Codebook of variables

Timing of citizenship acquisition and immigrants' children educational outcomes: a family fixed-effects approach

Marie Labussière

Contents

Identifiers 1
Educational variables 1
Naturalisation variables 2
Parental country of origin variables 3
Sociodemographic variables 4

Identifiers

rinpersoon	CBS Respondent's anonymous identifier	9 digit
rinpersoon_ma	CBS Anonymous identifier of the respondent's mother	9 digit
rinpersoon_pa	CBS Anonymous identifier of the respondent's father	9 digit
sibling_id_ma	Identifier for siblings, i.e., respondents registered with the same mother	
ind_in_SFE	The respondent is included in the sample of exposure discordant families	Time-constant dummy

Educational variables

cito_score_st_year	Respondent's Cito test score. If the respondent took the test multiple times, the score obtained at the first attempt is taken into account. Standardized by year *See detailed note about respondents' Cito test score on p.2 in the Supplementary Materials.*	Time-constant variable
track_4_cito	School track in which the respondent is registered the year after taking the Cito test. Based on CBS variable *onderwijssoortvo*.	Time-constant variable 1 – VMBO (vocational) 2 – Bridge 3 – HAVO (general) 4 – VWO (academic) 5 – PO (Practical education)
age_cito	Age at which the respondent takes the Cito test (for the first time).	Time-constant variable

Naturalisation variables

sum_nat_t5	Naturalisation **or** change from unknown to Dutch nationality is observed during the observation period	Time-constant dummy .a – atypical pattern .b – Dutch from birth
date_nat_t5	Date of naturalisation (based on *sum_nat_t5*). *Note: atypical naturalisation patterns (".a") represent 22 respondents (0.02% of the sample).* *See document "Data_construction.pdf" for more information.*	Time-constant variable 4-digit .a – atypical pattern .b – Dutch from birth .c – Does not naturalise over the observation period
age_nat_t5_wmiss	Age at naturalisation, censored at age 21, i.e., equals 21 is the respondent is still a foreign citizen at the end of the observation period. The variable is based on a recoding of *date_nat_t5:* - For those with atypical naturalisation patterns (previously ".a"), the last change in nationality is taken into account - For the Dutch from birth (previously ".b"), age is set to zero - For those who do not naturalise over the observation period (previously ".c"), age is set to 21, which is the maximum age at which respondents can be observed during the observation period	Time-constant variable
age_nat_13_censored	Age at naturalisation, censored at age 14, i.e., equals 14 if the respondent is still a foreign citizen at the age of 14. Based on *age_nat_t5_wmiss.* *See detailed note about respondents' age at naturalisation on p.3 in the Supplementary Materials.*	Time-constant variable 0 – 0 (Dutch from birth) 1 – 1 2 – 2 3 – 3 4 – 4 5 – 5 6 – 6 7 – 7 8 – 8 9 – 9 10 – 10 11 – 11 12 – 12 13 – 13 14 – 14 and above
age_nat_13_cat7	Age at naturalisation categorized in 7 clusters, censored at age 14, i.e., equals 14 if the respondent is still a foreign citizen at the age of 14. Based on *age_nat_t5_wmiss.*	Time-constant variable 0 – 0 (Dutch from birth) 1 – 1-2 2 – 3-4 3 – 5-6 4 – 7-10 5 – 11-13 6 – 14 and above
Nat_13_censored	The respondent naturalised before the age of 14. Based on *age_nat_t5_wmiss.*	Time-constant dummy

Parental country of origin variables

gbaherkomstgroepering_ma	Mother's country of origin (non-clustered), original CBS variable.	Time-constant variable 4-digit code (see SSB Catalogue)
gbaherkomstgroepering_pa	Father's country of origin (non-clustered), original CBS variable.	Time-constant variable 4-digit code (see SSB Catalogue)
country_of_birth_ma_agg	Mother's country of birth at the year of birth, grouped into 9 categories. It is based on CBS variable *gbaherkomstgroepering*, which gives the country of birth for first-generation respondents. *Note: If the mother's country of birth is not filled in the year of birth (e.g. if the mother's identifier is missing), it is replaced by later values.* We use a time-constant variable because a small share of respondents (<1%) change category due to EU enlargement.	Time-constant variable 1 – Afghanistan 2 – China 3 – Irak 4 – Iran 5 – Morocco 6 – Other EU countries 7 – Other non-EU countries 8 – Turkey 9 – Yugoslavia
country_of_birth_pa_agg	Father's country of birth at the year of birth, grouped into 9 categories. It is based on CBS variable *gbaherkomstgroepering*, which gives the country of birth for first-generation respondents. *Note: If the father's country of birth is not filled in the year of birth (e.g. if the father's identifier is missing), it is replaced by later values.* We use a time-constant variable because a small share of respondents (<1%) change category due to EU enlargement.	Time-constant variable 1 – Afghanistan 2 – China 3 – Irak 4 – Iran 5 – Morocco 6 – Other EU countries 7 – Other non-EU countries 8 – Turkey 9 – Yugoslavia
EU_child	At least one parent is from a EU country during the year Based on *country_of_birth_ma_agg* and *country_of_birth_pa_agg*.	Time-varying dummy
EU_sum_child	At least one parent is from a EU country over the observation period	Time-constant dummy

Sociodemographic variables

gender	Gender Original CBS variable *gbageslacht*.	Time-constant variable 1 – Male 2 – Female
first_born_FULREG	Birth order of the respondent, calculated on the full population registers.	Time-constant variable 1 – first-born 2 – non first-born
date_birth_educ	Year of birth, based on CBS variable *gbageboortejaar*.	Time-constant variable 4-digit
date_birth_cat	Year of birth, grouped into 9 groups, based on CBS variable *gbageboortejaar*. *Note: the very few students born before 1995 or after 2003 are clustered in the nearest available category.*	Time-constant variable 1 – 1995 or before 2 – 1996 3 – 1997 4 – 1998 5 – 1999 6 – 2000 7 – 2001 8 – 2002 9 – 2003 or after
secm_pa/ma_agg2_cito	Socioeconomic status of the respondent's father (*pa*) or mother (*ma*) the year the respondent took the Cito test Based on CBS variable *secm*.	Time-constant variable 1 – Employee or self-employed 2 – Reception of benefits 3 – No declared income 4 – Missing SECM
q3_st_disp_income_cito	Terciles of standardised disposable household income, measured the year the respondent took the Cito test. Based on CBS variables *hvrgestinkh* before 2011 and *inhgestinkh* from 2011.	Time-constant variable 1 – First tercile of income 2 – Second tercile of income 3 – Third tertile of income
homeowner_alo_cito	At least one parent was registered as a homeowner the year the respondent took the Cito test. Based on CBS variables *hvrehalg* before 2011 and *inhehalgrpl1* from 2011.	Time-constant dummy
parent_ed_level_miss	Highest educational degree among parents *See detailed note about the measurement of immigrants' level of education on p.2 in the Supplementary Materials.*	Time-constant variable 1 – Low education (primary education) 2 – Middle education (completed some secondary education) 3 – High education (bachelor or higher) 4 – Missing education
type_household_cito	Type of household in which the respondent lived in the year he or she took the Cito test. Based on CBS variable *typhh*.	Time-constant variable 1 – Two-parent household 2 – Single-parent household
nbr_children_cito	Number of minor children registered in the household the year the respondent took the Cito test. Based on CBS variable *aantalkindhh*.	Time-constant variable Range 1-7