Effect Heterogeneity: JSP in Swiss [1]

Jasmine. Hao[1]

[1]University of Hong Kong

ECON 6083: Machine Learning

[1]This section is based on [Knaus et al., 2022].

Knaus et al. (2022)

Knaus, M. C., Lechner, M., & Strittmatter, A. (2022). Heterogeneous employment effects of job search programs: A machine learning approach. *Journal of Human Resources*, 57(2), 597-636.

- **Research Question**
  What are the **heterogeneous** employment effects of job search programmes (JSPs) for unemployed workers?

Findings

Main results of this paper are:

• **Substantial effect heterogeneity** of Swiss JSPs during the **first six months** after the start of participation.

- **Negative effects** for most participants.
- Heterogeneity is **strongly** related to **unemployed characteristics**.
  - Disadvantaged labour market characteristics benefit more (lower lock-in effects and lower indirect programme costs).
  - Foreigners benefit more (less access to informal job search networks compared to locals).
- Only **little** heterogeneity for **caseworker characteristics**.

• **No substantial effect heterogeneity beyond six months** after the start of training.

Unemployed in Swiss

- Unemployed persons **have to register** at the regional employment agency closest to their home.
- Unemployed persons can participate in different active labour market policies, and JSP is the most common choice ($> 50\%$).

What are JSPs?

- JSPs provide training in effective job search and application strategies (e.g., training in resume writing).
- JSPs are relatively short (about 3 weeks).
- During JSPs, participants need to continue to search for jobs.

JSPs in Swiss

- Federal eligibility rules are rather vague.
- **Regional employment agencies** have a large degree of **autonomy**.
  - Caseworkers make the decision to assign unemployed persons to a training course based on information about the persons.
  - The omission of such variables generally contaminates estimation and inference results based on the selected set of variables.

Pros about JSPs

- Participants may improve the visibility of suitable job vacancies and the efficiency of the application process, which may improve employment stability.
- Duration of unemployment is decreased.

Cons about JSPs

- Participants may accept jobs with low matching quality, which may reduce employment stability.
- Intensive monitoring of JSPs may reduce informal job search, which might be a more efficient strategy than formal job search for some unemployed persons.

Empirical evidences suggest that Swiss JSPs have negative employment effects, which taper off one year after the start of participation (see Gerfin & Lechner (2002), Lalive et al.(2008)).

Data

Including all individuals who are registered as unemployed at a Swiss regional employment agency in 2003 aged between 24 and 55.

- From different unemployment insurance databases (AVAM/ASAL) and social security records (AHV).
- Including regional labour market characteristics, characteristics of unemployed persons, caseworker characteristics, and months employed since programme start.

In most cases, the **same caseworker** is responsible for the entire unemployment duration of his or her client.

- If this is not the case, the paper focuses on the first caseworker to avoid concerns about (rare) endogenous caseworker changes.

Excluding individuals who participate in other ALMPs within the first six months of unemployment.

- JSPs' participants vs. non-participants.

Descriptive Statistics

	Participants		Non-Participants		Std. Diff.
	Mean	S.D.	Mean	S.D.	
	(1)	(2)	(3)	(4)	(5)
Outcome: Months employed since programme start					
During first 6 months	1.21	1.93	1.94	2.44	23.29
During first 12 months	3.68	4.27	4.53	4.80	13.12
During first 31 months	15.30	12.49	15.59	12.85	1.60
During months 25 - 31	3.48	2.88	3.33	2.86	3.72
Characteristics of unemployed persons					
Female	0.45	-	0.44	-	0.58
Age (in 10 years)	3.73	0.88	3.66	0.86	5.59
Unskilled	0.22	-	0.23	-	1.80
Some qualification degree	0.60	-	0.56	-	5.19
Employability rating low	0.12	-	0.14	-	3.97
Employability rating medium	0.77	-	0.74	-	5.79
Employability rating high	0.11	-	0.12	-	3.62
# of unemp. spells in last 2 years	0.41	0.98	0.64	1.27	13.85
Fraction of months emp. in last 2 years	0.83	0.22	0.79	0.25	12.57
Past income (in 10,000 CHF)	4.58	2.02	4.16	2.05	14.50
Caseworker characteristics					
Female	0.45	-	0.41	-	6.94
Age (in years)	44.0	11.6	44.4	1.16	7.7
Tenure (in years)	5.54	3.23	5.86	3.31	6.84
Own unemp. experience	0.63	-	0.63	-	0.54
Vocational training degree	0.26	-	0.23	-	5.63
Local labour market characteristics					
German speaking REA	0.89	-	0.67	-	39.68
French speaking REA	0.08	-	0.25	-	33.30
Italian speaking REA	0.03	-	0.08	-	16.81
Cantonal unemployment rate (in %)	3.64	0.77	3.75	0.86	9.23
Cantonal GDP per capita (in 10,000 CHF)	5.13	0.92	4.92	0.93	15.75
# of caseworkers	989		1,282		

The Treatment Effects I

Using Rubin's (1974) potential outcome framework.

$$Y_i = Y_i^1 D_i + Y_i^0 (1 - D_i), \text{where}$$

- D_i: the binary treatment dummy, indicating JSP participation.
- Y_i^1: the potential outcome when individual i is participating in a JSP ($D_i = 1$).
- Y_i^0: the potential outcome when individual i is not participating in a JSP ($D_i = 0$).
- Y_i: the observed outcome of individual i.

Causal effect of D on Y for individual i is

$$\tau_i = Y_i^1 - Y_i^0.$$

The Treatment Effects II

How to identify group averages of τ_i (see Imbens & Wooldridge (2009))?

- Average treatment effect (ATE), $\tau = E[\tau_i]$.
- Average treatment effect on the treated (ATET), $\theta = E[\tau_i|D_i = 1]$.
- Average treatment effect on the non-treated (ATET), $\rho = E[\tau_i|D_i = 0]$.
- Conditional average treatment effects (CATE), $\tau(z) = E[\tau_i|Z_i = z] = E[Y_i^1 - Y_i^0|Z_i = z]$, where Z_i are exogenous pre-treatment variables.

The fundamental problem of causal analysis is that at least one potential outcome is unobservable.

- We cannot observe both Y_i^1 and Y_i^0.

The Treatment Effects III

Many ML methods consider effect heterogeneity in **randomized experiments**, which are expensive and minimally socially acceptable.

This paper considers a **selection-on-observables** identification strategy.

A promising approach to estimate group specific causal effects in nonexperimental contexts is the Modified Covariate Method (MCM).

MCM I

To introduce MCM, consider the stylised case where participation in a programme is randomly assigned to **50%** of the unemployed persons:

$$Y_i = Z_i\beta_s + D_i Z_i \delta + u_i, \text{ where} \tag{1}$$

- Z_i: constant term ($Z_{i0} = 1$) and pre-treatment variables (potentially related to the effect heterogeneity).
- $Z_i\beta_s$: linear approximation of the conditional expectation of the potential outcome under non-participation, $E[Y_i^0|Z_i = z] = z\beta_s$. Call this the main effects in the following.
- $D_i Z_i \delta$: provides linear approximation of the CATE. $\tau(z) = z\delta = E[Y_i^1 - Y_i^0|Z_i = z]$.

MCM II

Because of parsimony and robustness to misspecification of the main effects, we introduce MCM (see Tian et al. (2014)):

Transform the treatment dummy in model (1) into $T_i = 2D_i - 1$:

$$Y_i = Z_i\beta_t + \frac{T_i Z_i}{2}\delta + v_i. \tag{2}$$

Model (1) vs. model (2)

- Treatment indicator shifts from $D_i \in \{0, 1\}$ to $\frac{T_i}{2} \in \{-0.5, 0.5\}$.
- Coefficient vector δ does not change.
- Main effects changes, so:
  - model (1): $Cov(Z_{ij}, D_i Z_{ik}) = Cov(Z_{ij}, Z_{ik})E[D_i] = \frac{Cov(Z_{ij}, Z_{ik})}{2}$.
  - model (2): $Cov(Z_{ij}, T_i Z_{ik}) = Cov(Z_{ij}, Z_{ik})E[T_i] = 0$. So, we can estimate the coefficients β_t and δ in two separate regressions.

MCM III

We can estimate CATEs with the model

$$Y_i = \frac{T_i Z_i}{2}\delta + \epsilon_i.$$

- It is the baseline model of the MCM.
- It is suitable when **only the interaction effects** and **not the main effects** are of interest.

Identification

In addition to the pre-treatment variables included in the vector Z_i, we consider the possibility of confounding variables X_i.

- Z_i may be larger, smaller, partially, or fully overlapping with X_i.

Assumptions:

1. **Conditional independence**: $Y_i^1, Y_i^0 \perp D_i | X_i = x, Z_i = z$ for all values of x and z in the support of X and Z.
2. **Common support**: $0 < P(D_i = 1 | X_i = x, Z_i = z) < 1$ for all values of x and z in the support (of interest) of X and Z.
3. **Exogeneity of controls**: $X_i^1 = X_i^0$ and $Z_i^1 = Z_i^0$.
4. **Stable Unit Treatment Value Assumption (SUTVA)**: $Y_i = Y_i^1 D_i + Y_i^0 (1 - D_i)$.

Assumptions

Conditional independence requires potential outcomes are independent of programme participation conditional on the confounding pre-treatment variables.

- Including characteristics of the unemployed and the caseworkers.

Common support requires conditional probability to participate in a JSP is bounded away from zero and one.

- Trim observations below the 0.5 quantile of participants and above the 99.5 quantile of non-participants.

Exogeneity of controls requires exogeneity of confounding and heterogeneity variables.

- Only use control variables that are determined prior to the start of JSP participation.

SUTVA excludes spillover effects between participants and nonparticipants.

Theorem

Theorem: Under Assumptions 1-4, the following equality holds:

$$\begin{aligned}\tau(z) = E_{X|Z=z}[E(Y_i|D_i = 1, X_i = x, Z_i = z)|Z_i = z] \\ - E_{X|Z=z}[E(Y_i|D_i = 0, X_i = x, Z_i = z)|Z_i = z]\end{aligned}$$

Thus $\tau(z)$ are identified from observable data on $\{Y_i, D_i, Z_i, X_i\}_{i=1}^{N}$.

Search for Effect Heterogeneity I

Combine MCM with Inverse Probability Weighting (IPW), which is a standard approach to balance covariates in observational studies.

Estimate δ using Weighted Ordinary Least Squares (WOLS) by

$$\arg\min_{\hat{\delta}}[\sum_{i=1}^{N} \hat{w}(D_i, X_i, Z_i)\, T_i(Y_i - \frac{T_i Z_i}{2}\hat{\delta})^2] \text{ , where} \tag{3}$$

$\hat{w}(D_i, X_i, Z_i)$ is the IPW weights.

Z_i are potentially related to effect heterogeneity, consist of:

- Individual and caseworker characteristics, their second order interactions, up to fourth order polynomials, and logarithms of non-binary variables.
- Dummies for the 103 employment agencies, 29 category dummies for previous industry and 29 category dummies describing the previous job.
- 1,268 heterogeneity variables in total.

Search for Effect Heterogeneity II

85,198 unemployed persons with 1,268 heterogeneity variables, so this paper employs LASSO estimators:

$$\arg\min_{\hat{\delta}}[\sum_{i=1}^{N}\hat{w}(D_i,X_i,Z_i)T_i(Y_i-\frac{T_iZ_i}{2}\hat{\delta})^2]+\lambda\sum_{j=1}^{p}|\hat{\delta}_j|, \tag{4}$$

where we add a penalty term for the sum of the absolute values of the coefficients of the p variables appearing in Z.

- Do not penalize the constant $\hat{\delta}_0$.
- If $\lambda = 0$, then equation (4) is equivalent to equation (3).
- If $\lambda > 0$, then some coefficients are shrunken towards zero.
- **Bias-variance trade-off**: too **low penalties** lead to **overfitting**; too **high penalties** lead to models that **miss important variables**.

Search for Effect Heterogeneity III

This paper

- applies 10-fold cross-validation to find the penalty term λ with the best out-of-sample performance in terms of MSE.
- uses **Post-LASSO** estimates to calculate the MSE, and chooses λ that minimises the Post-LASSO MSE.
- avoids idiosyncratic within-sample effects by splitting the sample into **two equal sized parts**.
  - 1^{st} sample: training sample, to select the relevant effect heterogeneity variables
  - 2^{nd} sample: estimation sample, to estimate the WOLS model including all selected heterogeneity variables.
  - To improve precision and check the sensitivity of our results, this paper tries different splits and gets essentially similar CATEs.

Selected Variables

	Months employed during first 6 months after the start of participation		Months employed during first 12 months after the start of participation	
	Coef.	S.E.	Coef.	S.E.
	(1)		(2)	
Constant	-0.89***	(0.05)	-1.29***	(0.09)
# of unemp. spells in last two years	0.06	(0.12)	-	-
Unskilled × past income 0 - 25k	0.30***	(0.11)	0.53	(0.53)
Skilled w/o degree × same gender like CW	0.20	(0.21)	-	-
Skilled w/o degree × age difference between unemployed & CW	-0.01	(0.01)	-	-
# of unemp. spells in last 2 years × age of CW	0.00	(0.00)	-	-
# of unemp. spells in last 2 years × medium city size	-0.05	(0.06)	-0.13	(0.14)
# of unemp. spells in last 2 years × past income 0 - 25k	-0.04	(0.06)	-0.10	(0.14)
# of unemp. spells in last 2 years × prev. job unskilled	0.04	(0.05)	0.21*	(0.13)
# of unemp. spells in last 2 years × same gender like CW	-0.01	(0.05)	-	-
CW has own unemp. experience × prev. job unskilled	0.19**	(0.09)	0.34*	(0.21)
Foreigner with perm. residence permit × past income 25 - 50k	0.19	(0.12)	-	-
Small city × past income 50 - 75k	-0.16*	(0.09)	-0.26	(0.20)
Single household × no emp. spell last 2 years	-0.17**	(0.08)	-	-
Single household × prev. job unskilled	0.16	(0.11)	-	-
Prev. job primary sector × age difference between unemp. person & CW	-0.02**	(0.01)	-	-
Prev. job restaurant	-0.01	(0.12)	-	-
Prev. job tourist sector	-0.09	(0.12)	-	-
Unskilled × prev. job unskilled	-	-	-0.22	(0.64)
# of unemp. spells in last 2 years × unempl. & CW have primary education	-	-	0.19**	(0.08)
CW has vocational training degree × past income 50 - 75k	-	-	-0.13	(0.30)
Past income 25 - 50k × unskilled	-	-	0.14	(0.24)
# of emp. spells past 5 years × prev. job in primary sector	-	-	-0.24	(2.16)
Prev. job in primary sector × unskilled	-	-	-0.19	(0.53)
Regional emp. agency No. 44	-	-	-0.68	(0.52)
# of selected variables	17 of 1,268		13 of 1,268	

Interpretation

When using “Months employed during first 6 months after the start of participation" as outcome variable, Post-LASSO estimation selects 17 out of 1,268 potential variables, 5 of which are significant.

- Treatment effect increments by 0.3 months for unskilled workers with previous earnings below 25,000 CHF/yr.
- If all other selected variables equal 0, the predicted effect of JSP participation for unskilled workers with previous earnings below 25,000 CHF/yr is -0.89 + 0.3 = -0.59 months employment.
- Be cautious when interpreting the coefficients:
  - they are selected to maximise prediction power and might differ from the structural (causal) model.

Average Effects vs CATE

Figure: *

Average effects since participate in JSP

Months employed since start of participation	ATE		ATET		ATENT	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
	(1)		(2)		(3)	
During first 6 months	-0.80***	(0.02)	-0.82***	(0.02)	-0.80***	(0.02)
During first 12 months	-1.10***	(0.05)	-1.13***	(0.04)	-1.09***	(0.05)
During first 31 months	-1.14***	(0.14)	-1.20***	(0.13)	-1.12***	(0.15)
During months 25-31	-0.007	(0.03)	-0.011	(0.03)	-0.007	(0.04)

Figure: *

Aggregated CATEs since participate in JSP

Months employed since start of participation	Mean	Median	S.D.	Min.	Max.	Mean S.E.
	(1)	(2)	(3)	(4)	(5)	(6)
During first 6 months	-0.78	-0.84	0.25	-1.41	0.77	0.07
During first 12 months	-1.10	-1.20	0.32	-2.09	1.44	0.10
During first 31 months	-1.13	-1.25	0.60	-3.79	4.12	0.23
During months 25-31	-0.04	-0.05	0.06	-0.32	0.48	0.04

Aggregated CATEs are close to the estimated ATEs, meaning that the estimation of the **aggregated CATEs works well** on average.

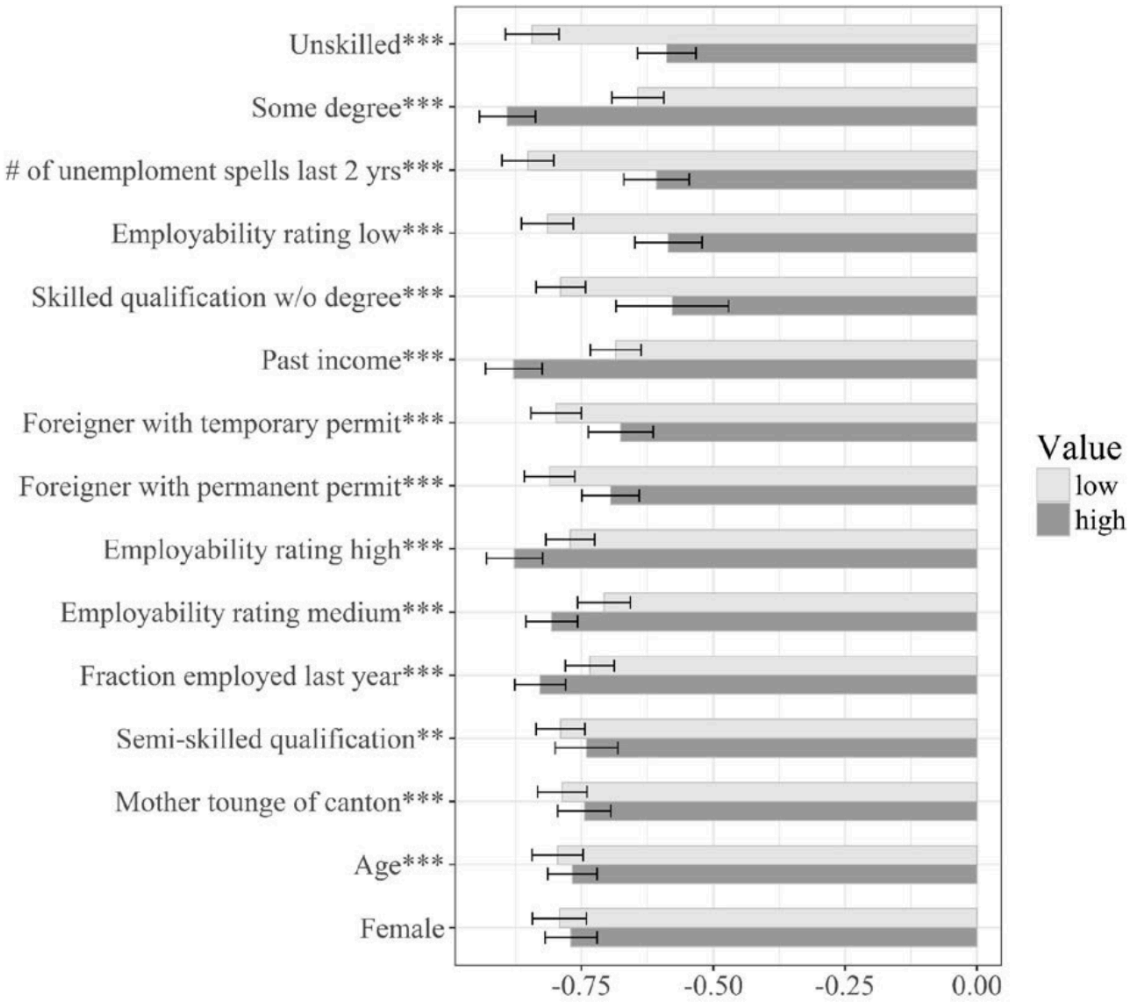
Effect Heterogeneity by Selected Variables

This paper averages CATEs by characteristics of unemployed persons and their caseworkers.

- For each characteristic, the sample is partitioned into **two mutually exclusive groups** (high $g = 1$, low $g = 0$), by using a binary characteristic itself or by discretising at the median of non-binary characteristics.
- The parameters $\bar{\tau}_{g=1}$ and $\bar{\tau}_{g=0}$ average the CATEs over all unemployed in the respective group.

Characteristics of Unemployed Persons I

CATEs on cumulated employment during the first 6 months after start JSP participation by characteristics of unemployed persons.



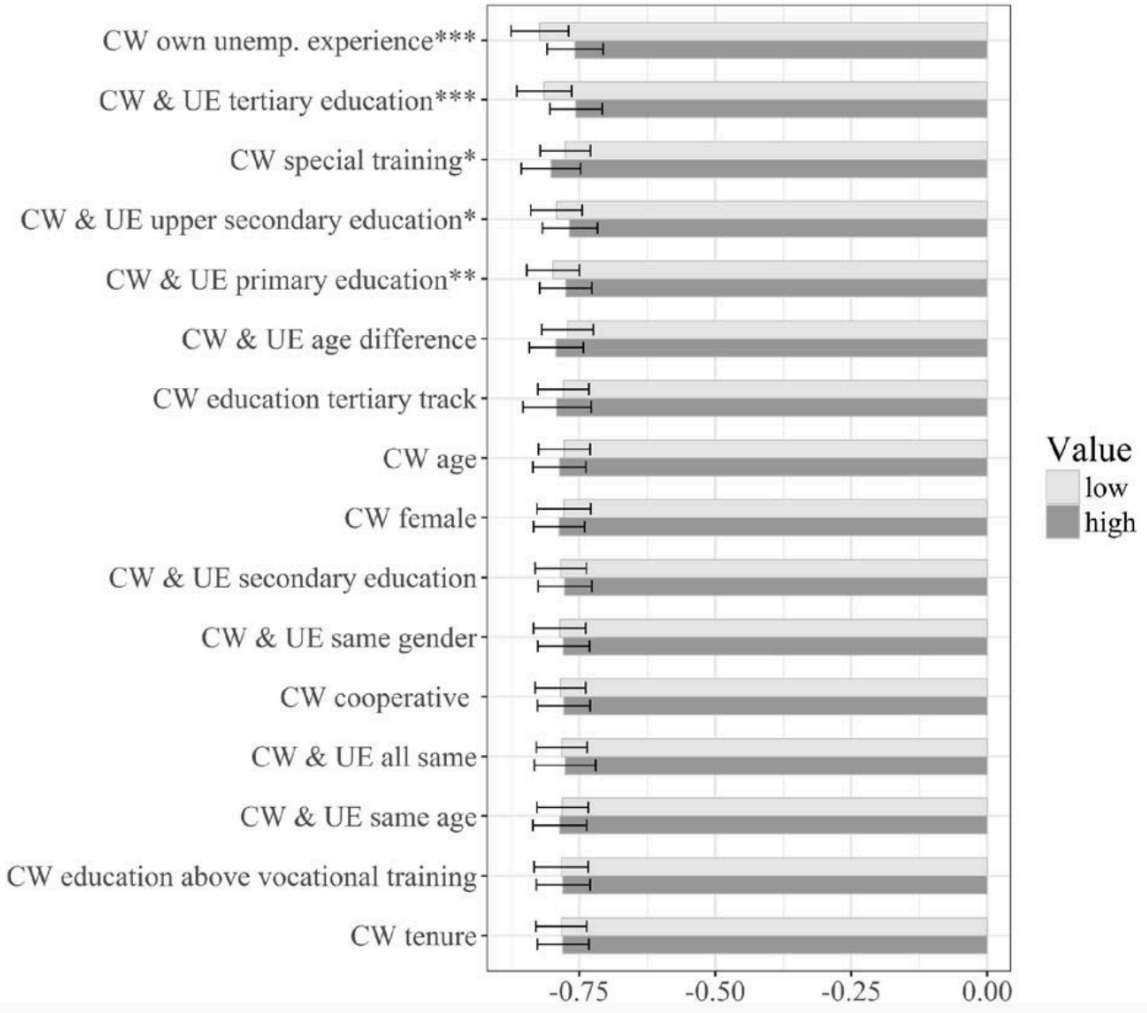
Characteristics of Unemployed Persons II

The top of the figure shows the larger effect heterogeneities, while the bottom shows little effect heterogeneity.

- Average effect of JSP for the **unskilled unemployed is longer** than for unemployed persons in other skill categories.
- Individuals with **some degree of education suffer more** from JSP on average than individuals with no degree.
- **Negative lock-in effect is much less pronounced** for unemployed persons with **lesser qualifications**.
- **Lock-in effects are less negative for foreigners**.
  - Foreigners have smaller network for an informal job search, so the formal job search strategy might be relatively successful for them.
- Little heterogeneity by gender and age.

Characteristics of Caseworkers I

CATEs on cumulated employment during the first 6 months after start JSP participation by characteristics of caseworkers.



Characteristics of Caseworkers II

Although there are some statistically significant differences, **characteristics of caseworkers are much less pronounced** than for the characteristics of unemployed persons.

- Most effect heterogeneity is observed by **caseworkers' own unemployment experience**, but the difference is only 0.07 months.
- The cooperativeness of caseworkers has no statistically significant influence on the effectiveness of JSP participation.
  - The authors would have expected this characteristic to be a good predictor for effect heterogeneity, because it might approximate different monitoring intensities of the caseworker.

References


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