economy with two sectors: *production* and *services*.
- There are equally numerous groups of workers: *high-school* educated and *college educated*.
- There is sectoral sorting:
  - 20% of the high-school educated work in services and 80% in production.
  - 80% of the college educated work in services and 20% in production.
- *Cognitive Booster Pills* increase cognitive productivy. Higher marginal effect for HS educated and services.

Table 1: Productivity increase after Pills

	Production	Services
High School	\$1.0	\$10
College degree	\$0.7	\$7

Heterogeneous TEs in education subgroups:

Why to keep confounding factors fixed? (illustrative example)

- Imagine a simple economy with two sectors: *production* and *services*.
- There are equally numerous groups of workers: *high-school* educated and *college educated*.
- There is sectoral sorting:
  - 20% of the high-school educated work in services and 80% in production.
  - 80% of the college educated work in services and 20% in production.
- *Cognitive Booster Pills* increase cognitive productivy. Higher marginal effect for HS educated and services.

Table 1: Productivity increase after Pills

	Production	Services
High School	\$1.0	\$10
College degree	\$0.7	\$7

Heterogeneous TEs in education subgroups:

- Treatment effect for high-school educated: $80\% \cdot \$1 + 20\% \cdot \$10 = \$2.79$
- Treatment effect for college educated: $20\% \cdot \$0.7 + 80\% \cdot \$7 = \$5.74$

Why to keep confounding factors fixed? (illustrative example)

- Imagine a simple economy with two sectors: *production* and *services*.
- There are equally numerous groups of workers: *high-school* educated and *college educated*.
- There is sectoral sorting:
  - 20% of the high-school educated work in services and 80% in production.
  - 80% of the college educated work in services and 20% in production.
- *Cognitive Booster Pills* increase cognitive productivy. Higher marginal effect for HS educated and services.

Table 1: Productivity increase after Pills

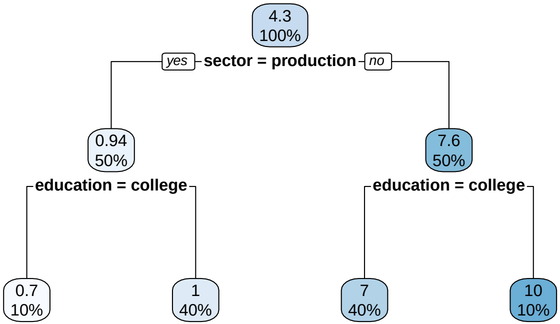
	Production	Services
High School	\$1.0	\$10
College degree	\$0.7	\$7

Heterogeneous TEs in education subgroups:

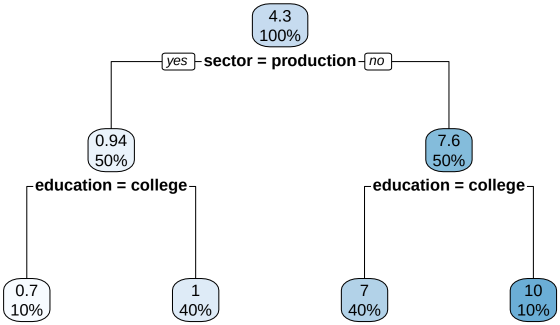
- Treatment effect for high-school educated: $80\% \cdot \$1 + 20\% \cdot \$10 = \$2.79$
- Treatment effect for college educated: $20\% \cdot \$0.7 + 80\% \cdot \$7 = \$5.74$

This simple accounting ignores compositional differences!

Why to keep confounding factors fixed? (illustrative example)



Why to keep confounding factors fixed? (illustrative example)



Case with two dummies is (intentionally) trivial. Real-world problems become more and more complicated if there are more variables. Which variables should be kept fixed and which can be ignored? ➔**task for the ML algorithm!**

Why do we need Machine Learning at all?

- Why not use interaction effects (or subsample estimations)?
- Which interaction effects to include?

  20.8 trillion! (20.8×10^{12}) possible interaction effects with our 18 variables
- Typical approach in the literature: Sample split, i.e. short vs long tenure

  Issue: tenure likely correlated with many other characteristics
- We estimate losses by varying one factor at a time, holding all other factors constant at the median.

References

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

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

Steven J Davis and Till Von Wachter. Recessions and the costs of job loss. *Brookings Papers on Economic Activity*, 43(2 (Fall)):1–72, 2011.

Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. Heterogeneous scarring effects of full-year nonemployment. *American Economic Review*, 107(5):369–73, 2017.

Martin Halla, Julia Schmieder, and Andrea Weber. Job displacement, family dynamic, and spousal labor supply. 2018.

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

Marta Lachowska, Alexandre Mas, and Stephen A. Woodbury. Sources of displaced workers' long-term earnings losses. *American Economic Review*, 110(10):3231–66, October 2020. doi: 10.1257/aer.20180652. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20180652`.

Johannes F Schmieder, Till von Wachter, and Jorg Heining. The Costs of Job Displacement over the Business Cycle and Its Sources: Evidence from Germany. 2020.