Commuting and Residency Choices of Young Adults - Results from Transition to Higher Education in Germany (Plans for analysis)

Dennis Oliver Kubitza

June 20, 2023

Contents

1 Disclaimer

This document describes the state of planed analysis before the first visit at the GESIS Institute in Cologne at the 22nd of June 2023, therefore before the analysis started. The analysis is heavly dependent on the available sample size and distributions of variables, missing values etc. I tried to account as good as possible for possible actions to reduce the set of observations to the most causal ones. It might be that these restrictions have to be relaxed or variables need to be replaced after inspection.

2 Overview

Economic theory about commuting behaviour is, in most cases, linked to the analysis of Work-To-Residence patterns. Labour Economical Researchers consider differences in payments for job/residence changes as an initial assumption for their commuting models and Urban Economists often assume free choices of mode of transportation and/or allocation of housing in a trade-off to time.

Young adults, like university students, are an interesting special case for these models, as they rarely earn any money and still have a limited time budget close to the normal labour force. In their transition to adulthood, they are also faced with a trade of between the leisure time allocation between their new place of living and their hometown. We know that individuals belonging to different for subgroups also experience these incentives differently, even at the most basic levels like gender.

Missing models and heterogeneous effects make it difficult in this setting to estimate relevant policies for reducing inequality in opportunities of individuals or between urbanised and non-urbanised regions regarding higher education.

- Are current subsidiaries of public transport enough to reduce spatial mismatch?
- Are measures to promote commuting already utilised to the maximum?
- Are housing subsidiaries an effective alternative to commuting subsidiaries and for whom ?

From a personal perspective a possible explanation is the Acceptable Travelling Time theory. (Dimitris Milakis et. al 2016), claiming that each individual considers a maximal threshold for tolerating commuting, dependent on their personal valuation on costs.

We present results from logistic regression model on a dataset of university students, starting their Bachelor's in economics. The MESARAS dataset (Weisser, 2016) (N=2589) was surveyed the first weeks of the semester in the introductory courses of 8 German Universities and contains information about personal traits, household information but also the mobility episodes of the participants. Using the information of the last and current place of residence, preferred and enrolled university, we can drive preferences about the place of living in relation to the position of the university.

Dependent variable in our initial regression model is the binary encoded movement decision close to the university. We are interested in the particular commuting time required pre-move as the explanatory variable. We include these hypothetical commuting times of each individual by querying Google MAPS API for different modes of transportation to assess the minimal cost of time. By that we can derive the probability for adolescents to move closer to their desired place of education.

For assessing the different dimensions of influences, we include personal information (gender, age), the family background (number of siblings, active partnership), available budget (loans, scholarships) together with the current accommodation costs, the regional average rent prices and labour market conditions. By utilising the federal educational system in Germany, we can directly compare the effects of a state implementing policies like public transport subscriptions for students.

3 Background / Motivation

Research Questions - Are there socioeconomic groups that are effected more by public subsidiaries for students than others ? - How do public transport subsidiaries interact with housing prices ? - Are current subsidiaries of public transport enough to reduce spatial mismatch for students? - Are public transport commuting times a reasonable choice for answering these questions or are other concepts like distances better for estimating ?

Hypothesis

1. The distribution of likelihoods for movements

   - are different for both genders. - are different for students with and without siblings. - are different for students with and without a current partnership - are different for students coming from an academic household

2. The availability of public transport subsidiaries for a route

   - decreases the chance that students move to their university cities and increases the chance of commuting - the effect increases with higher costs of living at the place of university

3. Models using other geographic measures than public transport commuting times are less plausible.

4. Models using other geographic measures than public transport commuting times reduces the size of effects of public subsidiaries and housing prices.

4 Data

4.1 Panel Dataset

With MESARAS, we have a Dataset of University Students studying Economics. Key characteristic is that it contains the geo-coded location of their previous homes. A full list of all available Variables can be obtained from https://dbk.gesis.org/dbksearch/download.asp?id=59684. The Survey Data (N=2589) was collected in the first weeks of the semester in the introductory courses of 8 German Universities. The Data was collected in 2013. Listed below are the one which seem intersting for our analysis. A more detailed description including reported outcomes of the survey is provided at https://dbk.gesis.org/dbksearch/download.asp?id=59673. Variables that have possibly unsufficient converage will be marked in red. As some variables are not available online a stay at the GESIS secure data rooms is necessary. (Figure **??**)

4.2 Coverage of student tickets

The coverage of student tickets for all moves has been manually researched. (Historic) Sources are provided in the Code for Data Gathering.

4.3 Infrastructure Information

We will use infrastructure information from INKAR citefurstadt and other sources at small regional units (Gemeindeverbände). These will be mapped to the ZIP Codes by identifieng the centroid of ZIP AREAS. Currently we have from INKAR:

- Population Density
- Average Age of Inhabitants
- Net Migration
- Percentage of Students
- Purchasing Power.

We did not found any indicator for cultural activity yet, that is suitable (like the number of cinemas or pubs). If we will find such indicators in the future, we will include them in the analyses.

4.4 Online APIs

The Distance Matrix API is a service that provides travel distance and time for a matrix of origins and destinations. The API returns information based on the recommended route between start and end points, as calculated by the Google Maps API, and consists of rows containing duration and distance values for each pair. We will use the public transport commuting times in two seperate forms: The centroid of ZIP areas and the centroid of the most populated parts of the ZIP Area.

4.5 Relevant Variables

Limit Statistical Population In the best case we will have enough observations for restricting ourselves to freshmans, who did not apply for other universities and ended up in their prefer ed subject and reported all relvant values. Depending on the number of remaining observations, we have to drop some restrictions. The first restriction to be lifted would be the application to other universities.

- *d301_302_303__I_begstud* indicator (detailed) for freshmen status
- *q30700a__otherapp* application for other studies at other universities
- *q30800__econpref* economic studies as preferred studies (lower bound)
- *q31300__resloc* current residential location (postal code or city) (Used for restrictions in terms of robustness checks)

Calculation of Dependent Variable We will calculate if there was move to the place of study directly before or at the start of the studies. Dependent on the encoding in the data we need to calculate that move from the variables below.

- *q31300__resloc* residential location
- *s003__unizip* ZIP-code for university cities (based on main station)
- *q10400__bplace* birth place (postal code, city or country)
- *q10600__uecplace* place of univ. entrance certificate attainment (postal code, city or country)
- *q20100__elschool* living at birth place at elementary school enrolement
- *q20101__elschoolloc* residence at elementary school enrolement (postal code, city or country)
- *q20200__resschool* occurrence of residential changes during school
- *q20201__res1school* destination of last residential change during school (plz, city or country)
- *q20202__res2school* destination of penultimate res. change during school (plz, city or country)
- *q20203__res3school* destination of antepenultimate res. change during school (plz, city or country)
- *q20300__respostschool* occurrence of residential changes after school (lower bound)
- *q20301__res1postschool* destination of last residential change after school (plz, city or country)
- *q20302__res2postschool* destination of penultimate res. change after school (plz, city or country)
- *q20303__res3postschool* destination of antepenultimate res. change after school (plz, city or country)

Independent Variables

- *q10100__gender* gender
- *q10300__siblings* number of siblings (lower bound)
- *d109_110__academichouse1* Indicator for academic parental household (at least one parent as academic) D
- *q11100__partner* partnership (lower bound)
- *d111__partner_comb* partnership status D
- *difference˙in˙accomodation˙cost* monthly accommodation costs difference (between personal budget and average accomodation cost for place of living)
- *student_ticket* Student Ticket available from place of living (pre move). (binary)

Interactions

- *student_ticket x difference˙in˙accomodation˙cost* Student Ticket available from place of living (pre move) x difference˙in˙accomodation˙cost

Controls The Controls will be imposed in several steps: We will use three seperate blocks of controls: Regional indicators, personal indicators from Mesaras and psychological Indicators.

- *q31700__budget* monthly available budget (lower bound)
- *d316_317__tot_budget* total monthly budget (lower bound) D
- *d316_317__tot_budget* total monthly budget (lower bound) D
- *q31801__finparents* share of current expenditures covered by parents (in
- *q10800__uecgrade* average grade of university entrance certificate (lower bound) transformed into four categories in the
- Population Density
- Average Age of Inhabitants
- Net Migration
- Percentage of Students
- Purchasing Power.

- *q40101__movetownrisk*
  moving to another town (in the same state) is risky
- *q40102__movestaterisk*
  moving to another state is risky (lower bound)
- *d60301r_60306__bf_extra*
  Big-5: Extraversion (lower bound) D
- *d60304r_60309__bf_neuro*
  Big-5: Neuroticism (lower bound)
- *d60305r_60310__bf_open*
  Big-5: Openness (lower bound)
- *d60303r_60308__bf_cons*
  Big-5: Conscientiousness (lower bound)
- *d60307r_60302__bf_agreea*
  Big-5: agreeableness (lower bound)

5 Research Design and Methods

- *Data* Limit on students enrolled on prefered university. Analyse binary movement decision, before starting studies: Only consider students that moved closer to the University as 1 for movement decision. Students that moved farther away from university are included as "Without a move" as they are still above their threshold after the move.
- *Regression* Logistic regression analysis with interpretation of average marginal effects. For Interactions we will need to do a visual analysis.
- *Descriptives* Within and between group variances (ICC1) and Boxplots of Movement (0-1) for each University (We want to do this to determine if Multilevel approach is justified)
- *Robustness (if applicable by ICC1*
  Multi-level logistic regression with level 1 - universities. (Fixed parameters) (dependent on the distribution of extreme values, we need to consider other link functions to account for asymmetric observations, Nelder (1989, Aitkin (1989))
- *Robustness* Re-estimate by restriction to distance (one sided- two sided) of pre-movement place of living.
- *Robustness* Re-estimate by using distances instead of public transport communication times
- *Robustness* Use self reported commuting times.