SPES’s Job Seeker Profiling
ETF Conference, Solna, 2024-05-22
Petter Helgesson, Analyst at SPES
Scope of this presentation

- SPES’s current usage of statistical profiling
- Available data
- The underlying statistical model
- A little on measuring model performance
SPES’s current usage of profiling

- Profiling central part of assessment support tool for recommending *Prepare and Match*
  - Since 2019 with major upgrade 2023 (when Prepare and Match 2 was introduced)

- Prepare and Match:
  - Support program delivered by private providers
  - Private providers receive reimbursement in two parts: baseline + result based
  - Directed towards mid-segment of job seekers

- The profiling tool estimates job finding probability to help in finding this mid-segment
  - + determines reimbursement to providers (A, B, C)

- Also: increasing usage as a follow-up-tool: e.g., how much support is given to those who need it the most
In practice

The caseworker receives a recommendation

1. Prepare and match yes/no and most important factors behind this recommendation
2. Reimbursement level

The caseworker always makes the final decision (except for the reimbursement level)

*Note to interpreters: the screenshot is in Swedish and serves as an illustration, the content is not necessary to translate*
Available data

- Based primarily on registration data
  - + population registration

- Collected in a relatively similar fashion for tens of years
  - Self-registration has been introduced and dominates more and more – normally checked by caseworker at first meeting

- Some more substantial changes in definitions (primarily classification of occupations) 2010-2014
  - We use data from 2015 and on

- Explanatory variables
  - Length of unemployment spell
  - Job searcher category
  - ALMPs (in model training)
  - Country of origin
  - Occupations
    - Searching for
    - Experience in
    - Relevant education
  - Disability (leading to reduced working capacity)
  - Unemployment fund
  - Education
    - Level
    - Orientation
  - Municipality
  - Sex
  - Age

- Outcomes
  - Timing and reason for deregistration
How the model is constructed
A simple option and it’s drawback

- Covariates at one point in time
- Outcomes (0/1) at another point in time
- Train model which estimates probability of a positive outcome at the second point in time
  - Arbitrary classification model can be used
- Drawback: If individuals take part of ALMPs between the start date and the outcome date, the individuals’ inherent abilities are confounded with the help/locking-in they receive
  - Important for us!
- Answers “what is the job finding probability including any ALMP effects?”

What’s the outcome?

Time since start

e.g., 6 months

Covariates

0
The selected model

- Answers another question: “What is the job finding probability excluding any ALMP effects?”
- How?
  - In principle: divide time into intervals, use the covariates at the start of each interval
  - Specifically: use a survival model (“piece-wise exponential”)
A few words on model performance
General comments

- Necessary to compare historical outcomes to potential predictions
  - Impossible to *observe* realized job chance – either you got a job or not: outcomes must be grouped

- A wealth of metrics – which to choose?
  - Often published for binary classifiers:
    - Accuracy: share of “correct” predictions - easy to understand but very sensitive to skewness of the problem at hand (including threshold)
    - ROC-AUC: how well sorted are the predictions?
  - Is the model well calibrated? I.e., is the share of positive outcomes in groups close to the mean prediction?
    - Meaningful results also on subgroup level (sex, country of origin, …) – can be used to ensure non-discrimination (with a certain definition)
    - Not straightforward to summarize into one single number
  - Concordance/c-statistic (survival models) – related to ROC-AUC but takes “continuous” time to an outcome into account
    - ...

- What are good values?
  - Comparisons to published metrics are somewhat problematic: different problems are different in difficulty
    - But perhaps the best practically feasible solution
  - Good option if feasible: set up fair comparison to caseworker predictions
    - With the same data, the model will perform better, but the caseworker may observe additional information

- Who should understand the results?
  - How can they be communicated to this audience?
Appendix
(to help concretize the performance metrics)
Accuracy (as a binary classifier)

- Binary classification: close to/far from labor market
  - Coarse classification of job finding probabilities above/below threshold
- Compared to outcomes
- Accuracy - share of "correct" predictions in our case:
  76 % ("qualified chance": 59 %)
Ranking performance measure: ROC-AUC

- Compares all pairs where one individual has got a positive outcome and the other has not: what fraction of these pairs are correctly ranked by the model?
  - Can be corrected for effect of ALMPs
- In our case: 81.5%
Well calibrated?
Well calibrated groupwise? Non-discrimination
## Comparison to other models

<table>
<thead>
<tr>
<th>LAND</th>
<th>Accuracy [%]</th>
<th>ROC-AUC [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>76 (conservative)</td>
<td>81,5 (adjusted)</td>
</tr>
<tr>
<td>PES 2020-2023</td>
<td>68 (PES 2023*: 74,9)</td>
<td></td>
</tr>
<tr>
<td>IFAU (2007)</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>-</td>
<td>80</td>
</tr>
<tr>
<td>Germany</td>
<td>84 - 85 (AF 2023*: 85,4)</td>
<td>70 – 77 (PES 2023*: 79,9-80,9)</td>
</tr>
<tr>
<td>Ireland</td>
<td>69 - 86</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td>-</td>
<td>63 – 83</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>67</td>
<td>76</td>
</tr>
<tr>
<td>Austria</td>
<td>80 - 85</td>
<td></td>
</tr>
</tbody>
</table>