The Importance of Language Training

Mette Foged
University of Copenhagen, Rockwool Foundation

Edith Zink
University of Copenhagen

This document summarizes the regression equations, non-response weighting and heterogeneity analysis for an information treatment about the perceived importance of Danish language proficiency for earnings in the Danish labor market. This is done following the recommendations in Haaland et al. (2023). Half of the sample will be treated with information on the returns to Danish skills.

Our survey will be merged with the Danish population register data. For information that is retrieved from this data source, we specify the exact registers in parentheses. Documentation on these registers can be found online on Statistics Denmark's website:

https://www.dst.dk/extranet/forskningvariabellister/Oversigt%20over%20registre.html, last accessed January 7, 2025.

1   Non-response weighting

Register data gives us access to a wide range of individual characteristics for both respondents and non-respondents. Our survey targeted the population of Ukrainian refugees in Denmark. We would like to ensure that our results are representative of that population. Therefore, we will conduct balance tests using characteristics at arrival:

- Age at immigration (VNDS and BEF)[1]

- Gender in the first entry in the population register (BEF).

- Indicator equal to one for individuals with an immigration spell in Denmark before Feb. 24, 2022, i.e. date of the Russian invasion into Ukraine (VNDS).

- Indicator equal to one for individuals with admission to Denmark according to the Special Law for People Displaced from Ukraine versus other admission criteria as their first admission category related to their immigration on or after Feb. 24, 2022 (OPHG)

---

[1]Entry is defined as the date of registering as immigrated to Denmark, `HAEND_DATO` from the VNDS registry. Birthdate comes from the first data entry in BEF.

- Children registered in the first entry in the population register (BEF).[2]

- Indicator for partner in Denmark (co-habitation or marriage) based on the first entry in the population register (BEF).

- Month of immigration fixed effects (VNDS).

In the case of imbalances, we will estimate entropy weights using the characteristics at arrival (see list above) following Hainmueller (2012) and apply them in our estimations. We will report both weighted and unweighted treatment effects.

2 Balance across treatment arms

We will conduct balance test to ensure that there are no statistically significant differences between treatment and control group before the treatment group was treated.

The list of variables for these balance tests is:

1. From the register data:

   - All characteristics at arrival listed above

   - Age of children if present (BEF).

   - Enrollment, accumulated hours, highest completed CEFR level in language training in the month before treatment (DKDB)

   - Employment indicator (at least one hour of paid work), hours worked, earnings, industry in the month before treatment (BFL)

   - Outmigration (VNDS)

2. From our survey data:

   - Prior beliefs

   - Indicator for self-reported symptoms of PTSD/CPTSD, ITQ and DSO scores

   - Difficulties to concentrate

   - Level of education before migration

   - Return-intentions

---

[2]We have access to quarterly draw of the population register BEF. So, this information will be based on the first draw of a quarterly BEF the individual appears in. Depending on the exact dates of immigration this can be at slightly different times since immigration for refugees.

- Proficiency in Danish

- Proficiency in English

- Family remaining in Ukraine (partner, children, parents)

- Family also in Denmark (partner, children, parents)

- Children's enrollment in daycare/school.

This list covers all variables that we will use in the heterogeneity analysis of our treatment effects.

3   Outcomes

1. Survey outcome

   - **Main outcome:** Intended hours to invest into learning Danish.

     Based on your current situation, how many hours per week would you like to invest in learning Danish, where 0 means no time and 37 hours represent the equivalent of a full-time job?
     – slider vom 0 to 37 –

2. Register outcomes measured on monthly level followed up until one year after experiment (checked balance before treatment)

   - From migration register (VNDS)

     - Outmigration

   - From language training registry (DKDB)

     - Indicator for enrollment

     - Hours of language training attended (assigned hours - hours absent)

     - Highest completed CEFR level (track by module mapped into CEFR levels)

   - From taxable earnings registry (BFL)

     - Indicator for being employed, at least one hour of paid employment per month.

     - Monthly earnings

     - Monthly hours

4   Estimation of main effects

We will compute results for all/unbalanced and a balanced sample of individuals who do not die or outmigrate within one year from responding to the survey.

Baseline model for respondent i:

$$outcome_i = \alpha + \beta\, treated_i + \varepsilon_i \tag{1}$$

where $outcome_i$ is one of the outcomes listed in section 3 for individual i, and $treated_i$ is an indicator variable measuring treatment status of individual i. β captures the difference in means between treated and control and due to randomization of treatment this is the average treatment effect, ATE. We expect $\beta \geq 0$.

We will report β and also show β scaled with relevant belief updating to get closer to a behavioral elasticity.

Heterogeneneous Effects Model:

$$outcome_i = \alpha + \beta_1\, X_i + \beta_2\, treated_i + \beta_3\, treated_i \times X_i + \varepsilon_i \tag{2}$$

where X_i is one of the variables we test heterogeneity in treatment effects for. Full list of variables that we use is provided below in section 5.

5 Heterogeneity

Note that the information treatment was implemented in wave 3 of DARECO (Danish Refugee Cohort). For some respondents, we will have replies from earlier waves which we will use for some of the heterogeneity analysis, i.e. whenever we want a measure closest to arrival in Denmark.

We group potential reasons for heterogeneity in our treatment effects into:

1. Demographic information at arrival:

   - Age at immigration (BEF)

   - Gender in the first entry in the population register (BEF)

   - Highest attained education level before migration from first survey response

   - Labor market participation before migration from first survey response

2. Time constraints:

   - Indicator for employment (at least one hour of paid work) and hours worked (BFL) in the month before the survey

- Presence of small children/partner in Denmark from the month before the survey (BEF)

3. Attachment of the host country

   - Stated return intentions from survey wave 3.

   - Enrollment of children in school/day-care from survey wave 3.

   - Family remaining in Ukraine from survey wave 3.

   - Time since immigration until survey response based on date of registration (VNDS)

4. Language-learning aptitude:

   - PTSD/CPTSD symptoms from survey wave 3.

   - Difficulties concentrating from survey wave 3:

     "Contrating on doing something for then minutes?", "Learning a new task, for example, learning how to get a new place?", "I have been thinking clearly"

5. Perceived relevance of provided information:

   - Field of occupation (language intensive or not) measured by industry (BFL)

   - English proficiency from survey wave 3.

   - Prior beliefs from survey wave 3.

References

Haaland, I., Roth, C., and Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, 61(1):3–40.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46.