SOEP

The Socio-Economic Panel

872

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP-Core v35 – MIGSPELL and REFUGSPELL: The Migration-Biographies

Klaudia Erhardt, Paul Schmelzer, and SOEP Group

DIW SOEP

Running since 1984, the German Socio-Economic Panel (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:
Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at **http://www.diw.de/soepsurveypapers**

Please cite this paper as follows:

Klaudia Erhardt, Paul Schmelzer, and SOEP Group. 2020. SOEP-Core v35 – MIGSPELL and REFUGSPELL: The Migration-Biographies. SOEP Survey Papers 872: Series D. Berlin: DIW/SOEP

SOEP-Core v35 – MIGSPELL and REFUGSPELL: The Migration-Biographies

Klaudia Erhardt, Paul Schmelzer, and SOEP Group

MIGSPELL and REFUGSPELL: The Migration-Biographies[1]

Klaudia Erhardt and Paul Schmelzer

MIGSPELL

MIGSPELL is derived from the migration biographies, which are collected from each new respondent to the IAB-SOEP Migration Survey, samples M1 and M2. The moves of foreign-born migrants, as well as of German-born respondents, were captured through a complex loop structure within the questionnaire, with the number of loops limited to 15.

The original files of the SOEP-distribution are provided in "wide" format. To create MIGSPELL, the original data has been transformed into spell format. Each migration that actually took place is represented by a spell. Additionally, a spell is generated for the period from birth to first move (if there has been any), or from birth to interview date (if the respondent has not moved). As a result, all respondents in the IAB-SOEP migration samples are represented in the MIGSPELL dataset, not just those who actually had moved to another country. In the Table 1 we report variables that were constructed for MIGSPELL data.

---

[1] For detailed information of MIGSPELL see
https://www.diw.de/documents/publikationen/73/diw_01.c.605593.de/diw_ssp0538.pdf

Table 1:   Variables in MIGSPELL

Variable	Variable label	Description
Identifiers		
cid or hhnr	Original Household Number	Identifier throughout the SOEP
pid or persnr	Never Changing Person ID	Identifier throughout the SOEP
mignr	Running no. of move	MIGPELL-specific identifier. The combination of persnr and mignr identifies a spell. Runs from 0 because first spell is not a move but episode from birth to first move or interview date, respectively.
sample1	Sample Indicator	Sample indicator throughout the SOEP
Time-invariant Characteristics		
sample1	Sample Indicator	
intyear	Interview year	Year when migration biography was surveyed
migfilter	German-born/abroad-born	
bcountry	Country of birth	
lastmig	Year of last move to Germany	= starty of the last spell of a person (only foreign-borns)
lastmig_imp	Year of last move to Germany (imputed version)	= starty_imp of the last spell of a person (only foreign-borns)

Time-variant Characteristics

starty	Start year	Start year of an episode. Birth year in first spell of a respondent.
startmo	Start month	Start month of an episode. Birth month in first spell of a respondent.
starty_imp	Start year imputed	Same as starty, with imputed missing values and birth year updated with birth year from PPFAD
startmo_imp	Start month imputed	Same as startmo, with imputed missing values and birth month updated with birth month from PPFAD
start	Start (months from jan 1900)	generated from starty_imp and startmo_imp
end	End (months from jan 1900)	= start of next spell - 1
stype	Spell type	Move to Germany / move abroad
country	Country of the stay	
status1	Legal background of entry	source variables .
status2	Status at entry	source variables
ostatus	Status at entry: unstandard. answers	derived from open answers of level "other" of status2
jobpr	Job agreement at entry	source variables .
lfgroup	Labor force group	source variables .
move	Type of move	to Germany / back to birth country / to another country
nmtype	Type of next move	to Germany / back to birth country / to another country
tcountry	Target country of next move	
staytime	Duration of stay (months)	end - start + 1 (generated from imputed date variables)

Technical Variables

censor	Censor	Censor Variable generated from original dates
censor_imp	Censor imput. version	Censor Variable generated from imputed dates
nspells	Number of spells	Number of moves of a respondent + 1 (= mignr(max) + 1)
f_birth-date_corr	Flag: birth-date updated with birth-date from PPFAD	= 1 if starty_imp and/or startmo_imp of first spell of a person is updated with PPFAD
f_starty_imp	Flag: starty imputed	= 1 if missing starty has been imputed
f_startmo_imp	Flag: startmo imputed	= 1 if missing startmo has been imputed

The date-variables in MIGSPELL

With spell data, the information on start date and end date of each episode is crucial. The migration biographies constitute a special kind of spell data, without parallelities and gaps: Because the questionnaire does not allow for reporting more than one place of habitation at the same time, the spells of a person cannot be parallel to each other.

The participants were only asked at what time they moved to a certain country, but not at what time they left. As one necessarily has to stay at one place or another, the episodes are successive to each other, and the end of a spell could be derived from the start of the next spell. Therefore the migration biographies have no gaps.

However, sometimes respondents were not able to recollect the exact dates of their moves, or they named dates that contradicted the time sequence of moves, which led to missing values in the date variables. The respondents were not forced to state a date, and time contradictions were not immediately clarified during the interview, so these kinds of missings were to be expected.

In order to make those spells accessible for data analyses, missing values were replaced with imputed values to the extent possible.

From the imputed versions of startyear and startmonth, the start and `end` variables were generated. These contain integers counting the months that have passed since January 1, 1900, (i.e. including January). This is in contrast to the SOEP standard, where the counting of time begins January 1, 1983, and also to the Stata standard, where counting begins January 1, 1960.

We changed the zero-point of the time scale in MIGSPELL because, otherwise, the SOEP-convention on missing codes (which are integers < 0 and > -10) would have conflicted: Some migration biographies begin earlier than January 1983 or 1960, which results in negative values if we had used the SOEP or Stata standard. In certain cases, missing and valid codes would then have become indistinguishable.

If you are using Stata for data analyses and want to benefit from the inbuilt Stata time and date functions, you can transform the `start` and `end` variables into the Stata standard by sub-tracting 721. Before doing so, you must replace the SOEP missing values with values that are distinguishable from valid values, such as the Stata missing codes:

```
gen start_stata = start
gen end_stata = end
recode start_stata end_stata (-1 = .a) (-3 = .b)
replace start_stata = start_stata - 721 replace
end_stata = end_stata - 721 format start_stata
end_stata %tmCCYY_mon
```

REFUGSPELL[2]

Summary description of REFUGSPELL

REFUGSPELL is derived from the migration biographies, which are collected from each new respondent of the IAB-BAMF-SOEP Refugees Survey, samples M3, M4 and M5. The reported moves of the respondents were captured through a loop structure of the questionnaire, with the number of loops limited to 15. In comparison to MIGSPELL, the loop structure of the migration biographies in REFUGSPELL is much less complex.

The original files of the SOEP-distribution are provided in "wide" format. To create REFUGSPELL, the original data has been transformed into spell format. Each migration that actually took place is represented by a spell. Additionally, a spell is generated for the period from birth to first move (if there has been any), or from birth to interview date (if the respondent has not moved). In the Table 2 we report variables that were constructed for REFUGSPELL data.

---

[2] For detailed information of constructing REFUGSPELL see
https://www.diw.de/documents/publikationen/73/diw_01.c.605593.de/diw_ssp0538.pdf

Table 2:   Variables in REFUGSPELL

Variable	Variable label	Description
Identifiers		
hhnr	Original Household Number	Identifier throughout the SOEP
persnr	Never Changing Person ID	Identifier throughout the SOEP
mignr	Running no. of move	REFUGPELL-specific identifier. The combination of persnr and mignr identifies a spell. Runs from 0 because first spell is not a move but episode from birth to first move or interview date, respectively.
sample1	Sample Indicator	Sample indicator throughout the SOEP
Time-invariant Characteristics		
intyear	Interview year	Year when migration biography was surveyed
migfilter	German-born/abroad-born	in REFUGSPELL: only abroad-born (see note (1) on page 22)
bcountry	Country of birth	
lastmig	Year of last move to Germany	= starty of the last spell of a person
lastmig_imp	Year of last move to Germany (imputed version)	= starty_imp of the last spell of a person (see note (2) on page 22)
lastmig_refug	Year of last move to Germany (extra question)	For example for wave 33 information is used from Question 34, variable bgpr_l_3401 of bgp_refugspell

Time-variant Characteristics

starty	Start year	Start year of an episode. Birth year in first spell of a respondent.
startmo	Start month	Start month of an episode. Birth month in first spell of a respondent.
starty_imp	Start year imputed	Same as starty, with imputed missing values and birth year updated with birth year from PPFAD
startmo_imp	Start month imputed	Same as startmo, with imputed missing values and birth month updated with birth month from PPFAD
start	Start (months from jan 1900)	generated from starty_imp and startmo_imp
end	End (months from jan 1900)	= start of next spell - 1. See note (3) on page 22
stype	Spell type	Move to Germany / move abroad
country	Country of the stay	
rexit	Reasons for emigration	source variables
move	Type of move	to Germany / back to birth country / to another country
nmtype	Type of next move	to Germany / back to birth country / to another country
tcountry	Target country of next move	
staytime	Duration of stay (months)	end - start + 1 (generated from imputed date variables)
Technical Variables (time variant and time invariant)		
censor	Censor	Censor Variable generated from original dates
censor_imp	Censor imput. version	Censor Variable generated from imputed dates
nspells	Number of spells	Number of moves of a respondent + 1 (= mignr(max) + 1)
f_birth-date_corr	Flag: birth date updated from PPFAD	= 1 if starty_imp and/or startmo_imp of first spell of a person is updated with PPFAD if a later interview gave different birth-date information
f_bio_corr	Flag: corrections in migbio	= 1 for all spells of a respondent, If one or more spells have been removed from the migration biography.
f_starty_imp	Flag: starty imputed	= 1 if missing starty has been imputed
f_startmo_imp	Flag: startmo imputed	= 1 if missing startmo has been imputed

| f_bcountry_iffy | Flag: info birth country dubious | see note (1) below |
| f_short | Flag: more than one spell begin in same month | = 1 if the next spell begins in the same month |

The date variables in REFUGSPELL

The description of the date variables in MIGSPELL also applies fully to the date variables in REFUGSPELL.