

# Description of the variables: STRAT1, STRAT2 and SAMPOINT

Martin Spiess

Longitudinal Data and Microanalysis

DIW Berlin

Koenigin-Luise-Str. 5

14195 Berlin

Germany

November 13, 2001

## 1 General

The three variables STRAT1, STRAT2 and SAMPOINT are generated for the estimation of variances using program packages like STATA or SUDAAN, that use strata and cluster information at the first wave of each subsample. The variables STRAT1, STRAT2 and SAMPOINT can be found in the file VARIANZ. A further variable that may be of interest is the identifier of the interviewers (INTNR) also located in the file VARIANZ. Due to german laws, the values of all three variables are randomly assigned, however, under the restrictions described below.

## 2 Variable STRAT1

The variable STRAT1 identifies the strata from which the primary sample units (PSU's, similar to voting districts) were selected. For subsamples A,C,D,E and F these are given by regional strata, defined by Bundesland (federal state), Regierungsbezirk (administrative district) and Gemeindetyp (type of commu-

nity). Note that these units may change over time. For subsample B, strata are defined by the nationality of the head of the household.

The strata in variable STRAT1 are coded as follows: Each Bundesland received a number between 20 000 000 and 180 000 000, which is the same for every first wave of each subsample, each Regierungsbezirk received a number between 1 000 and 999 000, differing over the first waves, and each Gemeindetyp received a number between 1 and 999, again differing over the first waves. These values were added to give the values of the variable STRAT1. The exception is for households in subsample B, where a number between 1 and 5 codes different strata.

Although the strata are different over the first waves of the different subsamples, they may be geographically the same or largely overlapping areas. This is, however, reflected in the values of the variable STRAT1: similar values identify geographically similar strata.

### **3 Variable STRAT2**

This variable is missing for subamples A and C–F. Values are given for subsample B only. The values are generated in the same way as for subsamples A and C–F in STRAT1. However, these values do not reflect real strata used to select the sample, but merely identify geographically neighbouring PSU's (artificial strata).

### **4 Variable SAMPOINT**

This variable identifies the primary sampling units (PSU's) from which the households were selected. For subsample D there are no primary sample units, so the corresponding values of SAMPOINT are missing.