The German Socio-Economic Panel Study

Information on the SOEP *County-Level Data*

September 18, 2006

C. Katharina Spieß and Annalena Dunkelberg

Summary

The SOEP county-level data offer users a dataset containing a household identifier for the SOEP household, the county ("*Kreis* / *kreisfreie Stadt*") in which the SOEP household lives, and further information at the county level. The dataset contains not only the original county code but also a recoded county code that facilitates longitudinal analysis at the county level. Further county-level information includes additional geographic information for the years 1985 to 2004 and selected socio-economic indicators for the years 1995 to 2002.

Information on the recoded county codes

The dataset contains the original county codes provided to the SOEP group by the survey institute *Infratest Sozialforschung*. These county codes were cleaned, that is, all codes identified as containing errors were corrected. The county codes are available from the year 1985 on.

Along with the original (cleaned) county code, the data set contains a recoded county code. The recoded county codes differ from the original county codes particularly for the years prior to 1997: for these years, the county codes were changed in order to enable the merging of external information longitudinally. In other words, the recoding renders the SOEP county codes compatible with county codes from external data

sources. From the year 1997 on, the original and the recoded county codes are virtually identical; all exceptions are listed in the following under points 2 and 3.

Specifically, the following changes were made:

1. Longitudinal analysis is particularly difficult in East Germany given the extensive redistricting at the county level carried out in this region[1]. Therefore it is impossible to integrate information from external sources (relating to different county boarders) since the county codes in question are thus incompatible. For this reason, the East German county codes were recoded to correspond with the current county codes. This is the case for the years prior to 1997.
2. The original SOEP county codes 11100 and 11200 (East and West Berlin) were recoded for the years 1990 to 2002 to county code 11000 (Berlin as a whole).
3. County codes 3253 and 3201 (Hannover region and Hannover city) were recoded to county code 3241 (Hannover as a whole).

The recoded county codes should be used when the user plans to merge external information with the SOEP data. We would like to point out in particular that the recoded county codes do not relate to the geographical delimitations of the particular counties or regions valid at that particular point in time but rather to the *current* definition of the area. We cannot guarantee that all changes in counties, regions or districts, or changes in the county codes, have been registered correctly. Inaccuracies in the use of recoded county codes must thus be dealt with through the content and definition of the research question.

The data set also contains a variable with the county name assigned on the basis of the information from “Statistik regional“[2] and additional information from the German Federal Office for Building and Regional Planning.

---

1 Cf. e.g. *Blach, Antonia und Jacek Jonetzko* (1999): Die Gebietsreform der neuen Länder: Folgen für die Laufende Raumbeobachtung des BBR, Working Paper 5/1999 of the Federal Office for Building and Regional Planning, Bonn.

2 Cf. Statistik regional, Daten und Informationen der Statistischen Ämter des Bundes und der Länder, edition 1998 to 2004 (on CD-Rom).

Information on the data from outside sources

Geographical information

Additional geographical information on the SOEP households from outside sources was merged with SOEP data via the recoded county code. This information includes the longitude and latitude of the county in which the SOEP household lives. The longitude and latitude of the county was, with only a few exceptions, assigned based on the county seat of the particular county in question. The variable *$ort* is the location at which the coordinates were measured. The longitude and latitude of the county seat was assigned using the Microsoft program package "*MS Autoroute*"[3].

Based on the longitudinal and latitudinal data *(variables: $laeng, $breit*), the distances between two counties can be calculated. The distance between county A with the latitude $breit_A$ and longitude $laeng_A$ and county B with the coordinates $breit_B$ and $laeng_B$ is computed using the following formula, where 6378 km² corresponds to the circumference of the earth.

$$D = 6378km^2 * \cos^{-1}(\sin(breit_A) * \sin(breit_B) + \cos(breit_A) * \cos(breit_B) * \cos(laeng_A - laeng_B)$$

Using the statistical program STATA, for example, the distance can be calculated using the following syntax[4]:

```
display 6378*acos(sin(breitA*_pi/180)*sin(breitB*_pi/180)+ cos(breitA
*_pi/180)*cos(breitB*_pi/180)* cos((laengA-laengB)*_pi/180))
```

Here the coordinates for $breit_A$, $laeng_A$, $breit_B$ and $laeng_B$ have to be inserted. Using the STATA package "circstat", it is possible to make many calculations for circular data such as coordinates.

Socio-economic information

Furthermore, using the recoded county codes, socio-economic information on the county level can be merged to the SOEP households. The selected indicators are intended as a first step in providing interested SOEP users with some central

[3] Cf. Microsoft AutoRoute 2005 (CD-Rom Ausgabe).

information on the counties. The indicators include data on the total surface area, population density, income, number of employed persons making compulsory social insurance contributions, and unemployment.

The total surface area (*$flaeche*) measures the geographical area in square kilometers, and was registered in the 12th month of the year in question. The variables *$bev_g, $bev_dt* and *$bev_al* measure the total population, the German and foreign population, in each case also measured in the 12th month of the year in question. The average income of the private households per year, including non-business-related private organizations measured in euros per resident, contains the variable *$eink*. The variable *$sozvs* contains the number of employed persons making compulsory social insurance contributions in the county. This information is for the sixth month of the particular year. The variables *$alq_1* to *$alq_4* reflect the unemployment rates for the particular quarter measured as a percentage. All this information is available from "*Statistik regional*", mainly for the years from 1995 on, and can be integrated into the SOEP dataset on this basis[5].

Information on the variables in the data set

The following table lists the variables contained in the dataset "Kreise.dta" as well as each variable label and the data source. The data set itself is provided in STATA format. "$" is the wave-specific prefix (waves A to U). Missing data from the outside sources (e.g. "Statistik regional") are set at "-2".

---

4 The result is the distance, measured in kilometers.

5 For the East German counties, this information was merged for the years 1995 and 1996 based on the recoded county codes, which are not necessarily completely identical with the counties as defined in the years 1995 and 1996.

Variable name	Variable label	Source
Hhnr	Original household number	SOEP original
hhnrakt	Current household number	SOEP original
$hhnr	Wave-specific household number	SOEP original
$kkz	Original SOEP county code	SOEP original
$kkz_rek	Recoded county code	SOEP change
$kreis	County name	"Statistik regional"
$ort	County of coordinate measurement	SOEP addition
$breit	Latitude	MS Autoroute
$laeng	Longitude	MS Autoroute
$flaeche	Surface area in square kilometers	"Statistik regional"
$bev_g	Total population	"Statistik regional"
$bev_dt	German population	"Statistik regional"
$bev_al	Foreign population	"Statistik regional"
$eink	Income of private households, euros/resident	"Statistik regional"
$sozvs	Employees paying compulsory social insurance	"Statistik regional"
$alq_1	Unemployment rate, 1st quarter	"Statistik regional"
$alq_2	Unemployment rate, 2nd quarter	"Statistik regional"
$alq_3	Unemployment rate, 3rd quarter	"Statistik regional"
$alq_4	Unemployment rate, 4th quarter	"Statistik regional"

For data protection reasons, SOEP county-level data can be used only via *SOEPremote,* both at and outside DIW Berlin. For questions on the use of *SOEPremote,* please contact Jan Göbel, jgoebel@diw.de. For questions on the content of this dataset, contact Prof. Dr. C. Katharina Spieß, Tel.: 030/89789-254, e-mail: kspiess@diw.de.

We are grateful for suggestions on this documentation and on the use of the SOEP county-level data. Please direct your comments and ideas to: Prof. Dr. C. Katharina Spieß, email: kspiess@diw.de.

Selected studies carried out using SOEP county codes

Baumgartner, Hans J. (2004): Are There Any Class Size Effects On Early Career Earnings in West Germany ?, *DIW Discussion Papers* 417, Berlin.

Büchel, Felix und C. Katharina Spieß (2002): Müttererwerbstätigkeit und Kindertageseinrichtungen - neue Ergebnisse zu einem bekannten Zusammenhang, in: *Vierteljahrshefte zur Wirtschaftsforschung*, 2002 (71), 96-114).

Hank, Karsten (2003): The Differential Influence of Women's Residential District on the Risk of Entering First Marriage and Motherhood in Western Germany, in: *Population and Environment*, 25 (1), 3-21.

Hank, Karsten und Michaela Kreyenfeld (2003): A Multilevel Analysis of Child Care and Women 's Fertility Decisions in Western Germany, in: *Journal of Marriage and Family,* 65 (3), 584-596.

Hank, Karsten, Michaela Kreyenfeld und C. Katharina Spieß (2004): Kinderbetreuung und Fertilität in Deutschland, in: *Zeitschrift für Soziologie,* 33 (3), 228-244.

Hunt, Jenny (2004): Are Migrants More Skilled than Non-Migrants? Repeat, Return and Same-Employer Migrants, in: *Canadian Journal of Economics,* (37), 830-849).

Jürges, Hendrik (2005): The Geographic Mobility of Dual-Earner Couples: Does Gender Ideology Matter*?, DIW-Discussion Paper* 474.

Kreyenfeld, Michaela und Karsten Hank (2000): Does the availability of child care influence the employment of mothers? Findings from western Germany, in: *Population Research and Policy Review,* 19 (4), 317-337.

Rehdanz, K. and Maddison, D. (2005): Der Wert des Klimas für Haushalte in Deutschland, erscheint in: G. Grözinger/W. Matiaske (Hrsg.): *Deutschland regional: Sozialwissenschaftliche Daten im Forschungsverbund*.

Rehdanz, Katrin. and David Maddison. (2004): *The Amenity Value of Climate to German Households*, FNU-39, Centre for Marine and Climate Research, Hamburg University, Hamburg. Auch erschienen als Nota di Lavoro 57.2004, Milano: Fondazione Eni Enrico Mattei, Milano und als DIW Discussion Paper 414, German Institute for Economic Research, Berlin.

Spieß, C. Katharina und Felix Büchel (2003): Effekte der regionalen Kindergarteninfrastruktur auf das Arbeitsangebot von Müttern, in: W. Schmähl (Hrsg.), *Soziale Sicherung und Arbeitsmarkt*, Berlin, S. 95-126.

Wrohlich, Katharina (2004): Child Care Costs and Mothers' Labor Supply: An Empirical Analysis for Germany, in: *DIW Discussion Papers* 412, Berlin.