pairfam

Panel Analysis of Intimate Relationships and Family Dynamics

Regional Data in the German Family Panel (pairfam)

Claudia Schmiedeberg

December 2015

Funded as long-term project by the German Research Foundation (DFG)

Cite as:
Schmiedeberg, Claudia (2015): Regional Data in the German Family Panel (pairfam). pairfam Technical Paper No. 07. https://doi.org/10.5282/ubm/epub.91989

pairfam Technical Paper

07

1. Introduction

Over the past years, spatial aspects of social phenomena have become more and more a central focus of social scientists. The idea that individual opportunities and behavior patterns depend on characteristics of one's surroundings is not new, but due to the increasing availability of spatial data and the development of tools for spatial analysis, a "spatial turn" in the social sciences has been observed.

To enable spatial analysis using the German Family Panel (pairfam), spatial data has been made available for pairfam users. All spatial data refer to the location of respondents' main residence. The main residence refers to the address where the survey agency has contacted the respondent.

In the following, a documentation of the data available for spatial analysis of pairfam data is provided. The documentation refers to Release 6.0 of the data (Brüderl, Hank et al., 2015). For an introduction to pairfam see Huinink et al. (2011), for details regarding sampling, data collection, and data editing, see Brüderl, Schmiedeberg et al. (2015) and Brüderl, Hajek et al. (2015).

2. Regional data in the Scientific-Use-File (SUF)

The scientific use file (SUF) contains the *federal state* (*bula*) of the respondent's place of residence. In addition, basic information about the place of residence is given such as community size (*gkpol*, 7 categories) and 10 types of settlement structures (*bik*, BIK-type). This data is included for each panel wave. Data on federal state, municipality size, and settlement structure is included also in the SUF for the anchor person's parents.

Table 1: Community size classes

1	1,000 - 2,000 inhabitants
2	2,000 - 5,000 inhabitants
3	5,000 - 20,000 inhabitants
4	20,000 - 50,000 inhabitants
5	50,000 - 100,000 inhabitants
6	100,000 - 500,000 inhabitants
7	500,000 + inhabitants

Table 2: Settlement structures

0	City Center - population 500,000+
1	Periphery - population 500,000+
2	City Center - population 100,000-500,000
3	Periphery - population 100,000-500,000
4	City Center - population 50,000-100,000
5	Periphery - population 50,000-100,000
6	Region - population 20,000-50,000
7	Region - population 5,000-20,000
8	Region - population 2,000-5,000
9	Region - population < 2,000

To facilitate mobility analyses, in the generated datasets (bio_mob and bio_rtr) *migration distance* for all moves within the observation period of the panel as well as with respect to retrospective data have been included. Migration distance is calculated using geographically referenced data (coordinates) of consecutive places of residence (for details, see Brüderl, Hajek et al., 2015).

3. Geographically Referenced Data (Geodata)

As the concept of Geodata may be unusual/new for many social scientists, we shortly introduce the main aspects in the following. For a more detailed description of research potential, challenges, and methods, see Hintze and Lakes (2009) and Logan (2012).

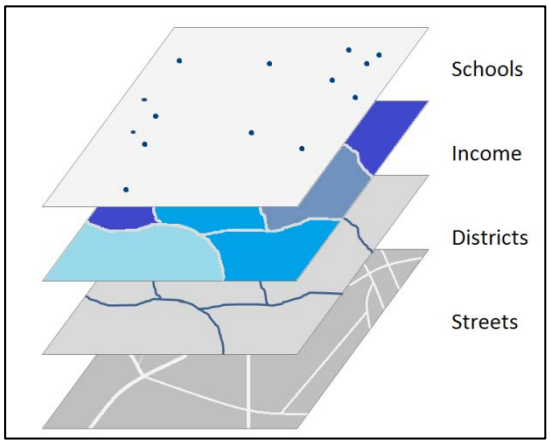
3.1. Short Introduction to Geodata and Geospatial Analysis

Geographically referenced objects are defined by attribute data and spatial data. In case of household addresses, these attributes may be the common socio-demographic characteristics of a household. But in general, all types of objects can be geographically referenced, such as firms, streets, trees, or lakes, and accordingly, attribute data may vary widely from, say, dimensions and year of construction for a bridge to headcount, working atmosphere and the CEO's name for a company. Spatial data define the location of the object, in case of indirect spatial reference mainly via postal addresses, in case of direct spatial reference via two- or three-dimensional geographical coordinates. By geocoding and reverse geocoding postal addresses can be converted to geographic coordinates and vice versa.

For the storage, management and analysis of Geodata a large number of systems, called *Geographic Information Systems (GIS)*, have been developed over the past decades, and today, both commercial and open source systems are available for social scientists (e.g. ArcGIS by the company Esri or GeoDA).

Spatial data is organized in the GIS in form of map layers, so that separate files (layers) contain the different types of geographic features (see figure 1). Data can be either vector data, stored as series of geographic coordinates, or raster data, consisting of a geo-referenced rectangular grid of pixels. Pixel size is not fixed and determines the spatial accuracy of the raster picture. Data can be restructured by raster-to-vector transformation.

Figure 1: Layers in a GIS



Using GIS, a variety of analytical methods can be applied. By *map overlay*, for instance, two or more spatial datasets are combined, just like superimposing two maps of the same region, to study the relationships between them. By overlay analysis, for instance, noise maps of airports can be used to determine the most burdened residents. *Proximity analysis* is used to establish the distance between objects, for instance, the distance of schools to the next mobile phone mast. A further proximity analysis tool is buffers, which are used to identify objects within a given distance, e.g. the schools within a 200m buffer of mobile phone masts. Moreover, methods such as neighborhood analysis, spatial cluster analysis, or Geographically Weighted Regression (GWR) are possible.

3.2. Geodata in pairfam

For each pairfam wave, *geographic coordinates of anchor respondents' main residences*[1] exist. Using these coordinates, pairfam data can be combined with geodata from external sources. As data protection prohibits any form of reverse geocoding of pairfam data to locate respondents, pairfam geographic coordinates are not accessible for users. Therefore, the pairfam user service will merge pairfam data with data provided by the user based on geographic coordinates so that the user can work with the resulting dataset. This means, the researcher prepares a dataset containing geographic coordinates of the entities of interest (e.g. child care facilities or nuclear power plants) and specifies the merging algorithm (e.g. the distance from the anchor's residence to the next child care facility, or a binary variable indicating if an anchor lives nearer than 20 km to a nuclear power plant). The pairfam user service will generate a dataset containing the anchor id and the specified geographic variables. This dataset can then be merged to pairfam data. If the variables potentially enables the researcher to locate respondents (e.g. because values allow to identify certain municipalities), the researcher will have to work with this dataset on a special workplace at one of the pairfam locations. Otherwise, if the possibility to locate respondents is excluded, the dataset will be sent to the researcher.

In addition, the *official municipality key* (*gkz*, Gemeindekennziffer) and the *official district key* (*kkz*, Kreiskennziffer) are available for the anchor respondents' main residence for each panel wave. The municipality key consists of 8 digits, of which the first two indicate the federal state (Bundesland), followed by one digit indicating the administrative districts (Regierungsbezirke) and further two digits indicating the county (rural and urban districts, Stadt- und Landkreise). The last three digits give the municipality (except for urban districts which are not further subdivided). The district key consists of the first five digits of the municipality key. In Germany, 402 (rural and urban) districts and 11,220 municipalities existed as of December 2012 (Statistisches Bundesamt, 2014). For instance, data of the fifth wave of pairfam (2012/2013) contains data from 336 rural and urban districts identified by the district key, and from 965 municipalities identified by the municipality key. For two respondents, no information about their district or municipality is available. On average, 22 respondents from the same district and 8 respondents from the same municipality were surveyed in the fifth wave, with a maximum of 264 respondents in Berlin. Berlin has different municipality keys for East and West Berlin, i.e. Berlin is treated as two municipalities with 138 and 126 respondents in wave 5, respectively.

Analyses with *official municipality / district keys (gkz / kkz)* are possible at special work stations at any pairfam location. Thereby, the procedure is as follows: The researcher prepares a dataset containing the official municipality or district keys and the variables of interest (e.g. sex ratio or unemployment rates). At the pairfam work station, this dataset can be merged to a dataset containing the anchor id as well as the *gkz* and *kkz* of their main residence (separately for each panel wave). The resulting data can be combined with the 'normal' pairfam data of the current release available on the work station. Alternatively, if it is impossible to identify a respondent's place of residence with the geographic information merged (e.g. as it is impossible to identify a municipality

[1] In 99% of the cases, addresses could be referenced on house level. The remaining cases were referenced on municipality level, and a very small number of cases (e.g., 14 in wave 1) could not be referenced due to inconsistencies in the address data. It should be noted that addresses are not necessarily identical to the residences recorded in the interview, but refer to the addresses the survey institute uses to contact respondents.

by its sex ratio with one decimal), the pairfam user service may merge the prepared dataset with the *gkz/kkz* and send the resulting dataset back to the researcher so that the researcher does not have to travel to one of the pairfam locations.

A detailed example of the procedures is given by Hensel, Kreyenfeld, and Walke (2015).

4. microm data

Data of the first five waves is enriched with data delivered by microm consumer marketing. This data stems from a number of sources such as the federal statistical offices, governmental agencies (e.g. employment agency), and market research companies (e.g. Sinus, Gruppe Nymphenburg). While the primary purpose of this micro-geographical data is commercial, the variables can be of use for sociological research as they provide respondents' neighborhood context.

In the pairfam dataset microm data from the years 2008-2012 are included, corresponding to the first five waves of the panel. Data is available in separate cross-sectional datasets for each of the panel waves, i.e. microm1.dta, microm2.dta, etc., containing the microm variables, the panel wave, and the anchor ID so that the microm datasets can be merged with the anchor datasets of the respective waves.

Data for the first DemoDiff wave is available in a separate dataset (microm1_DD.dta) as data corresponds to the fielding year of the wave. Starting with wave 3, the microm datasets include both pairfam and DemoDiff respondents.

Due to data protection issues, microm data is not included in the SUF, but can be accessed for analyses on-site at any pairfam location.

Data is available on the house level (microcells/microm-segments), street level, for small districts ("market cells" in wave 1-4 and "PLZ8-districts" in waves 1-5), as well as on the postal code and municipality levels. Houses with more than 5 households are microcells on their own, otherwise houses in the same street with similar structures are pooled to reach the minimum of 5 households. On average, microcells encompass 7.5 households. The street level refers to sections of one side of the street (i.e. only odd or even street numbers) with approximately 27 households. Market cells and PLZ8-districts are constructed by segmentation of postal code areas and contain about 500 households on average.

Table 3: Overview of microm variables

	House level	Street level	District level	Municipality level
Sociodemographic data	X	X	X	X
Building data	X	X	X	X
Mobility data	X		X	X
Life stages	X	X	X	X
microm types & groups	X	X	x	X
microm Geo Milieus	X		X	X
Rented/owned dwellings*		X		
District size			X	
Population density*			X	
Age structure			X	X
Ethnicity*			X	
Unemployment rates			X	X
District types			X	
Pedestrian area, culture & leisure facilities*			X	X
Religious denominations				X

* only wave 1

Table 3 shows which data is available on which of the four aggregation levels. In the following sections, an overview of the variables available is given. It is based on the microm data manuals (microm Consumer Marketing, 2010, 2015). Variables are described on the most detailed geographic level available although in most cases data is available on all more aggregate levels (as specified in the following sections, respectively). In Section 4.1, each variable available on the house level is included, with variables from street level in section 4.2, in Section 4.3 variables on the district level, and in section 4.4 variables on the community level. Values in the tables are based on wave 1 data (excluding DemoDiff). Variables not included in wave 1 data are shown using wave 2 data. Variables with names starting with *"ha"* refer to house level, variables starting with *"st"* to street section level, variables starting with *"p8"* or *"mz"* to district level, and variables starting with *"gk"* to municipality level.

4.1. House level data

4.1.1. Sociodemographic data

Age structure is captured by the mean age of the household head (*ha_mso_k_alter*), the share of household heads below age 30 (*ha_mso_k_alter30*), and the share of household heads above 60 (*ha_mso_k_alter60*). The variables are categorical, with 5-year age groups for the first variable and 5%-categories in the latter two variables. At the aggregated levels (street, district, and municipality levels) the information is given in a set of metric variables indicating the number and the percentage of households with a household head in one of the age groups. For instance, the variable p8_mso_p_alter_1 gives the percentage of households in the district with household heads below age 36. Data is available both for market cells and for PLZ8-regions in wave 1.

Table 4: Age of household head (*ha_mso_k_alter*)

	Mean age	Frequency	Precent
.	Incomplete data	969	7.81
1	≤35 years	1,074	8.66
2	>35 to ≤40 years	1,447	11.67
3	>40 to ≤45 years	2,065	16.65
4	>45 to ≤50 years	2,184	17.61
5	>50 to ≤55 years	1,804	14.55
6	>55 to ≤60 years	1,213	9.78
7	>60 to ≤65 years	827	6.67
8	>65 years	784	6.32
99	Only commercially used	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

Table 5: Household heads below 30 and above 60 *(ha_mso_k_alter30 / ha_mso_k_alter60)*

	Share of household heads…	… below age 30		… above age 60	
		Frequency	Percent	Frequency	Percent
.	Incomplete data	969	7.81	969	7.81
0	≤5%	1,476	11.90	2,509	20.23
1	>5% to ≤10%	941	7.59	1,143	9.22
2	>10% to ≤15%	1,070	8.63	1,129	9.10
3	>15% to ≤20%	1,142	9.21	1,106	8.92
4	>20% to ≤25%	1,091	8.80	964	7.77
5	>25% to ≤30%	1,081	8.72	914	7.37
6	>30% to ≤35%	917	7.39	799	6.44
7	>35% to ≤40%	794	6.40	616	4.97
8	>40% to ≤50%	1,143	9.22	953	7.68
9	>50%	1,743	14.05	1,265	10.20
99	only commercially used	35	0.28	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

The **number of children per household** (*ha_mso_k_kinder*) is available only for wave 1 and 4 on the house level. Instead of metric values, a variable with 9 categories is given. Data originate from two private data bases (Felicitas Adressen & Service GmbH and Verband der Vereine Creditreform).

Table 6: Children per household (*ha_mso_k_kinder*)

		Frequency	Percent
.	Incomplete data	953	7.68
1	lowest share	1,412	11.39
2	very low share	832	6.71
3	share far below average	964	7.77
4	share below average	1,085	8.75
5	share slightly below average	1,140	9.19
6	average share	1,250	10.08
7	higher-than-average share	1,390	11.21
8	Share far above average	1,550	12.50
9	highest share	1,790	14.43
99	only commercially used	36	0.29

Source: microm + pairfam anchor W1; N=12,402.

Share of immigrants (*ha_mso_k_ausland*) is based on an analysis of residents' first and second names conducted by microm. The variable gives the expected share of household heads with migration background in 9 categories ranging from 1=very low to 9=very high. On the aggregate levels for each of the nine categories metric variables give the number and percentage of households falling in the respective category, e.g. the metric variable *mz_mso_p_ausland_2* gives the share of households with a very low share of immigrants.

Table 7: Share of immigrants (*ha_mso_k_ausland*)

		Frequency	Percent
.	Incomplete data	969	7.81
1	lowest share	1,247	10.05
2	extremely low share	1,307	10.54
3	very low share	1,325	10.68
4	share far below average	1,352	10.90
5	share below average	1,316	10.61
6	share slightly below average	1,115	8.99
7	average share	1,165	9.39
8	higher-than-average share	1,263	10.18
9	highest share	1,308	10.55
99	only commercially used	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

Dominant family structure (*ha_mso_k_familie*) indicates the share of singles versus families in 9 equally sized categories (1=predominantly single households; 9=highest share of families with children). The information is based on information on household size and number of children. On the aggregate levels for each of the nine categories metric variables give the number and percentage of households falling in the respective category, e.g. the variable *mz_mso_p_familie_1* indicates the share of households in the district falling in the category "predominantly one-person households".

Table 8: Dominant family structure (*ha_mso_k_familie*)

		Frequency	Percent
.	Incomplete data	969	7.81
1	Predominantly one-person households	758	6.11
2	Share of one-person households far above average	910	7.34
3	Share of one-person households above average	1,127	9.09
4	Share of one-person households slightly above average	1,185	9.55
5	Mixed family structure	1,218	9.82
6	Share of families with children slightly above average	1,337	10.78
7	Share of families with children above average	1,499	12.09
8	Share of families with children far above average	1,597	12.88
9	Predominantly families with children	1,767	14.25
99	Only commercial use	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

Status (*ha_mso_k_status*) is a variable classifying households in 9 approximately equally sized groups according to their education and income (1=very low status; 9=very high status). Information is drawn from occupations and academic titles of telephone subscribers and data from a large private database (administered by the "Verband der Vereine Creditreform"). On the aggregate levels for each of the nine categories metric variables give the number and percentage of households falling in the respective category, e.g. on the district level the metric variable *mz_mso_p_status_2* gives the share of households with a very low status.

Table 9: Status (*ha_mso_k_status*)

	Status category	Frequency	Percent
.	Incomplete data	969	7.81
1	Lowest status	1,314	10.60
2	Very low status	1,273	10.26
3	Status far below average	1,256	10.13
4	Status below average	1,344	10.84
5	Status slightly below average	1,262	10.18
6	Average status	1,320	10.64
7	Status slightly above average	1,279	10.31
8	Status above average	1,294	10.43
9	Highest status	1,056	8.51
99	Only commercial use	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

Need for anonymity (*ha_mso_k_anonym*) is available only for waves 1-3 on house level. It is measured by an index of disclosed and withheld information about profession, full first names, and full postal addresses. Data is provided in nine categories.

Table 10: Need for anonymity *(ha_mso_k_anonym)*

		Frequency	Percent
.	Incomplete data	969	7.81
1	Lowest need for anonymity	1,392	11.22
2	Very low need for anonymity	1,399	11.28
3	Need for anonymity far below average	1,461	11.78
4	Need for anonymity below average	1,407	11.34
5	Need for anonymity slightly below average	1,271	10.25
6	Average need for anonymity	1,217	9.81
7	Need for anonymity slightly above average	1,198	9.66
8	Need for anonymity above average	1,113	8.97
9	Highest need for anonymity	940	7.58
99	Only commercial use	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

4.1.2. Building data

Type of building (*ha_mbe_k_haustyp*) gives information on the type of structure with respect to the number of housing units and companies. On the aggregate levels for each of the seven categories metric variable give the number and percentage of households falling in the respective category, e.g. the variable *P8_MBE_P_Haustyp_6* gives the share of households in the district located in high-risers (with ≥ 20 households).

Table 11: Type of building (*ha_mbe_k_haustyp*)

	Type of building	Frequency	Percent
.	Incomplete data	969	7.81
1	One/two family house in homogenously built street section	2,503	20.18
2	One/two family house in in-homogenously built street section	3,151	25.41
3	Multi-family house (3-5 households)	2,146	17.30
4	Multi-family house (6-9 households)	1,717	13.84
5	Housing block (10-19 households)	1,282	10.34
6	High-rise building (≥ 20 households)	445	3.59
7	Predominantly commercially used	154	1.24
99	Only commercially used	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

Type of Street (*ha_mbe_k_strtyp*) describes the share of residential and commercial use of the street section where the microcell is located. Commercial use of the street is determined on the basis of the number of jobs, freelancers, stores, and restaurants as well as the number of companies from objectionable industries. On the aggregate levels metric variable give the number and percentage of households falling in the respective category, e.g. the variable *st_mbe_p_strtyp_1* gives the share of households in the street characterized as located in a residential street.

Table 12: Type of street (*ha_mbe_k_strtyp*)

	Type of street	Frequency	Percent
.	Incomplete data	969	7.81
1	Residential street	6,149	49.58
2	Street with stores and services	1,396	11.26
3	Mixed type street	2,217	17.88
4	Business street	1,425	11.49
5	Street with predominantly commercial use	211	1.70
99	Only commercially used street	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

4.1.3. Mobility data

Volume of relocations (*ha_mmo_k_volumen*) refers to the number of relocations in relation to 1,000 households. This ratio is calculated directly for larger geographic areas such as communities or postal code districts. For the house level, statistical probabilities have been calculated by microm and projected from a higher level (with on average 26 households) to the microcells. No information is available on the street level. On the district and municipality level, the information is given in metric variables measuring the volume of relocations (e.g. *p8_mmo_k_volumen)*.

Table 13: Volume of relocations (*ha_mmo_k_volumen*)

	Volume of relocations	Frequency	Percent
.	Incomplete data	969	7,81
1	Lowest volume of relocations	1,184	9.55
2	Very low volume of relocations	1,414	11.40
3	volume of relocations far below average	1,275	10.28
4	volume of relocations below average	1,441	11.62
5	volume of relocations slightly below average	1,313	10.59
6	Average volume of relocations	1,288	10.39
7	volume of relocations slightly above average	1,171	9.44
8	volume of relocations above average	1,243	10.02
9	Highest volume of relocations	1,069	8.62
99	Only commercial use	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

Balance of relocations (*ha_mmo_k_saldo*) is the difference of population influx and outflux per 1,000 households. This ratio is calculated directly for larger geographic areas such communities or postal code districts. For the house level, statistical probabilities have been calculated by microm and projected from the higher levels to the microcells. No information is available on the street level. On the district and municipality level, the information is given in metric variables measuring the balance of relocations (e.g. *p8_mmo_k_saldo*).

Table 14: Balance of relocations (*ha_mmo_k_saldo*)

	Balance of relocations	Frequency	Percent
.	Incomplete data	969	7.81
1	Balance of relocations very negative	1,119	9.02
2	Balance of relocations negative	1,207	9.73
3	Balance of relocations slightly negative	1,315	10.60
4	Balance of relocations even	1,265	10.20
5	Balance of relocations slightly positive	1,340	10.80
6	Balance of relocations positive	1,292	10.42
7	Balance of relocations very positive	1,314	10.60
8	Balance of relocations extremely positive	1,325	10.68
9	Balance of relocations most positive	1,221	9.85
99	Only commercial use	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

Population fluctuation (*ha_mmo_k_fluktu*) integrates the volume and balance of relocations. Lowest fluctuation arises in the case of a small number of relocations and a positive balance of relocations, whereas highest fluctuation corresponds to a large number of relocations with a negative balance. This ratio is calculated directly for larger geographic areas such as communities or postal code districts. On the district and municipality level, the information is given in metric variables measuring the fluctuation (e.g. *p8_mmo_k_fluktu*). For the house level, statistical probabilities have been calculated by microm and projected from the higher levels to the microcells. No information is available on the street level.

Table 15: Population fluctuation (*ha_mmo_k_fluktu*)

	Fluctuation	Frequency	Percent
.	Incomplete data	969	7.81
1	Lowest fluctuation	1,297	10.46
2	Very low fluctuation	1,293	10.43
3	fluctuation far below average	1,347	10.86
4	fluctuation below average	1,362	10.98
5	fluctuation slightly below average	1,312	10.58
6	Average fluctuation	1,223	9.86
7	fluctuation slightly above average	1,250	10.08
8	fluctuation above average	1,247	10.05
9	Highest fluctuation	1,067	8.60
99	Only commercial use	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

Short and long distance relocations (*mz_mmo_k_nahquote / mz_mmo_k_fernvol*) are available for waves 1-4. The share of short distance relocations is calculated by the share of relocations over a distance of less than 5 kilometers on the total number of relocations. This variable can be used as a measure of attractiveness of a district. Long distance relocations refer to the number of relocations over a distance of more than 10 kilometers. Note that short distance relocations are given as a rate,

whereas long distance relocations are given as the total volume. These variables are available also on the district levels (PLZ8 and market cells), on the postal code level, and on the municipality level. No information is available on the street level.

Table 16: Short/long distance relocations

		Short distance relocations (*mz_mmo_k_nahquote*)		Long distance relocations (*mz_mmo_k_fernvol)*	
		Frequency	Percent	Frequency	Percent
.	Incomplete data	105	0.85	105	0.85
1	Lowest	1,537	12.39	1,208	9.74
2	Far below average	1,691	13.63	1,306	10.53
3	Below average	1,317	10.62	1,455	11.73
4	Slightly below average	1,443	11.64	1,489	12.01
5	Average rate	1,307	10.54	1,452	11.71
6	Slightly above average	1,401	11.30	1,293	10.43
7	Above average	1,274	10.27	1,421	11.46
8	Far above average	1,195	9.64	1,324	10.68
9	Highest	1,131	9.12	1,348	10.87
99	Only commercial use	1	0.01	1	0.01

Source: microm + pairfam anchor W1; N=12,402.

4.1.4. Life stages

This classification combines age and family structure, yielding the stages listed in the following table. In addition, for each of the groups in the table, three subgroups are built according to financial status (financially weak, financially sound, and financially strong).

Table 17: Age groups, types of family structure, and life stages

		Age groups			
		≤ 35 years	36-55 years	56-56 years	> 65 years
Family structure	Singles	Young singles	Singles		Unmated elderly people
	Couples	Young couples	Couples	Older couples	
	Families / Households with children	Young families with children	Families with children	Older multi-person households	

The dataset includes a categorical variable indicating the dominant life stage (*ha_mlp_k_lebphase*) and one indicating the dominant life stage combined with financial status (*ha_mlp_k_statuslp*), as well as a set of metric variables giving the probability for each life stage that a household of this life stage is in the respective building (*ha_mlp_p_jusingle, ha_mlp_p_jupaare*, etc.). microm does not provide information about how the variables were measured or constructed on house level. On the street and district levels (but not on the municipality level), the variables indicating the dominant life stage and the dominant life stage combined with financial status is given (*p8_mlp_k_lebphase* and

p8_mlp_k_statuslp) are available. In addition, for each of the life stages, variables on the aggregate levels (including municipality level) indicate the share of households characterized by the respective life stage.

Table 18: Life stages (*ha_mlp_k_lebphase*)

	Life stage	Frequency	Percent
.	Incomplete data	969	7.81
1	Young singles	672	5.42
2	Young couples	429	3.46
3	Young families with children	531	4.28
4	Singles	1,357	10.94
5	Couples	1,370	11.05
6	Families with children	4,349	35.07
7	Unmated elderly people	1,205	9.72
8	Older couples	1,128	9.10
9	Older multi-person households	357	2.88
99	Only commercial use	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

4.1.5. microm types and groups

These typologies resulted from a cluster analysis where information such as resident age and income, type and ownership of buildings is used. The 39 microm types (e.g. new high standard detached houses, not modernized old buildings) are combined into 11 groups (e.g. urban trouble spots, good addresses in medium-sized cities). Microm types are stored in the variable *ha_mty_k_mtyp*, microm groups in the variable *ha_mty_k_mgruppe*. Table 3 gives the distribution of microm groups in wave 1. For a more detailed description of the types, see microm (2014). On the aggregate levels metric variable give the number and percentage of households falling in each of the categories, e.g. the variable *st_mty_p_mgruppe_i* gives the share of households in the street characterized as rural population. Information is available both for microm types and groups.

Table 19: microm groups (*ha_mty_k_mgruppe*)

	Group	Original (German) group name	Frequency	Percent
.	Incomplete data	Unvollständige Information	969	7.81
1	High status city slickers	Statushohe Großstädter	433	3.49
2	Well-to-do citizens in suburban areas	Gutsituierte in stadtnahen Umlandgemeinden	1,329	10.72
3	Nice residential areas in medium-sized cities	Gute Wohngebiete in mittelgroßen Städten	1,161	9.36
4	Urban trouble spots	Städtische Problemgebiete	1,559	12.57
5	High-rises and basic rented apartments	Hochhäuser und einfache Mietwohnungen	1,456	11.74
6	Senior citizens in basic postwar buildings	Rentner in einfachen Nachkriegsbauten	784	6.32
7	Old buildings in rural areas	Alte Häuser auf dem Land	1,179	9.51
8	Blue-collar workers in small towns	Arbeiter in kleinen Städten	1,680	13.55
9	Senior citizens in suburban areas	Ältere Leute in Umlandgemeinden	1,327	10.70
10	Rural population	Landbevölkerung	490	3.95
99	Exclusively commercial use	rein gewerbliche Nutzung	35	0.28

Source: microm + pairfam anchor W1; N=12,402.

4.1.6. microm Geo Milieus®

microm Geo Milieus are a micro-geographic adaptation of the Sinus Milieus® (developed by the Sinus-Institute). Sinus Milieus describe groups of people with similar life-styles and views of life, including, for instance, value orientations as well as attitudes towards work, family, or – as the classification stems from market research – consumption. The Sinus model, which was developed using qualitative research methods, was revised in 2010 and now includes 10 milieus. By combining the existing microm data with data from research on Sinus Milieus, the microm Geo Milieus were created. The dataset includes a metric variable for each of the 10 milieus indicating the probability that a certain Sinus Milieu exists in the house (*ha_mgm_p_ket, ha_mgm_p_lib,* etc.) as well as a categorical variable (*ha_mgm_k_dom*) to identify the dominant milieu in the house. Thereby, the dominant milieu is the one that is most overrepresented in the house rather than simply the milieu with the largest share. Note that for the first two waves the old typology is available (*ha_mmi_s_etb* etc.). The variables are available on all aggregate levels.

Table 20: microm Geo Milieus (*ha_mgm_k_dom*)

	Milieu	Description	Dominant milieu	
			Frequency	Percent
.	Incomplete data		446	5.30
1	Established conservative	Responsibility and success ethics, leadership aspirations, class consciousness	936	11.13
2	Liberal intellectual	Liberal attitude, desire for self-determination, intellectual interests	633	7.52
3	High achiever	Efficiency-orientation, Avant-garde consumption styles, high multimedia competence	614	7.30
4	Movers and shakers	High geographic mobility, intense networking, creativity, looking for new experiences and solutions	442	5.25
5	New middle class	Acceptance of social order, desire for security and establishment in occupation and society	1046	12.43
6	Adaptive pragmatist	Success oriented, pragmatic, hedonistic and conventional attitudes, desire for social affiliation	743	8.83
7	Socio-ecological	High ecological and social conscience, anti-consumerism, and political correctness	565	6.72
8	Traditional	Desire for security and order, thriftiness, conformism	954	11.34
9	Precarious	Social disadvantages, few opportunities for advancement, fear for the future, resentments	834	9.91
10	Escapist	Importance of fun and experience, rejection of conventions and norms of the performance-oriented society	1,183	14.06
99	Only commercial use		16	0.19

Source: microm Data handbook, 2014; microm data + pairfam anchor W3; N=8,412.

4.2. Street level data

4.2.1. Rented & owned dwellings

Wave 1 includes data on households who rent or own their dwelling, respectively. This information is given as the number of households (*st_mwo_a_eigentum* / *st_mwo_a_mieter*) as well as the share of households on the total number of households in the street section (*st_mwo_p_eigentum* / *st_mwo_p_mieter*). The variables are available only on the street level.

Table 21: households with rented/owned dwellings (*st_mwo_a/p_eigentum/mieter*)

	Rented dwellings	Owned dwellings
Incomplete data (.)	98	98
Number of households		
Mean	54.63	26.67
Median	27	20
Range	0-855	0-285
Share of total number of households		
Mean	54.71	44.83
Median	54.00	45.71
Range	0-100	0-100

Source: microm + pairfam anchor W1; N=12,402.

4.3. District level data

Two district levels are included in the dataset because of a change in measurement units. Market cells are available for the first four waves whereas PLZ8-areas are available for waves 1-5. For many variables in waves 1-4 both levels are included.

4.3.1. District size

The variable *p8_p8t_w_flaeche* gives information as to the size of the district in square meters. Average district size in the fifth wave of the pairfam panel is 4,259,519 m², with a range from 5,531.48 m² in high-density districts to 60,796,800 m² (60.8 km²) in rural areas. The variable is available only for wave 5.

The variable *mz_ewa_a_gesamt* gives the number of residents in the district. On average, in the wave 1 data, districts embrace 1,338 residents (median: 1,275), with a range from 0 to 6939 residents. The variable is available for PLZ8 regions and market cells.

4.3.2. Population density

Wave 1 data contain an additional variable (*mz_reg_w_ewdichte*) measuring population density (with reference to the covered area in the district). Mean population density is 0.005 residents per m².

4.3.3. Age structure

Data includes the number of residents separately for both genders and 17 age groups (*p8_eag_a_m00bis03, …, p8_eag_a_m75undgr; p8_eag_a_w00bis03, …, p8_eag_a_w75undgr*). In the following table, mean values of each gender and age group are listed. Data is only available for PLZ8 regions in wave 5.

Table 22: Number of inhabitants per age group (*p8_eag_a_**)

	Women			Men		
	Number of persons		Mean share of total (%)	Number of persons		Mean share of total (%)
	Mean	Range		Mean	Range	
0-3 years	16.60	0-61	1,25	17.25	0-62	1,30
3-6 years	16.54	0-63	1,25	17.33	0-66	1,31
6-10 years	23.03	0-87	1,76	24.25	0-91	1,83
10-15 years	31.26	1-143	2,36	32.94	1-138	2,49
15-18 years	18.72	0-82	1,41	19.74	1-105	1,49
19-20 years	14.02	0-64	1,07	14.63	0-67	1,11
20-25 years	40.16	0-235	3,03	42.06	1-176	3,18
25-30 years	39.42	0-190	2,98	41.02	0-186	3,10
30-35 years	38.67	0-157	2,92	39.98	0-164	3,02
35-40 years	39.82	1-138	3,01	40.98	1-150	3,10
40-45 years	52.82	1-211	3,99	55.41	1-201	4,19
45-50 years	56.80	1-248	4,29	59.14	1-238	4,47
50-55 years	50.28	1-167	3,80	50.84	0-192	3,84
55-60 years	44.99	0-179	3,40	43.81	0-171	3,31
60-65 years	38.00	1-121	2,87	36.37	1-127	2,75
65-75 years	78.62	1-317	5,94	69.78	2-280	5,27
75 years and older	74.52	2-293	5,63	44.06	0-151	3,33
Total (men + women)	1,323.86	25-4,007				
Incomplete data (.)	129					

Source: microm + pairfam anchor W5; N=7,201.

4.3.4. Ethnicity

For wave 1, the shares of residents of different ethnicities within the total population of the district (market cell) are given. This information is gained by analyses of first and second names regarding their linguistic origin (using international indexes of names and lists of linguistic origin of names).

Table 23: Ethnicities

Share of residents with names from…	Mean	S.D.	Range	Incomplete data (.)
Germany (*mz_met_p_deutschl*)	93.44	6.21	34.63-100	111
Italy (*mz_met_p_italien*)	.71	.81	0-13.57	111
Turkey (*mz_met_p_tuerkei*)	2.17	3.98	0-62.05	111
Greece (*mz_met_p_griechen*)	.48	.60	0-9.55	111
Spain, Portugal, South America (*mz_met_p_spanport*)	.21	.32	0-12.87	111
Balkan states (*mz_met_p_balkan*)	.73	1.02	0-13.21	111
Eastern Europe (*mz_met_p_osteurop*)	.66	.75	0-9.55	111
Sub-Saharan Africa (*mz_met_p_afrika*)	.11	.22	0-3.31	111
Non-European Islamic countries (mz_met_p_islam)	.19	.37	0-7.94	111
South/East/Southeast Asia (*mz_met_p_asien*)	.08	.18	0-3.44	111
Other (*mz_met_p_uebrige*)	.94	.61	0-16.74	111
Late repatriates (*mz_met_p_spaetaus*)	.24	.38	0-4.49	111

Source: microm + pairfam anchor W1; N=12,402.

4.3.5. Unemployment rates

Unemployment rates are provided by the German Federal Employment Agency (Bundesagentur für Arbeit). The dataset contains the unemployment rate in the district (*p8_alq_p_quote*), an index of the district level unemployment quotes in relation to the general German quote (*p8_alq_i_quotebrd*), and an index of district level unemployment in relation to West and East Germany, respectively (*p8_alq_i_quotewo*). Districts in West (East) Germany are set in relation to the West (East) German unemployment rates, West (East) Berlin is treated as West (East) Germany. In addition, a categorical variable (*p8_alq_k_quote*) is included, ranging from 1 (lowest unemployment rate) to 7 (highest unemployment rate).

Table 24: Unemployment categories (*p8_alq_k_quote*)

	Unemployment categories	Range of unemployment rates	Frequency	Percent
.	Incomplete data		3,726	30.04
1	Lowest unemployment rate	0 - 2.89	2,064	16.64
2	Unemployment rate below average	2.90 - 4.70	1,562	12.59
3	Unemployment rate slightly below average	4.71 - 6.52	1,373	11.07
4	Unemployment rate about average	6.53 - 8.32	814	6.56
5	Unemployment rate slightly higher than average	8.33 - 10.87	971	7.83
6	Unemployment rate higher than average	10.87 - 14.49	955	7.70
7	Highest unemployment rate	14.5 - 27.82	937	7.56

Source: microm + pairfam anchor W1; N=12,402.

4.3.6. District type

Districts are classified according to their location and function. The 15 types are categorized into 7 groups. For the classification, variables such as type of building, population density, and industries of the companies located in the district were used. The dataset contains a variable indicating the type of district (*p8_p8t_k_p8typ*) where the microcell is located, as well as the group to which the type of district belongs (*p8_p8t_k_p8gruppe*). This data is available only for wave 5.

Table 25: District types and groups

Group	Type of district	Original (German) name	Freq.	Percent
Urban centers	City centers	Stadtzentren	32	0.44
	District centers / town centers	Neben- und kleine Zentren	100	1.39
	Shopping, culture, leisure areas	Einkaufs-, Kultur- und Freizeitbereiche	120	1.67
	Small centers	Kleinstzentren	373	5.18
Densely populated areas	Inner-city high-density habitation	Innerstädtisches, hochverdichtetes Wohnen	370	5.14
	Densely populated metropolitan housing areas	Großstädtisch geprägte verdichtete Wohngebiete	422	5.86
	Densely populated small-town housing areas	Kleinstädtisch geprägte verdichtete Wohngebiete	644	8.94
Habitation in urban peripheries	Peripheral metropolitan housing areas	Großstädtisch geprägte Wohngebiete in Randlagen	652	9.05
	Habitation in outer conurbation areas	Wohnen im städtischen Umland	832	11.55
	Peripheral small-town housing areas	Kleinstädtische Wohngebiete in Randlage	687	9.54
Rural centers	Small centers in rural areas	Kleine Zentren im ländlichen Raum	645	8.96
Holiday areas	Holiday and leisure areas	Ferien- und Freizeitgebiete	94	1.31
Rural areas	Habitation in rural areas	Wohnen im ländlichen Raum	1,769	24.57
	Agricultural areas	Landwirtschaftliche geprägte Gebiete	232	3.22
Commercial areas	Commercial areas	Gewerblich geprägte Gebiete	229	3.18

Source: microm + pairfam anchor W5; N=7,201.

4.3.7. Pedestrian area, cultural & leisure facilities

For wave 1 data on the existence of a pedestrian area in the district (*mz_reg_k_fgflag*), the length of the pedestrian area (*mz_reg_w_fglgm*), and the share of the pedestrian area in relation to all streets of the district (*mz_reg_p_fgstr*) is given. 5.6 percent of the districts in the dataset have a pedestrian area (N=695). Table 26 includes only those districts with a pedestrian area.

Table 26: Pedestrian area (*mz_reg_w_fglgm* / *mz_reg_p_fgstr*)

	Mean	S.D.	Range
Length of pedestrian area	330.79	369.34	7-3,147
Share of pedestrian area on district streets	.11	.15	.002-1

Source: microm + pairfam anchor W1; N=695.

In addition, wave 1 data includes the share of cultural and leisure facilities of all households and companies in the district and in the municipality (*mz_reg_p_kultur, gk_reg_p_kultur*).

Table 27: Cultural & leisure facilities (*mz_reg_p_kultur, gk_reg_p_kultur*)

	Mean	S.D.	Range	Incomplete data (.)
Share of cultural and leisure facilities: district	.0001	.0005	0-.345	116
Share of cultural and leisure facilities: municipality	.090	.149	0-1.8	1,780

Source: microm + pairfam anchor W1; N=12,286.

4.4. Community level data

4.4.1. Religious denominations

Data covers the two main Christian denominations, Roman Catholic and Protestant. Other Christian denominations, other religions, and religionless inhabitants are classified as "other". The data include the number of inhabitants in the community where the house is located (*gk_kon_a_gesamt*), the number of inhabitants of each of the three groups (*gk_kon_a_evangel*, *gk_kon_a_roemkath*, *gk_kon_a_sonstige*, for Roman Catholic, Protestant, and Other, respectively), as well the percentage of the groups to the total number of inhabitants (*gk_kon_p_evangel*, *gk_kon_p_roemkath*, *gk_kon_p_sonstige*). In addition, for waves 1 and 5 an index value is included indicating the share in the community in relation to the German mean (gk_kon_i_evangel, *gk_kon_i_roemkath*, *gk_kon_i_sonstige*). The variables are based on census data and are updated using data from the Federal Statistical Office and the German Catholic and Protestant churches.

Table 28: Religious groups

	Mean	S.D.	Incomplete data (.)
Protestant (*gk_kon_p_evangel*)	31.71%	17.63	4,334
Roman Catholic (*gk_kon_p_roemkath*)	30.26%	25.01	4,334
Other (*gk_kon_p_sonstige*)	38.03%	20.57	4,334

Source: microm + pairfam anchor W1; N=12,402.

5. References

Brüderl, J., Hank, K., Huinink, J., Nauck, B., Neyer, F. J., Walper, S., . . . Wilhelm, B. (2015). The German Family Panel (pairfam): ZA5678 Data file Version 6.0.0. *GESIS Data Archive, Cologne.* doi:10.4232/pairfam.5678.6.0.0.

Brüderl, J., Hajek, K., Herzig, M., Huyer-May, B., Lenke, R., Müller, B., . . . Schumann, N. (2015). *pairfam Data Manual: Release 6.0*. Munich.

Brüderl, J., Schmiedeberg, C., Castiglioni, L., Arránz Becker, O., Buhr, P., Fuß, D., . . . Schumann, N. (2015). *The German Family Panel: Study Design and Cumulated Field Report (Waves 1 to 6): Release 6.0* pairfam Technical Paper No. 01. Munich.

Hensel, T., Kreyenfeld, M., & Walke, R. (2015). *Guidelines for linking contextual factors and survey data: an application with data from the German Family Panel (pairfam)* MPIDR Technical Report No. TR-2015-005. Rostock.

Hintze, P., & Lakes, T. (2009). *Geographically Referenced Data in Social Science. A Service Paper for SOEP Data Users.* SOEP Data Documentation No. 46. Berlin.

Huinink, J., Brüderl, J., Nauck, B., Walper, S., Castiglioni, L., & Feldhaus, M. (2011). Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. *Zeitschrift für Familienforschung - Journal of Family Research, 23,* 77–101.

Logan, J. R. (2012). Making a Place for Space: Spatial Thinking in Social Science. *Annual Review of Sociology, 38,* 507–524.

microm Consumer Marketing. (2010). *microm Datenhandbuch: Arbeitsunterlagen für microm MARKET & GEO*. Neuss.

microm Consumer Marketing. (2015). *microm Datenhandbuch 2015*. Neuss.

Statistisches Bundesamt. (2014). *Daten aus dem Gemeindeverzeichnis: Verwaltungsgliederung in Deutschland am 31. 12. 2012*. Aktualisiert auf Zensusdaten mit dem Stand vom 10.04.2014 im April 2014. Wiesbaden.