Economic Effects of Gaining a Qualification in Later-life

– PRELIMINARY –

FINN LATTIMORE* — Reserve Bank of Australia
DANIEL STEINBERG — Gradient Institute
ANNA ZHU — RMIT University, IZA

Gradient Institute Reading Group
24th August 2022

*This work was performed while the author was working at the Gradient Institute. Views expressed in this paper are those of the author and do not reflect those of the Reserve Bank of Australia

Aims

- What is the long-term impact of obtaining a new qualification (on total earnings and wages)?
- Who tends to benefit?

Motivation: Gains in education for mature-age students

- Re-training and up-skilling
  - critical in light of automation and IT-skills-biased economy (Autor et al., 2008; Acemoglu and Autor, 2011)
  - facilitates career-change
- Expectation for re-training is across the whole income and age distribution
  - but existing literature focuses on younger ages (below age 25)
  - or on one type of setting such as in community colleges
- Total earnings versus wages

Idea and Contributions

- Focus on mature-age learners (age 25 or above)
- Earnings (both total and hourly)
- Returns to gaining different types of degrees and across the income distribution
- Machine Learning (ML) techniques applied to detailed and nationally representative data

Benefit 1 Construct a comparable counterfactual group

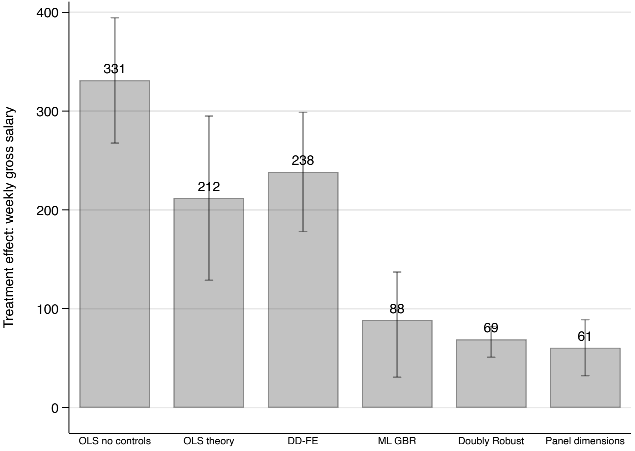
Benefit 2 A-theoretical approach to finding heterogeneous treatment effects

Why does ML Achieve those Benefits?

- Detect patterns in our (very) rich data e.g. important variables, functional forms and interactions
  - So what? Reduces mis-specification bias and can construct better counterfactuals
- Reduce information redundancy
  - So what? Reduces variance in estimators

Preview of Findings

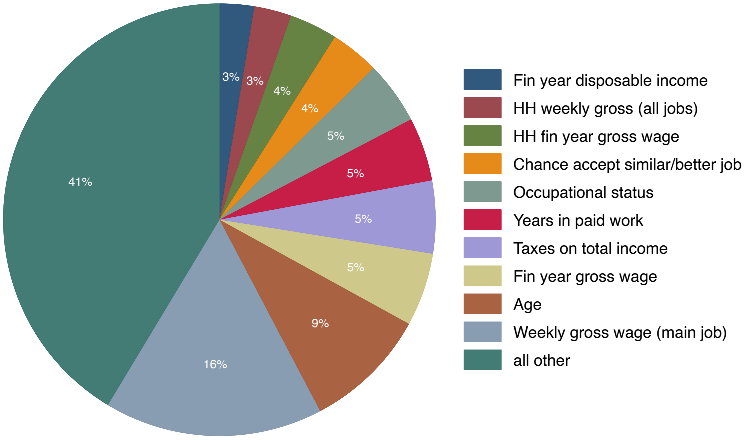
A ML-approach estimates a positive but smaller return to further education than traditional approaches



Notes: HILDA; sample size: 5,441 observations; 63,044 observations (Panel Dimension)

Preview of Findings

Effect sizes depend on starting income and age



DATA: HILDA

Several benefits to using HILDA data

- covers a long time-span (of nearly 2 decades, starting in 2001 and we use the wave 19 release)
- longitudinal data that details the year in which an individual started and completed further education
- wide range of background information on survey respondents

Sample

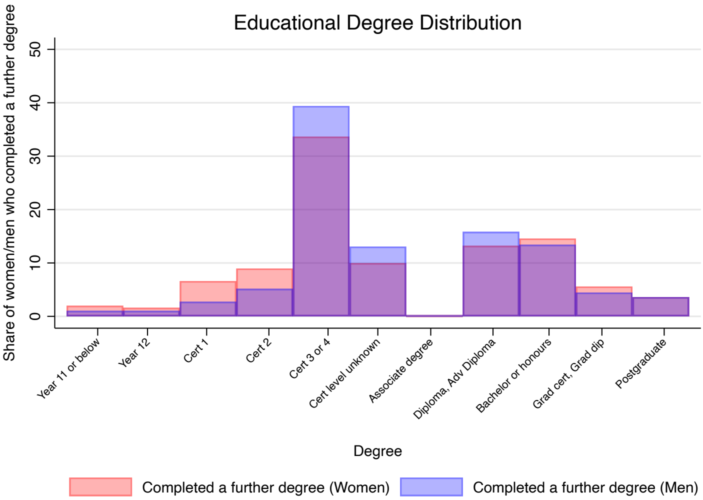
Our analysis sample includes:

- everyone who was 25 or above in 2001
- respondents not currently studying in 2001
- those observed in both 2001 and 2019 (and available information for Treatment and Outcome variables) and completed study by 2017
- Number of people in final sample: 5,441

Defining our Variables

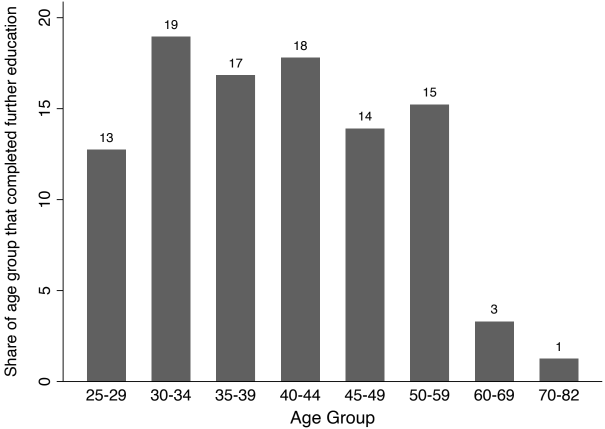
- Treatment indicator: obtaining an additional qualification between 2002 to 2017 (binary variable; 1,398 treated)
- Outcome variables: Earnings (total and hourly) in 2019
- Covariates or features: defined in 2001. Selected using a LASSO procedure.

Degree types by sex



Notes: HILDA; sample size: 5,441 observations

Degree completion by age (at 2001)



Notes: HILDA; sample size: 5,441 observations

Estimation goal and challenges

Aim: to estimate the causal returns to a new qualification

- The missing data problem

i	T	Y	Y(1)	Y(0)	Y(1) - Y(0)
1	0	0		0	?
2	1	1	1		?
3	1	0	0		?
4	0	0		0	?
5	0	1		1	?
6	1	1	1		?

Notes: Made-up table

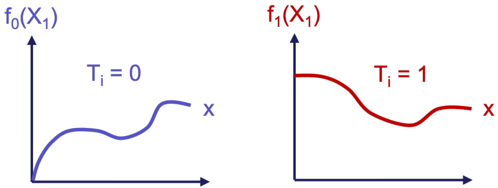
$$E[(Y_1)–(Y_0)] = E(Y_1)–E(Y_0) \neq E(Y_1|T = 1)–E(Y_0|T = 0) \quad (1)$$

How can ML help fill in the missing information?

- Standard ML predicts Y; alternatively, here we are interested in the causal parameter
- But no counterfactuals (no Individual Treatment Effects (ITEs))
- ML models can heavily penalise some variables that are weakly predictive of Y but strongly predictive of Treatment e.g. by shrinking their parameters
- Balance reducing this type of bias and increasing variance from including unnecessary variables in the model
- Adapt ML to fill in the missing data

T-learner

A potential solution



- Filling in the missing data blanks
  - two models: use 'treated' group to predict Y(1) and use 'control' group to predict Y(0)
  - for each person, we can plug their X's through the above two models to predict their outcomes when treated and when not treated
  - subtract difference of these two models to obtain estimated Individual Treatment Effects (ITEs)

Model selection

- Model classes: LASSO, Ridge, Gradient Boosting Regression (GBR)
  - LASSO and RIDGE are closely related to least squares but penalise complexity
  - GBR is a tree-based model: well-placed to pick up non-linearities
  - Tuning parameters are determined through cross-validation
  - GBR details gbr2

Model Selection

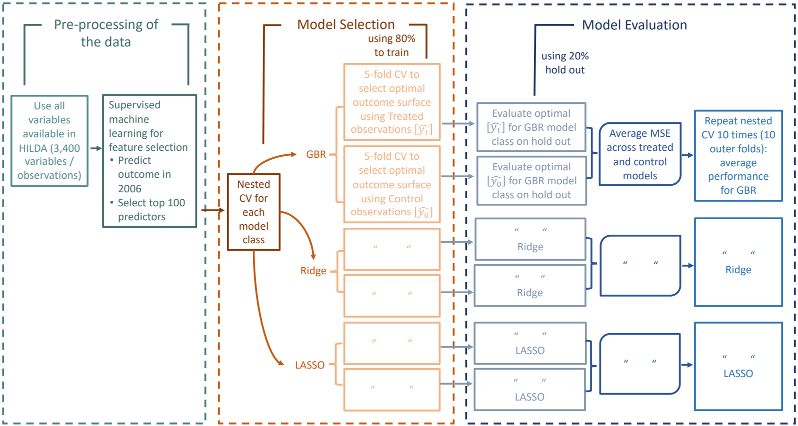


Table: Predictive Performance - Holdout Sample

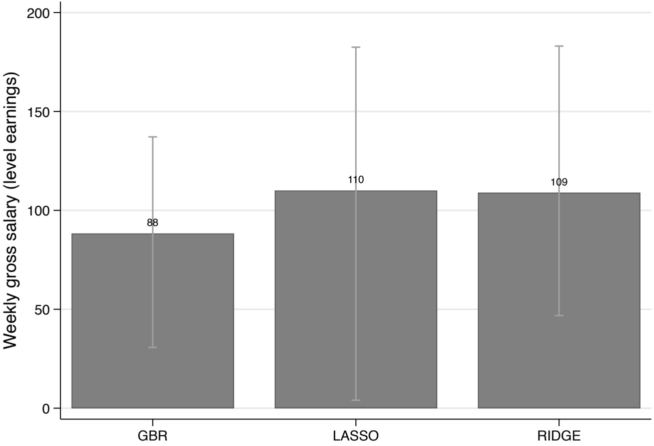
Model	Outcome surface	Out-of-Sample R-squared	R-squared SD
GBR	Treated	0.22	0.06
	Control	0.36	0.07
LASSO	Treated	0.15	0.09
	Control	0.32	0.05
Ridge	Treated	0.16	0.08
	Control	0.32	0.04

Notes: 5 fold CV performed on 80% train sample. All statistics presented in this table are based on the 20% holdout sample. Ten outer folds are used.

Inference

- Bootstrapping procedure
  - capture two sources of uncertainty: selecting the model and estimating the parameters
  - wider confidence intervals
- BS details explainer2

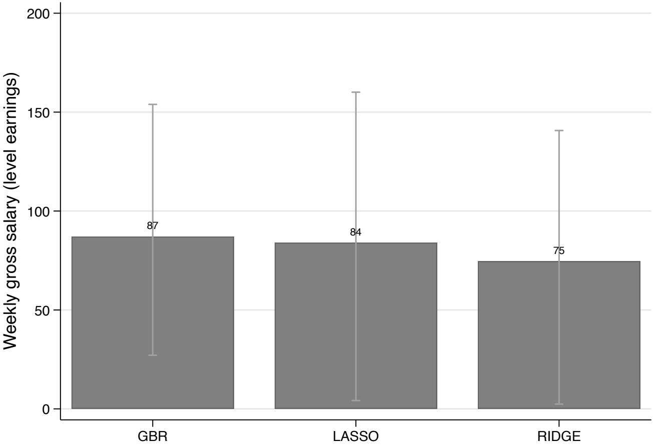
Results: Returns to Further Education



Notes: HILDA; sample size: 5,441 observations

GBR better Captures Non-linearities

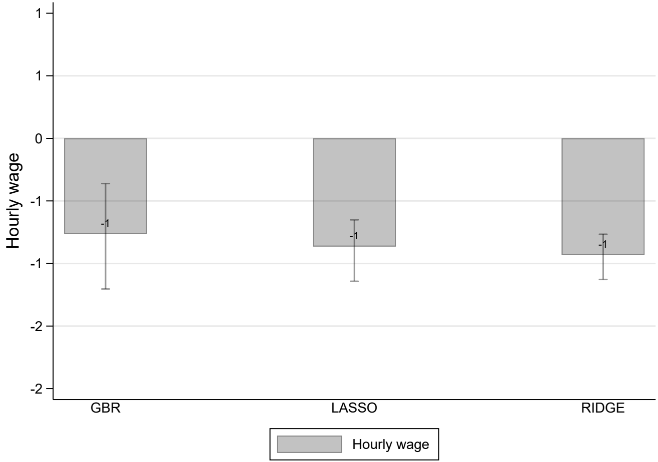
25 - 46-year-old (in 2001) sample



Notes: HILDA; sample size: 3,065 observations

Results: Wages

25 - 46-year-old (in 2001) sample



Improving on the T-learner

Doubly Robust

- Doubly Robust gives us two chances to get it right
- Previously, T-learner estimated the outcome of interest, given the treatment and the observable characteristics
- Here, we also estimate the probability of being treated, given the observable characteristics
- Combining above: robust to mis-specification of either model

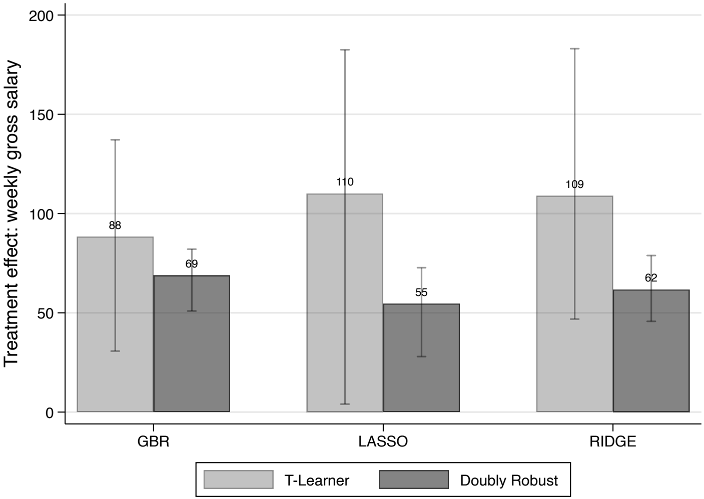
Doubly Robust

$$\hat{ATE} = \frac{1}{n}\sum_{i=1}^{n}[\frac{T_i(Y_i - \hat{\mu_1}(X_i))}{\hat{p}(X_i)} + \hat{\mu_1}(X_i)] - \frac{1}{n}\sum_{i=1}^{n}[\frac{(1-T_i)(Y_i - \hat{\mu_0}(X_i))}{1-\hat{p}(X_i)} + \hat{\mu_0}(X_i)]$$

- where:
  - $\hat{p}(X_i)$ is an estimation of the propensity score (using logistic regression)
  - $\hat{\mu_1}(X_i)$ is an estimation of E[Y|X,T=1] (using any ML model)
  - $\hat{\mu_0}(X_i)$ is an estimation of E[Y|X,T=0] (using any ML model)

Results

Doubly Robust

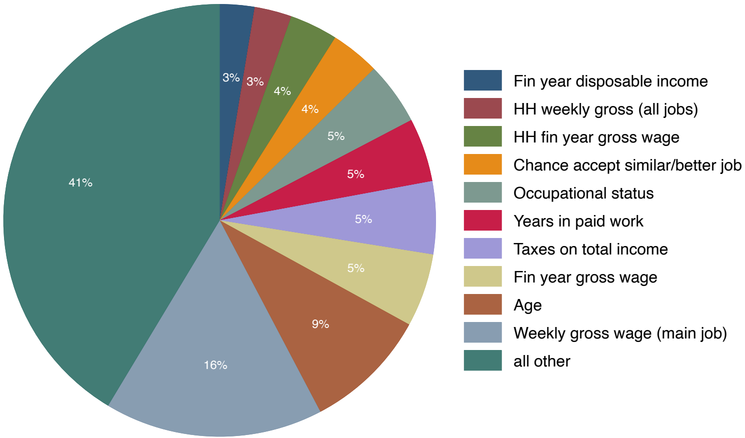


Notes: HILDA; sample size: 5,441 observations

- Is there a positive, long-run return to acquiring further education in later-life? ✓
- which groups were most affected?

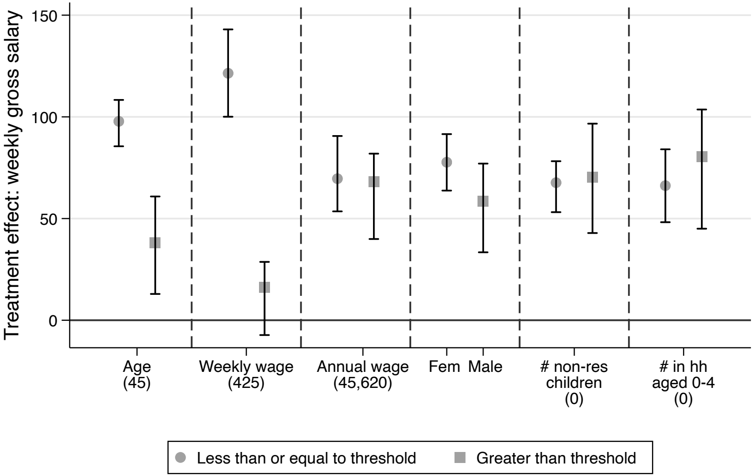
Sub-group analysis

Effect sizes depend on starting income and age



Sub-group analysis

Effect sizes depend on starting income and age



Note: Thresholds in parantheses

Summary of Other Results

- Acquiring an additional qualification may increase earnings through a number of potential mechanisms such as through getting a job, and switching occupations or industries
- Largest earnings gains are associated with acquiring an undergraduate degree or above and for technical subject areas
- Little evidence of well-being or mental health benefits

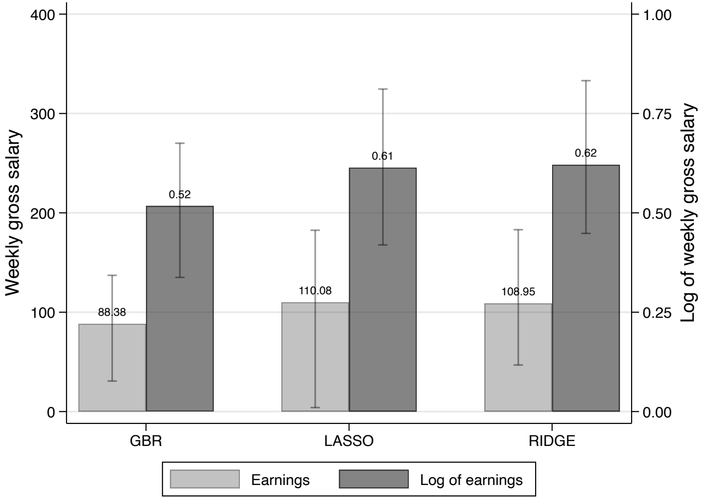
Conclusions

- Economic benefits to gaining an additional qualification in later-life
- An estimated gain of approximately $80 per week in gross earnings, which represents roughly 7 percent of the Average Weekly Gross Earning for the average worker in Australia
- Largest earnings gains are associated with younger learners and those with lower starting incomes

Appendix

Results: Returns to Further Education

Levels and Logs



Notes: HILDA; sample size: 5,441 observations

Sensitivity Analysis

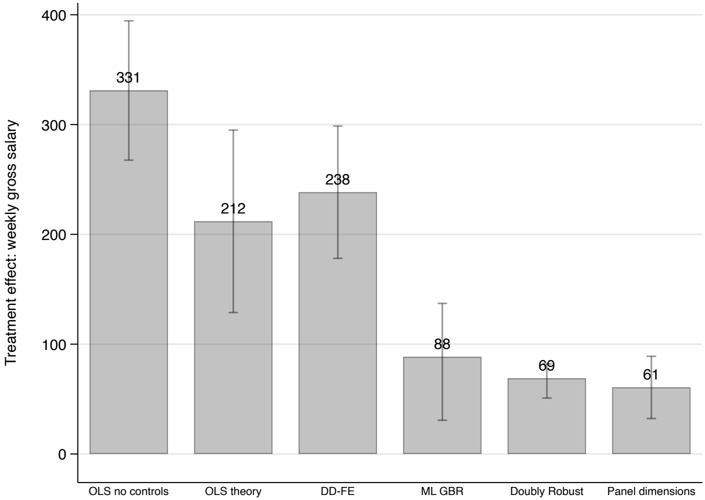
Defining features differently

We run sensitivity analysis on our approach

- Defining features differently: in year/s before study start
- Dealing with dynamic selection into further study
- Comparing the earnings of those who complete further study to the earnings of similar non-students who displayed the same paths (in earnings and other factors) as the student group before study began
  - Sample: 25 and above
  - Started studying in 2003 - 2017
  - In HILDA in two consecutive waves before study start
  - Could be in top-up sample or new joining household member (different from main analysis)
  - Started but did not complete are in the Control group
  - Control group are repeated: given a theoretical 'study-start' time stamp at every wave; standard errors adjusted

Sensitivity Analysis

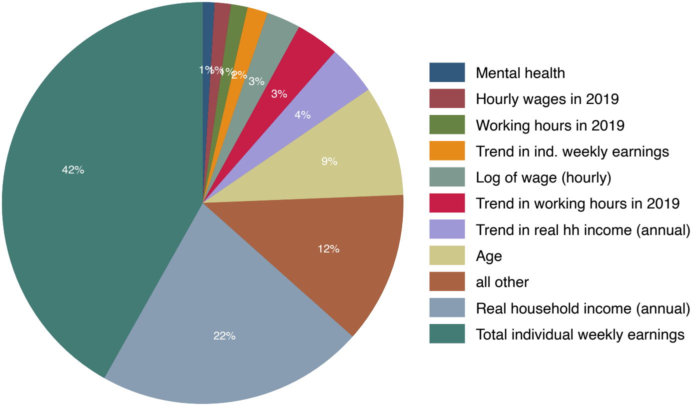
Across all Methods



Notes: HILDA; sample size: 5,441 observations; 63,044 observations (Panel Dimension)

Sensitivity Analysis

Important Trend Predictors

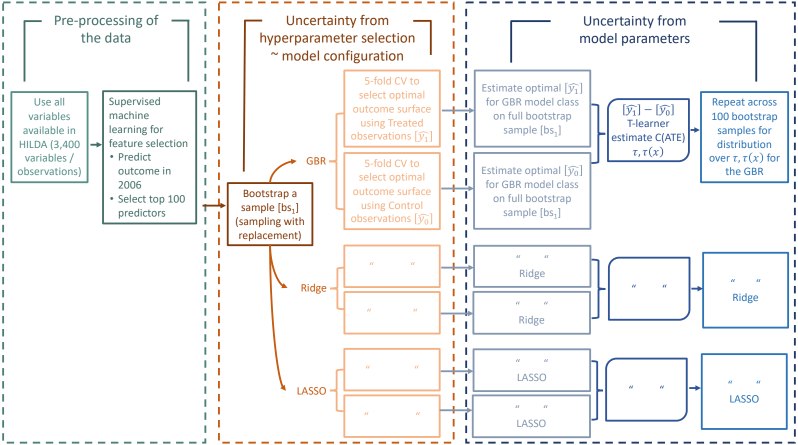


Notes: HILDA; sample size: 5,441 observations; 63,044 observations (Panel Dimension)

GBR details

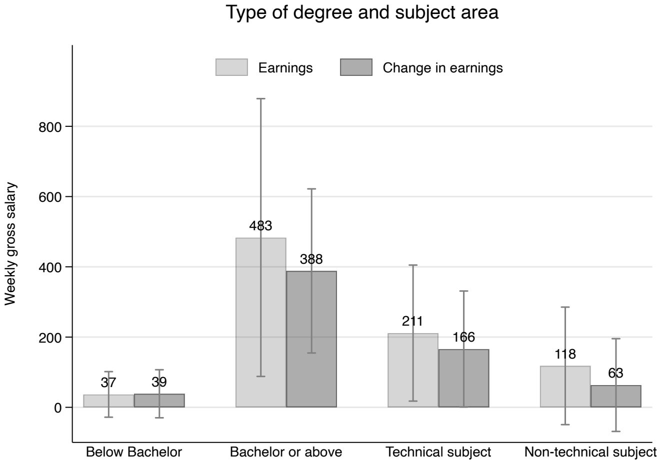
- In GBR, we sequentially fit small trees (i.e. you use the residuals from the previous tree to grow your next tree).
- Then we add a shrunken version of the new tree to the existing function and then update residuals and repeat
  - Number of trees (iterations)
  - Shrinkage – pace of learning
  - Number of splits (depth of tree) Back

Model Inference



◂ Back

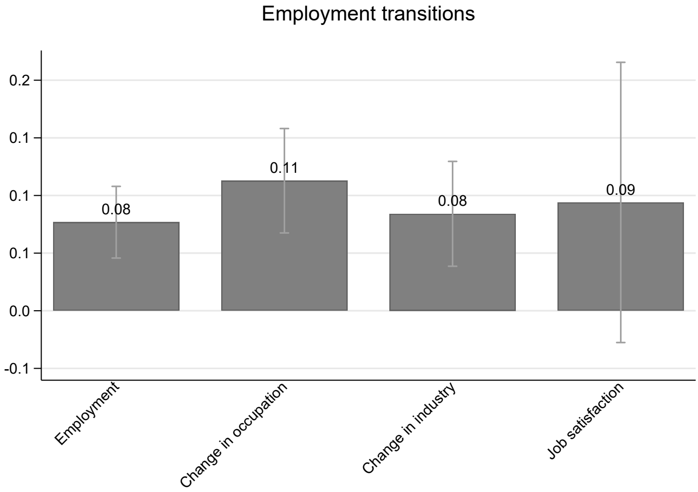
Higher-order degrees and technical subjects yield stronger effects



Notes: BB: 1,037; BA: 306; T: 681; NT: 662 *Technical*: STEM, Architecture, Ag and Env, Medicine, other health related studies and nursing, management and commerce and law. *Non — technical*: Education, Society and culture (includes economics!), Creative arts, and Food, hospitality and personal services

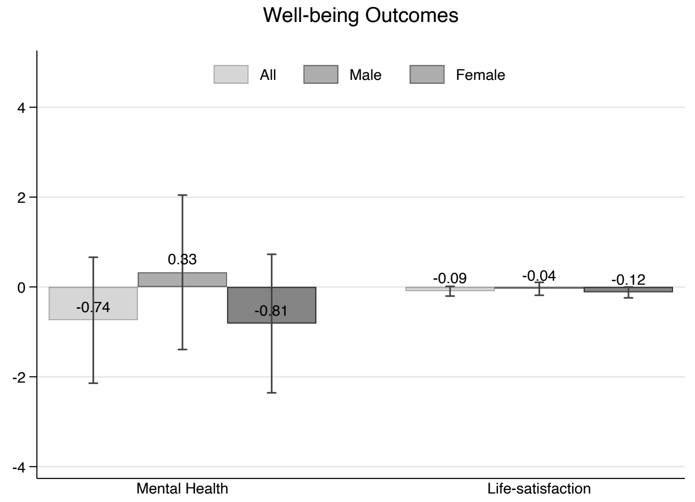
- Is there a positive, long-run return to acquiring further education in later-life? ✓
- Does the degree type and subject area matter? ✓
- Who was affected?
- Did it change long run labour market attachment?
- Any effects on mental health and wellbeing?

A larger workforce and a more mobile one



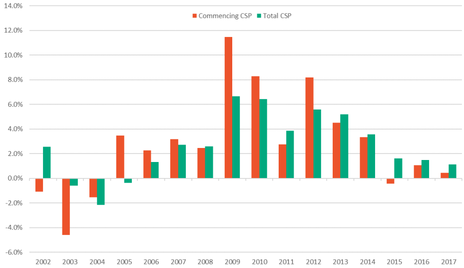
Notes: HILDA; ML model: LASSO

Small to no effects on wellbeing



Notes: HILDA; sample size: 5,441 observations; ML model: LASSO

Govt funding



Notes: DET, uCube