Replication package overview

This document provides detailed instructions on how to reconstruct the data used, and replicate the findings presented in the manuscript „From ethnic prejudice to employment discrimination: The role of small firms as mediators". The submission folder contains all program codes in Stata. Non-proprietary data and replication materials can be found in the folders.

Table below provides an overview of data used in this paper.

**There are two types of data**

* non-proprietary data that the authors can make accessible directly: uploaded data;
* proprietary data that are accessible for replication in the Research Room (RR) jointy operated by the Hungarian Central Statistical Office (CSO) and the Databank of the Center for Economic and Regional Sciences (CERS).

|  |  |  |
| --- | --- | --- |
| **Table, Figure, Appendix**  **in the manuscript** | **Data source** | **Data access** |
| Table1, Table2, AppA | Hungarian Labor Force Survey | Research Room (proprietary data) |
| Table3 | District level data compiled from primary data sources | uploaded |
| Fig1 | Tárki Omnibus Surveys, national opinion poll data 2000-2011 | uploaded |
| Table4, Table8 | Tárki Life Course Survey | Research Room (proprietary data) |
| Table5, Table9,  AppB, AppC, AppD | Hungarian Census 2011 | Research Room (proprietary data) |
| Table7 | Hungarian Census 2001 | Research Room (proprietary data) |

**Proprietary data access for refereeing and replication**

This research was conducted in a Research Room (RR) jointly operated by the Hungarian Central Statistical Office (CSO) and the Databank of the Center for Economic and Regional Sciences (CERS). For a description of the RR and the rules of work, visit <https://adatbank.krtk.mta.hu/en/kutatoszoba/>

**Background**. The CCTV-controlled RR is available for projects led by researchers at the CERS and the Center for Social Sciences. Other researchers, including assistants, are also eligible as approved project participants. Any output produced in the RR must undergo an output-checking procedure conducted by the CSO. Copying data from the servers and computer screens is strictly forbidden.

**Access for referees.** If a referee wants to check the data and computations, the Databank starts a Referee Project in which the referee acts as a project participant. The referee can work in the RR personally or by hiring a local assistant. The required data will be uploaded to the referee's personal folder. The names of the referees and assistants are kept confidential and will not be shared with the authors.

**Access for replication.** The person willing to replicate the results can join the author’s project and attend the RR in person.

For further information, contact the Databank: [adatkeres@krtk.mta.hu](mailto:adatkeres@krtk.mta.hu)

**There are six folders in the replication package**

* Hungarian Labor Force Survey
* LFS\_readme.pdf
* Analysis
* District level
* DISTRICT\_readme.pdf
* Data
* Analysis
* Tárki Omnibus Surveys, national opinion polls 2000-2011
* TOS\_readme.pdf
* Data
* Tárki Life Course Survey
* LCS\_readme.pdf
* Analysis
* Census 2011
* CENSUS2011\_readme.pdf
* Processing
* Analysis
* Census 2001
* CENSUS2001\_readme.pdf
* Processing
* Analysis

**Labor Force Survey (LFS)**

The LFS is a quarterly rotating panel conducted by the Hungarian Central Statistical Office (CSO) since 1992. The survey provides information on the population's economic activity, following Eurostat guidelines.

The sampling units are dwellings. Each person in the dwelling are asked about gender, age, place of birth, ethnicity, household characteristics, and other essential attributes. An 'activity questionnaire' with detailed information on education, employment, employers, job search, and transfers targets people aged 15-74. The observed households varied between 22,000 and 34,000 in 1992-2024. The activity questionnaire covered 45,000-70,000 individuals per quarter. The sample is drawn using a multi-stage stratified sample design. Weights ensure that the sample is representative. Each cohort stays in the sample for six consecutive quarters. Individuals are identifiable across waves. For detailed information, click on 'LFS guide' at <https://adatbank.krtk.mta.hu/en/adatbazisok/elerheto-adatbazisok/>

Sample used in this paper: the panel covers 356,753 quarterly observations on 104,977 individuals. (9,865 observations on 3,301 Roma individuals). see *Supplementary materials SM5.*

Access to data is provided in the Research Room of CSO-CERS.

**District level data file**

District (NUTS4 / LAU1 / in Hungarian: járás) level data compiled from primary data sources. On primary data sources see *Supplementary materials SM5*.

|  |  |
| --- | --- |
| Variable | Primary data source |
| Districts id and name |  |
| Roma share 2011 | Census 2011 |
| men / women |
| Employment rate 2011 |
| Roma men / Roma women |
| non-Roma men / non-Roma women |
| Raw ethnic employment gap 2011 |
| Residual ethnic employment gap 2011 |
| Industry share 2011 |
| Agriculture |
| Construction |
| Services |
| Fraction of non-employed…. |
| Roma searching for jobs 2011 |
| non-Roma searching for jobs 2011 |
|  |  |
| Roma share 2001 | Census 2001 |
| men / women |
| Employment rate 2001 |
| non-Roma men / non-Roma women |
| Roma men / Roma women |
| Residual employment gap 2001 |
|  |  |
| Small firm share 2011 | Hungarian Wage Survey 2011, Hungarian Business Register 2011 |
| Small firm share 2001 | Hungarian Wage Survey 2001, Hungarian Labor Force Survey 1999-2003 |
| Jobbik share 2010 | National election results 2010 |
| Residual Jobbik share 2010 |
| Q1, … Q5 quintiles by residual Jobbik share 2010 |
| Outstanding hard-currency loabs / population 2012 | BISZ Central Credit Information System |
| Sales revenues of the employers of district residents 2008 | Admin3 |
| year-2008 data / year-2010 data |
| Interschool segregation index: mean of 1985, 1989, 1992 school years | Ministry of Public Education and Culture 1985, 1989, 1992 |

district.dta file can be found in the Data folder.

**Tárki Omnibus Surveys, national opinion polls 2000-2011**

Tárki Omnibus Surveys 2000–2011 measure social and political attitudes in Hungary among the adult population in national representative samples. Individual data are accessible from Tárki Social Research Institute: <https://adatbanktest.tarki.hu/cgi-bin/katalogus/tarkiser_en.pl>

see *Supplementary materials SM5.*

National data of Figure 1 can be found in the Data folder in excel form.

**Tárki Life Course Survey (LCS)**

The LCS 2006-2012 is a panel survey that follows 10,000 youths on an annual basis, beginning in the fall of 2006 and ending at the 2011/12 school year. The survey sampled regular students who participated in the National Assessment of Basic Competencies (NABC) as primary school graduates (8th grade students) in May 2006. (These administrative data cover the entire population of 8th grade students, excepting those absent on the day of testing.) The questions in the first wave of the TLCS in 2006 focused on the respondents’ family structure, financial situation, early childhood experiences, medical and school history and plans for secondary school. Subsequent waves of the survey primarily concentrated on school careers and the mechanisms underlying student dropout.

<https://adatbanktest.tarki.hu/cgi-bin/katalogus/tarkiser_en.pl>

In this paper we use data from the Youth Questionnaire of the wave 4, 2009 (questions 5-9):

<http://adatbanktest.tarki.hu/adatbank-h/katalog/dokument/kerdoiv/h84_ques_complementary.pdf>

see *Supplementary materials SM5.*

Access to data is provided in the Research Room of CSO-CERS.

**Census 2011**

see *Supplementary materials SM5.*

Access to data is provided in the Research Room of CSO-CERS.

**Census 2001**

see *Supplementary materials SM5.*

Access to data is provided in the Research Room of CSO-CERS.