Combining the SOEP ongoing panel (subsamples A–E) with the new subsample F

Martin Spiess

Longitudinal Data and Microanalysis

DIW Berlin

Koenigin-Luise-Str. 5

14195 Berlin

Germany

November 13, 2001

1 Derivation of the weight

In this paper the derivation of the weight w to calculate estimates

$$\hat{\theta} = w\,\hat{\theta}_1 + (1 - w)\,\hat{\theta}_2, \qquad 0 \le w \le 1,$$

is described, where $\hat{\theta}_1$ is an estimator of θ, the population parameter, using the ongoing panel in 2000, $\hat{\theta}_2$ is an estimator of the same quantity using the first wave of the new sample (subsample F), and w is a weight to be constructed.

The theoretical description is given in Spiess and Rendtel (2000). They also describe the application of the theory to the combination of subsamples A–D and F.

The derivation of w follows those given Spiess and Rendtel (2000), except that it is assumed that the portion of variance due to the 'time effects' of the ongoing sample (A–E) compared to the portion of the variance due to 'extended design stage effects', i.e. group, design and nonresponse effects at the stage of 'drawing' the sample is negligible. For the derivation of the weight, note that both samples,

i.e. A–E and F are both considered as to be drawn in the same year, i.e. the year 2000 (c.f. Spiess and Rendtel, 2000). Therefore, we have

$$w_{y,\text{max}} \approx \frac{n_1}{n_2 + n_1}. \tag{1}$$

In 2000 in subsamples A–E und F there are 7030 and 6004 observed private households, respectively. Within these households there lived 13616 and 11048 adults (aged sixteen and older) and 3376 and 2773 children, respectively. This yields $w_H = .539$ (for households) and $w_P = .552$ (for persons) and, after rounding, $w_H = w_P = .55$ and $1 - w = .45$.

Reference

Spiess, M. & Rendtel, U. (2000). *Combining an ongoing panel with a new cross-sectional sample.* Deutsches Institut für Wirtschaftsforschung, Diskussionspapier Nr. 198.