Using Analysis of Gini (ANOGI) for Detecting Whether Two Subsamples Represent the Same Universe

The German Socio-Economic Panel Study (SOEP) Experience

JOACHIM R. FRICK
JAN GOEBEL
DIW Berlin

EDNA SCHECHTMAN
Ben Gurion University, Israel

GERT G. WAGNER
DIW Berlin

SHLOMO YITZHAKI
Hebrew University, Jerusalem

A wildly discussed shortcoming of panel surveys is a potential bias arising from selective attrition. Based on data of the German Socio-Economic Panel Study (SOEP), the authors analyze potential artifacts (level, structure, income inequality) by comparing results for two independently drawn panel subsamples started in 1984 and 2000. They apply ANOGI (analysis of Gini) techniques, the equivalent of ANOVA performed with the Gini coefficient. They rearrange, reinterpret, and use the decomposition in the comparison of subpopulations from which the different samples were drawn. Taking into account indicators for income, significant differences between these two samples with respect to income inequality are found in the first year, which start to fade away in Wave 2 and disappear in Wave 3. The authors find credible indication for these differences to be driven by changes in response behavior of short-term panel members rather than by attrition among members of the longer running sample.

Keywords: *panel studies; survey research; inequality decomposition; Gini*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 34, No. 4, May 2006 427-468
DOI: 10.1177/0049124105283109

1. INTRODUCTION

Most population surveys assert the claim that they are representative of the underlying population universe. While this is, in reality, already a very ambitious goal on its own, *panel* surveys that follow households, families, and individuals over time have to cope with the problem of adequately covering any *changes* occurring in the underlying population since the time of the original sampling. For example, immigration may not be represented within an ongoing panel survey if recent immigrants founded new households, which by definition did not have a positive sampling probability in the ongoing sample. Thus, new subsamples that complement a panel might be required. Furthermore, due to panel attrition and the need to control for eventual selectivity within this process, long-running panel surveys especially may require being complemented by additional subsamples that serve two functions: First, such "refreshment" samples help to stabilize the number of observations, and second, they provide a benchmark for the analysis of eventual selectivity due to panel attrition and changes in response behavior.

When analyzing economic well-being or income distribution issues on the basis of such micro-data, any undetected selectivity—for example, given by a "middle-income bias"—creates a bias in the estimates of income inequality measures.

The question to be answered in this article can be presented in the following general way: Given the existence of several independent samples (within one survey), do they represent the same population or universe? A common practice to answer this question is to look at the differences in various parameters among the populations with respect to the variable of interest. For example, one may wish to compare the moments of the distributions (means, variances, Ginis, medians, etc.). The problem with a methodology such as this one is that there are only a few moments that are usually being compared, and therefore, possible differences in other moments may not be detected. The methodology suggested in this article is based on a decomposition of a measure of total variability to the contributions of subpopulations. Since the interest is in the inequality in economic well-being (as measured, for example, by income), the comparison between the subpopulations is done by decomposing the Gini of income of the overall population into the contributions of the subpopulations to the overall inequality. The advantage of the decomposition is that it reveals a new parameter (called the overlapping index), which shows how intertwined the subpopulations are. Hence, unlike the comparison of moments that rely on each distribution separately, the overlapping index is based on the entwined observations of all the distributions involved.

Intuitively, the methodology presented below can be referred to as ANOGI (analysis of Gini)—the equivalent of analysis of variance (ANOVA), performed with the Gini coefficient. The decomposition we follow is the one presented in Yitzhaki (1994). We rearrange and reinterpret that decomposition in order to use it in the comparison of the subpopulations from which the different samples were drawn.

In this article, we are *mainly* interested in the effect (possible bias) of attrition. For each year within a three-year period (2000 to 2002), we have two subsamples. Both of them are subsamples from a chain of panel data, intended to represent the entire population of Germany. The major difference in the subsamples is that the new one was started in the year 2000, while the main part of the old one was started as early as 1984. For the period since the original sampling took place, both panel subsamples carried on using the same set of straightforward follow-up rules. We are interested in seeing whether the subsamples come from the same population (i.e., no effect of long-term attrition) or if attrition causes a bias.

As a central result, our analyses for the first year reveal a significant difference in income stratification and inequality between the subsamples, while in the second and third years, the two subsamples overlap almost perfectly (i.e., the relevant substantive results converge rather quickly). When discussing reasons for the differences in the first year, we find indications that these are not to be attributed to attrition in the "old" sample but rather to changes in the response behavior of the new subsample's members. In fact, there is some evidence that the answers of long-term panel respondents are of better quality than those of first-time respondents. In case of income data, this may reflect an improved record keeping of such households or a better knowledge and recall capacity resulting from the repeated interview experience.

We start the article with a detailed description of the underlying data and with a discussion of problems related to the representation of a population by means of different subsamples (Section 2). In

Section 3, we set up the ANOGI methodology, and Section 4 provides the estimators. In Section 5, we present results of inequality decomposition, and Section 6 concludes.

2. REPRESENTATION OF A POPULATION BY MEANS OF DIFFERENT SAMPLES

2.1. THE GERMAN SOCIO ECONOMIC PANEL STUDY (SOEP)

Established in 1984, the German Socio-Economic Panel Study (SOEP) is one of the main tools for social science and economic research for Germany, as well as for international comparisons (see Wagner, Burkhauser, and Behringer 1993; Haisken-DeNew and Frick 2003).[1]

In principle, the universe of the SOEP sample includes the entire resident population of Germany. As the SOEP started before the reunification of Germany occurred, the first subsamples of the SOEP in 1984 were only conducted in West Germany. We summarize the various subsamples in Table 1. A detailed description can be found in Appendix A.

In 2000, the starting year of the supplementary innovation Subsample F, all in all, 24,586 adult individuals participated in the SOEP survey, which covered 13,258 households and included 6,659 children younger than 16 years of age.

2.2. POTENTIAL ARTIFACTS CAUSED BY DIFFERENT SUBSAMPLES

The realization of different independent samples, which have the aim of representing the same population universe, can cause problems due to different, mostly fieldwork-related, reasons. First, it is possible that they belong to different universes because the sample frames were not by intention but in fact different with respect to specific sampling procedures. Second, initial response rates may differ across the various subsamples. Third, methodological problems occurring during the fieldwork can result in a couple of survey artifacts. All such specific problems could have occurred with the SOEP, especially because in the year 2000, the "old" Subsamples A through E were

TABLE 1: Description of the Socio-Economic Panel Study (SOEP) Subsamples

Sample	*Starting Year*	*Sample Size*[a] *(Number of Households in Starting Year)*	*Comments*
A	1984	4,528	"West German" sample
B	1984	1,393	Oversampling of foreigners
C	1990	2,179	"East German" sample
D	1995	522	Immigrants since 1984
E	1998	1,067	Supplementary sample
F	2000	6,052	Supplementary Innovation sample
Cross-section 2000		13,258 (6,052)	All samples A-F (thereof F)
Cross-section 2001		11,947 (4,911)	All samples A-F (thereof F)
Cross-section 2002		11,468 (4,586)	All samples A-F (thereof F)

a. Due to attrition and the follow-up of newly founded households in case of split-offs, the cross-sectional sample size in any year after the starting wave deviates from the initial sample size.

true panel subsamples with more than one wave and in fact different with respect to the numbers of waves.

Different Sampling Procedures

There are two structural problems of sampling households in Germany, as in many other countries: first, the representation of foreigners/immigrants and, second, households living in institutions. In the SOEP, the procedures of handling those subpopulations were changed over time: The sampling procedures of Subsamples A and B as well as F were slightly different for the first waves. Whereas in A and B, Germans and foreigners are surveyed by different methods, allowing for a theoretically proper representation of the five major immigrant groups (represented by Subsample B), this superior procedure was not possible for Subsample F. Subsamples C and D are samples with a special focus on East Germans and recent immigrants, respectively, while E is a small sample that was drawn by basically the same procedure as Subsample F. A more detailed description of these procedures is provided in Appendix B.

Response Rates

Given the massive confrontation with telephone surveys, ad hoc interviews by marketing companies, and so on, it is becoming increasingly problematic to motivate individuals to participate in population surveys. As such, the older subsamples in SOEP clearly show higher initial response rates (e.g., Subsamples A and B with 61 and 68 percent, respectively) than newly introduced subsamples such as Subsample F with only 52 percent (see Table 2).[2] This phenomenon also applies to longitudinal response rates over (two or) three waves; it does not matter whether one looks at the subsample-specific first two or three waves or at the exact same time period (i.e., calendar years; e.g., 2000 through 2002).

Surveying Artifacts

While in the year 2000, Subsample F is a "fresh" cross section, Subsamples A to E consist of panel samples of varying durations. Therefore, Subsamples A to E and F could represent different populations,

TABLE 2: Cross-Sectional and Longitudinal Response Rates in the Socio-Economic Panel Study (SOEP) by Subsample (in Percentages)

Initial (cross-sectional) Response Rate in Wave 1	
Sample A (1984)	61
Sample B (1984)	68
Sample C (1990)	70
Sample D (1994/95)	> 55
Sample E (1998)	54
Sample F (2000)	52
Longitudinal Response Rate (balanced panel as a percentage of starting wave's population)	
Over two years (calendar years 2000–2001 and 2001–2002)	
Samples A-E	89–93
Sample F	78–87
Over three years (calendar years 2000–2002)	
Samples A-E	80–86
Sample F	< 69
Over sample-specific first two waves (Waves 1–2)	
Samples A-E	81–92
Sample F	78
Over sample-specific first three waves (Waves 1–3)	
Samples A-E	72–88
Sample F	69

SOURCE: Authors' calculation from SOEP 2000–2002.

first, if it is not possible to correct for panel attrition and, second, if response behavior changes over time ("panel effects"). SOEP data providers control for Subsamples A to E in an appropriate manner by means of weighting (see Rendtel 1995; Rendtel, Wagner, and Frick 1995), but over the course of time, other panel effects certainly cannot be ruled out. First, conventional wisdom dictates that respondents can change their true behavior, which in turn produces some sort of bias in the result.

A second important effect may be provided by an increasing familiarization of the respondents with the survey instrument, mostly a questionnaire, which minimizes errors in the answers and improves quality of the collected information. As such, due to this "learning effect," panel data yield a more realistic measure than data from a single cross-sectional sample of the very same population. The second effect especially takes place in the field of income surveying. We know that in the course of time, the share of missing values on income variables (due to item nonresponse) declines and that there is a special reason for different "answering styles."

A third problem can be created by a mix of interview modes, which is necessary, at least in Germany, to ensure that respondents remain willing to participate over the long term.[3] Since the late 1990s, computer-assisted personal interviewing (CAPI) has been introduced gradually by means of a controlled experiment to complement the conventional paper-and-pencil questioning technique. CAPI was used for the first time in Subsample E in 1998, and after a successful testing phase, this survey method was also introduced in the existing SOEP Subsamples A through D in 2000. In Subsample F, this methodology was used from the very first wave, yielding an overall share of CAPI interviews of 28 percent in 2002.

Methodological research focusing on sustainable mode or technology effects[4] caused by the introduction of CAPI shows a mixed picture. While in the case of the British Household Panel Study (BHPS), no significant effects were found (see Laurie 2000), for the SOEP, Schräpler, Schupp, and Wagner (2005) found some indications that CAPI increases the probability of item nonresponse on income questions but also that it reduces the probability of unit nonresponse in the subsequent wave. Thus, the variation in the use of CAPI across SOEP Subsamples A through D, E, and F yields a potential for survey artifacts.

3. ANOGI: THE METHODOLOGY

Let y_i, $F_i(y)$, $f_i(y)$, μ_i, and p_i represent the income, the cumulative distribution, the density function, the expected value, and the share of subpopulation i in the overall population, respectively.[5] Let $s_i = p_i \mu_i / \mu_u$ denote the share of group i in the overall income, where subscript u denotes the union of the populations, Y_u, from which all the subsamples are drawn.[6] The overall population is composed of the union of the subpopulations. That is, $Y_u = Y_1 \cup Y_2 \cup \cdots \cup Y_n$.

Note that

$$F_u(y) = \sum_i p_i F_i(y). \tag{1}$$

That is, the cumulative distribution (or ranks) of the overall population is the weighted average of the cumulative distributions of the subpopulations, weighted by the relative sizes of the populations.[7] The formula of the Gini used in this article is the following (Lerman and Yitzhaki 1984, 1989[8]):

$$G = \frac{2\,\mathrm{cov}(y, F(y))}{\mu}, \tag{2}$$

which is twice the covariance between income y and the rank $F(y)$, standardized by mean income μ.[9] The Gini of the entire population, G_u, can be decomposed as

$$G_u = \sum_{i=1}^{n} s_i G_i O_i + G_b, \tag{3}$$

where O_i is the overlapping index of subpopulation i with the entire population (explained below), and G_b measures the between-group inequality. Equation (3) decomposes the Gini of the union into two related components: intra- and intergroup components, connected in a way that is relatively complicated. Note that while in ANOVA, the total variability is partitioned into inter- and intravariances, in ANOGI, we have inter- and intra-Ginis, but in addition, there is an extra parameter, which is the overlapping index. We will return to this implication following the explanation of the individual components.

3.1. THE OVERLAPPING INDEX AND ITS PROPERTIES[10]

The index of overlapping is the one that distinguishes the decomposition of the Gini (ANOGI) from the decomposition of the variance (ANOVA) and is the main reason for expressing a preference for the methodology suggested in this article over ANOVA.

Overlapping should be interpreted as the inverse of stratification. Stratification is a concept used by sociologists. We follow Lasswell's

(1965) definition: "In its general meaning, a stratum is a horizontal layer, usually thought of as between, above or below other such layers or strata. Stratification is the process of forming observable layers, or the state of being comprised of layers. Social stratification suggests a model in which the mass of society is constructed of layer upon layer of congealed population qualities" (p. 10).

According to Lasswell (1965), perfect stratification occurs when the observations of each population (in our case, subsample of the SOEP) are confined to a specific range of income and the ranges of incomes do not overlap. An example of a perfect stratification is the division of the society into deciles. Stratification plays an important role in the theory of relative deprivation (Runciman 1966), which argues that stratified societies can tolerate greater inequalities than nonstratified ones (Yitzhaki 1982). In our case, this property plays an important role because it tells us whether the different subsamples represent different strata.

One can rarely find a perfect stratification, and an index describing the degree of stratification is called for. The index of overlapping is actually an index describing the extent to which the different subpopulations are stratified. In this article, the goal is to evaluate the overlapping (i.e., nonstratification) and check whether the several income distributions, based on several independent panel subsamples, represent the same universe.[11]

Formally, overlapping of the overall population by subpopulation i is defined as

$$O_i = O_{ui} = \frac{\text{cov}_i(y, F_u(y))}{\text{cov}_i(y, F_i(y))}, \tag{4}$$

where, for convenience, the index u is omitted and cov_i means that the covariance is according to distribution i, that is,

$$\text{cov}_i(y, F_u(y)) = \int (y - \mu_i)(F_u(y) - \overline{F}_{ui}) f_i(y) dy, \tag{5}$$

where $\overline{F}_{ui}$ is the expected rank of population i in the union (all observations of population i are assigned their union's ranks $F_u(y)$, and $\overline{F}_{ui}$ represents their expected value).[12,13] The overlapping (4) can be further decomposed to identify the overlapping of subpopulation

i with all subpopulations that comprise the union. In other words, total overlapping of subpopulation i, O_i, is composed of overlapping of i with all subpopulations, including group i itself. This further decomposition of O_i is

$$O_i = \sum_j p_j O_{ji} = p_i O_{ii} + \sum_{j \neq i} p_j O_{ji} = p_i + \sum_{j \neq i} p_j O_{ji}, \qquad (6)$$

where $O_{ji} = \frac{\mathrm{cov}_i(y, F_j(y))}{\mathrm{cov}_i(y, F_i(y))}$ is the overlapping of group j by group i.

The properties of the overlapping index O_{ji} are the following:

(a) $O_{ji} \geq 0$. The index is equal to zero if no member of the j distribution lies in the range of distribution i (i.e., group i is a perfect stratum).

(b) O_{ji} is an increasing function of the fraction of population j that is located in the range of population i.

(c) For a given fraction of distribution j that is in the range of distribution i, the closer the observations belonging to j to the expected value of distribution i, the higher O_{ji}.

(d) If the distribution of group j is identical to the distribution of group i, then $O_{ji} = 1$. Note that by definition, $O_{ii} = 1$. This result explains the second equality in (6). Using (6), it is easy to see that $O_i \geq p_i$ is a result to be borne in mind when comparing different overlapping indices of groups with different sizes.

(e) $O_{ji} \leq 2$. That is, O_{ji} is bounded from above by 2. This maximum value will be reached if all observations belonging to distribution j that are located in the range of i are concentrated at the mean of distribution i. Note, however, that if distribution i is given, then it may be that the upper limit is lower than 2 (see Schechtman 2005). That is, if we confine distribution i to be of a specific type, such as normal, then it may be that the upper bound will be lower than 2, depending on the assumption on the distribution.

(f) In general, the higher the overlapping index O_{ji}, the lower O_{ij} will be. That is, the more group j is included in the range of distribution i, the less distribution i is expected to be included in the range of j.

Properties (a) to (f) show that O_{ji} is an index that measures the extent to which population j is included in the range of population i. Note that the indices O_{ji} and O_{ij} are not interrelated by a simple relationship. It is clear that the indices of overlapping are not independent.

3.2. BETWEEN-GROUP COMPONENT G_b AND ITS PROPERTIES

As will be seen later, we are interested in two alternative parameters representing between-groups Gini. We start with the one appearing in equation (3). The between-groups inequality G_b is defined in Yitzhaki and Lerman (1991) as

$$G_b = \frac{2\,\mathrm{cov}(\overline{Y}, \overline{F}_u)}{\mu_u}. \tag{7}$$

G_b is twice the covariance between the mean incomes of subpopulations and the subpopulations' mean ranks in the overall population, divided by overall expected income. That is, each subpopulation is represented by its mean income and by the mean rank of its members in the overall distribution. The term G_b equals zero if either the mean incomes or the mean ranks are equal for all subpopulations. In extreme cases, G_b can be negative, which occurs when the mean income is negatively correlated with mean rank.

One may argue that G_b is not really a Gini coefficient because it can be negative. An alternative between-groups Gini (G_{bp}) was defined by Pyatt (1976); Mookherjee and Shorrocks (1982), Shorrocks (1984), and Silber (1989) also follow Pyatt. In this definition, the between-groups Gini is based on the covariance between mean income in each subpopulation and its rank among the mean incomes of subpopulations. The difference between the two definitions is in the rank that is used to represent the group: Under Pyatt's approach, it is the rank of the mean income of the subpopulation, while according to Yitzhaki and Lerman (1991), it is the mean rank of all members. Generally, it can be shown that

$$G_b \leq G_{bp}. \tag{8}$$

The upper limit is reached and (8) holds as an equality if the ranges of incomes that groups occupy do not overlap (i.e., perfect stratification).

Having explained the different components, we now present a variation of decomposition (3) that will be used in this article as

$$G_u = \sum_{i=1}^{n} s_i G_i + \sum_{i=1}^{n} s_i G_i (O_i - 1) + G_{bp} + (G_b - G_{bp}). \tag{9}$$

TABLE 3: A Summary of Analysis of Gini (ANOGI) Components in Comparison to Analysis of Variance (ANOVA)

Components Parallel to ANOVA	*Formula*	*Range*
Intragroup	$IG = \sum s_i G_i$	$0 \leq IG \leq G$
Between-groups-Pyatt	$BG_p = G_{bp}$	$0 \leq BG_p \leq G_u$
	Additional information	
Overlapping effect on intragroup	$IGO = \sum s_i G_i (O_i - 1)$	
Overlapping effect on between-groups	$BGO = G_b - G_{bp}$	$-BG_p - IGO - IG \leq BGO \leq 0$

For the benefit of readers who are interested in a quick comparison with ANOVA, a summary table of ANOGI is shown in Table 3. The four components can be divided into two types: those that carry equivalent information to ANOVA (when using Gini instead of the variance as a measure of variability) and those with additional information.

3.3. SUMMARY OF THE DECOMPOSITION COMPONENTS

3.3.1. Components That Are Parallel to ANOVA

For a given overall inequality, G_u:

Intragroup component (IG). A weighted average of groups' Ginis. It reaches the lower limit if all intragroup Ginis are equal to zero. It reaches the upper limit if all groups are identical (identical to mean square error [MSE] in ANOVA).

Between-groups component, based on Pyatt (BG_P). It measures between-groups inequality, assuming a complete stratification. It reaches the lower limit, zero, if the means of all groups are equal (identical to mean square between [MSB] in ANOVA). It reaches the upper limit if all groups are concentrated at their means.

3.3.2. Additional Components

The effect of overlapping on intragroup component (IGO). This term "revises" the contribution of each subpopulation to intragroup variability, provided that inequality in the group is greater than zero. If the subpopulation and the overall population are equally distributed, then there is no revision to its contribution ($O_i = 1$). However, if a subpopulation forms a stratum in the population ($O_i < 1$), then its contribution to the intragroup component is reduced, while its contribution to between-groups is increased. On the other hand, if the scatter of the ranks of group members is larger than that of the population ($O_i > 1$), the contribution of the group to the intragroup is increased, while its contribution to between-groups is decreased.

The effect of overlapping on the between-groups component (BGO). The effect of overlapping on the between-groups component occurs only if the expected values of the subpopulations are not all equal. It is always nonpositive because overlapping reduces the ability to distinguish between the groups. It reaches the upper limit (zero) if the ranges occupied by the different groups do not overlap. Note, however, that the combined effect of the between-group inequality and the impact of overlapping on it can be negative if the means of the groups are negatively correlated with the means of the ranks. This possibility occurs if, for example, the population is composed of two groups, with one group composed of a majority of poor people and a few very rich people, while the second group is composed of the middle class. In this case, the expected income of the first group is high (because of the few rich), while its expected rank is low (because of the majority of poor people), making the correlation negative.

In the empirical application, we seek to find out whether all the intragroup Ginis are equal and whether the second, third, and fourth terms all converge to zero. We can interpret the convergence to zero as follows.

$G_{bp} = 0$ implies that all expected values are equal, $(G_b - G_{bp}) = 0$ implies that the expected ranks of the subpopulations in the overall population are equal, and $\sum_{i=1}^{n} s_i G_i (O_i - 1) = 0$ implies that each subpopulation overlaps with the entire population. Comparison of the Ginis ensures that intragroup variabilities are the same.

Clearly, we are using terms that are connected. However, each parameter adds insight, and there is no redundancy or double counting because the sum of all of them adds up to the overall Gini, and one can produce examples where one term is equal to zero and the others are not. The advantage of ANOGI over ANOVA is that the decomposition of Gini adds a new parameter to the existing inter- and intraterms—namely, the overlapping index. Hence, not only are the equivalents of the first and second moments examined, but the extent of population intertwining is also considered.

4. ESTIMATION AND TESTING

The decomposition (9) involves four parameters, that need to be estimated from the data: G_i, O_i, G_b, and G_{bp}.

The estimation technique used here is based on U-statistics. For each parameter, a kernel of the proper degree is found, and then a U-statistic is constructed. The advantage of dealing with U-statistics is that they are unbiased estimators, and their limiting distributions are normal under regularity conditions (see, e.g., Randles and Wolfe 1979; Hoeffding 1948). Also, the jackknife method for variance estimation works well for U-statistics (see Shao and Tu 1995; Arvesen 1969; Schechtman and Wang 2004). Since the estimation procedures were already detailed elsewhere, we chose to provide the estimators here and refer the reader to the relevant literature for details.

(a) Estimation of G_i

Let $Y_1, \ldots, Y_{ni}$ be a random sample from subpopulation i, with a distribution function $F_i(y)$. Then a U-statistic for estimating G_i, which is an unbiased estimator, is given by

$$\hat{G}_i = \frac{2}{n_i(n_i - 1)} \sum_{i<j} |y_i - y_j|, \tag{10}$$

where n_i is the sample size coming from subpopulation i.[14]

(b) Estimation of O_i

Recall that the numerator of O_i is a covariance, which can be expressed as a function of three means, as shown below. The denominator is simply the Gini of subpopulation i. Therefore, we represent

O_i as

$$O_i = O_{ui} = \frac{\text{cov}_i(y, F_u(y))}{\text{cov}_i(y, F_i(y))} = \frac{E_i(yF_u(y)) - E_i(y)E_i(F_u(y))}{G_i}$$
$$= \frac{\theta_1 - \theta_2 \cdot \theta_3}{G_i}. \tag{11}$$

Each mean is estimated by a U-statistic, and hence, the estimator of O_i is a function of four (dependent) U-statistics. Let

$$\hat{\theta}_1 = \frac{1}{n_i n_u} \sum_{j=1}^{n_i} y_j (\#y's \leq y_j),$$
$$\hat{\theta}_2 = \bar{y},$$

and

$$\hat{\theta}_3 = \frac{1}{n_i n_u} \sum_{j=1}^{n_i} (\#y's \leq y_j),$$

where n_u is the size of the entire population; then, combining the pieces together, the estimator of O_i is (see Schechtman 2005)

$$\hat{O}_i = \frac{\hat{\theta}_1 - \hat{\theta}_2 \cdot \hat{\theta}_3}{\hat{G}_i}. \tag{12}$$

(c) Estimation of G_b

The parameter G_b is defined in equation (7), where $\overline{F}_u$ is the vector of average ranks of the members of the n subpopulations, ranked within the entire population. The denominator of G_b can easily be estimated by the sample mean. The numerator can be written as a function of three expectations:

$$\text{cov}(\overline{Y}, \overline{F}_u) = E(\overline{Y}\overline{F}_u) - E(\overline{Y})E(\overline{F}_u).$$

The estimators of $E(\overline{YF}_u)$ and $E(\overline{F}_u)$ involve the sample version of $\overline{F}_u$. Let $\overline{F}_{u_t}$ be the tth component of $\overline{F}_u$, then $\hat{\overline{F}}_{u_t} = \frac{\sum(\#y \leq y_i)}{n_u n_t}$, where the summation is over $y_i \in$ subpopulation t, $t = 1, \ldots, n$, and $E(\overline{Y}\ \overline{F}_u)$ is estimated by $\frac{1}{n}\sum_{t=1}^{n} \overline{Y}_t \hat{\overline{F}}_{u_t}$.

(d) Estimation of G_{bp}

The parameter G_{bp} is actually a Gini of the vector of means. Therefore, its estimator is basically the same as $\hat{G}_i$, after replacing Y_i by $\overline{Y}_i$.

As mentioned above, the estimators are U-statistics or functions of several U-statistics. Therefore, inference can be made, using the fact that their limiting distributions are approximately normal under regularity conditions. The only missing link here is a way to estimate the variances, which are difficult to obtain analytically. We therefore estimated the variances using the jackknife method. The method, which can be best described as "delete one at a time," can be generally explained as follows: Given a sample $X_1, \ldots, X_n$ of size n and a sample statistic $g(X)$, whose variance needs to be estimated, follow the two steps:

1. Calculate n values $g_i(X), i = 1, \ldots, n$, where $g_i(X)$ is $g(X)$, computed for the original sample after deleting X_i (i.e., based on $(n - 1)$ observations).
2. Use the n values $g_i(X)$ to estimate the variance of $g(X)$ by $\frac{n-1}{n}\sum(g_i(X) - \overline{g}_.(X))^2$, where $\overline{g}_.(X)$ is the average of $g_1(X), \ldots, g_n(X)$. For details, see, for example, Shao and Tu (1995).

The case of jackknifing a two-sample statistic is a bit more complicated, and we will not go into details here. The interested reader can find the details in Arvesen (1969) and Schechtman and Wang (2004).

5. RESULTS OF ANOGI COMPARING DIFFERENT SUBSAMPLES

This section provides empirical results of the decomposition of the Gini by different SOEP subsamples ("old" Subsamples A through E vs. "new" Subsample F) for two different income variables. Related to the theoretical considerations in Sections 3 and 4, we would expect the following results from the empirical application if both subsamples represent the same population or universe ("=" means no significant difference):

- Mean income: $\mu_{AE} = \mu_F$
- Mean rank: $\overline{F}_{AE} = \overline{F}_F = 0.5$
- Gini coefficient: $G_{AE} = G_F$

- Overlapping index: $O_{AE} = O_F = 1$
- Between-groups inequality: $G_b = G_{bp} = 0$

Any *significant* deviation from these results would have to be interpreted as an indication that the two subsamples do *not* represent the same population.

To analyze whether our results on the income distribution are driven by selective attrition or by changing the answering behavior of respondents (see Section 2.2), we complement this analysis by using another objective variable (namely, years of education[15]) and by a subjective variable (namely, life satisfaction). Education is a noncomplex concept that is not very difficult for respondents to report. However, there is some evidence that answers to questions on satisfaction vary in quality during the time span of a panel, as over the course of the first (three) waves respondents learn to deal with this complex subjective concept better (see Landua 1991).

We analyze two income concepts: (1) annual postgovernment household income (i.e., posttax posttransfer[16]), which is a generated variable based on an explicit aggregation of various income components (labor income, capital income, private and public transfers such as pensions, child allowances, social assistance, etc.) across household members, and (2) the monthly disposable household income ("screener") asked in the household questionnaire.[17] Missing data due to item nonresponse in the components of *annual* income are imputed by means of longitudinal and various cross-sectional techniques. Recent studies provide evidence that using only cross-sectional data for imputation of missing data in panel surveys is inferior to using longitudinal data (see Spiess and Goebel 2004; Frick and Grabka 2005).[18]

In stark contrast to the annual income figures, the monthly screener variable is not imputed in case of missing data; the share of item nonresponse here ranges between 5 and 10 percent. The question of the "income screener" itself appears to be a rather simple one.[19] However, it is not easy to give a proper answer because the respondent, in most cases the household head, must calculate the net income from different income sources and, in case of larger households, across several household members.

The variable "years of education" is analyzed only for the prime age population (ages 25–55). As is the case for income, this variable is also an objective variable, describing an important social and demographic dimension. However, responding to questions about educational attainment appears to be not as complex and sensitive as is the case for income.

Finally, we also make use of the subjective measure "satisfaction with life in general," which becomes an increasingly important indicator in socioeconomic analyses (as a proxy for utility) as well as in psychological research. An important advantage of this concept is that in the answers of the respondents, there are almost no item nonresponses. However, there are "panel effects," in the sense that respondents learn to handle the question more sensitively over the course of time.

In Tables 4 through 6, we present the results of income distribution analyses and the Gini decomposition (ANOGI) for annual income in the two SOEP Subsamples A through E and the new Subsample F. In general, a statistically significant difference between the Gini coefficients is found in 2000. However, the difference disappears thereafter (values in parentheses give standard errors according to jackknife estimators). This difference is in need of an explanation, given that the two subsamples are intended to represent the (same) population of individuals living in households in Germany.[20] The long-running Subsamples A through E are in fact a conglomerate of five different population subgroups (see Section 2), which partly have been added to cope with changes in the German population caused by reunification in 1990 and by ongoing immigration, while Subsample F represents just *one* big enlargement subsample drawn in 2000.[21]

Average income in Subsample F is lower in 2000, adapts to the level of Subsamples A through E in 2001, and is almost identical in 2002. The mean rank for Subsamples A through E in the overall distribution (normalized between 0 and 1) decreases from each period to the next (0.514 to 0.5). Accordingly, the mean rank for Subsample F increases from 0.483 to 0.5. The group-specific Gini coefficients are only significantly different for the first wave of Subsample F. Inequality between groups is extremely low in all three years: In 2000, it starts at 0.22 percent of overall inequality and disappears completely in 2002.

The overlapping information shows that the identification of these two subsamples as distinct "groups" in terms of their position in the income distribution is only given in 2000. If the overlap component is larger than 1, the distribution has the highest relative densities at the tails of the other group-specific distribution, which is the case for Subsample F in 2000.

The other components of equation (9), like the two definitions of between-groups Gini, show the same pattern of convergence, which enables us to safely conclude that the difference in the samples disappear in the third year.

Possible reasons for the significant differences in the first year could be panel attrition, the imputation models used to adjust for item nonresponse, or respondent behavior effects.

- If attrition causes the distinctiveness of the two subsamples, then those who dropped out from the survey were systematically different from those persons who were willing to further participate, something already notable in the second wave.
- The explanation via the imputation of missing values could be a relevant issue if the assumption of missing at random (MAR[22]) does not hold or the imputation model (or parts of it) was not correctly specified.
- Last, but not least, response behavior may cause the significant differences in the results for 2000, if there were changes in the behavior due to learning effects in using and answering a complex questionnaire and/or by an improved personal relationship between respondent and interviewer[23] that enhanced confidence.

To differentiate between the different causes, we introduce two amendments into the analysis: (1) an income concept that is not influenced by any imputation strategy in case of missing information (i.e., the monthly income "screener") and (2) a balanced panel design that considers only those observations that were part of the survey for three consecutive years (i.e., 2000 to 2002). If panel attrition caused the differences in 2000, then these differences should disappear when using a three-year balanced panel, in contrast to the cross-sectional population, especially in Wave 1. If the imputation procedure was causing the differences for 2000, then vanishing significant results if using a nonimputed income concept could be an indication in this direction.

Table 7 shows the comparison of the results for the annual income and the monthly screener income. By definition, the share of missing values for the annual income is zero, whereas the share of imputed

TABLE 4: Analysis of Gini (ANOGI) for Subsamples A-E and F (Year 2000)

(1) *Group*	(2) *Population Share* (P_i)	(3) *Income Share* (S_i)	(4) *Mean Income* (μ_i)	(5) *Mean Rank* (F_{io})	(6) *Gini* (G_i) (SE)	(7) *Overlapping Component* (O_i)(SE)
Subsamples A-E	0.55	0.56	33,601	0.514	0.2649 (.0028)	0.9831 (.0022)
Subsample F	0.45	0.44	32,285	0.483	0.2808 (.0026)	1.0182 (.0022)
Total	1.00	1.00	33,010	0.500	0.2722 (.0021)	—
Between-groups G_b	0.0006 (.0002) 0.22%					
Within group	0.2716 99.78%					
Between groups/max. between groups (G_b/G_{bp})	0.0607					
$G_b - G_{bp}$ (see equation (9))	– 0.0093					

SOURCE: Authors' calculation from SOEP 2000.
NOTE: Standard errors are given in parentheses.

TABLE 5: Analysis of Gini (ANOGI) for Subsamples A–E and F (Year 2001)

(1) *Group*	(2) *Population Share* (P_i)	(3) *Income Share* (S_i)	(4) *Mean Income* (μ_i)	(5) *Mean Rank* (F_{io})	(6) *Gini* (G_i) (SE)	(7) *Overlapping Component* (O_i) (SE)
Subsamples A–E	0.55	0.56	34,332	0.507	0.2657 (.0037)	0.9970 (.0023)
Subsample F	0.45	0.44	33,461	0.491	0.2667 (.0021)	1.0029 (.0023)
Total	1.00	1.00	33,941	0.500	0.2662 (.0022)	—
Between-groups G_b	0.0002 (.0001) 0.07%					
Within group	0.2660 99.93%					
Between groups/max. between groups (G_b/G_{bp})	0.0310					
$G_b - G_{bp}$ (see equation (9))	– 0.0062					

SOURCE: Authors' calculation from SOEP 2000.
NOTE: Standard errors are given in parentheses.

TABLE 6: Analysis of Gini (ANOGI) for Subsamples A–E and F (Year 2002)

(1) *Group*	(2) *Population Share* (P_i)	(3) *Income Share* (S_i)	(4) *Mean Income* (μ_i)	(5) *Mean Rank* (F_{io})	(6) *Gini* (G_i) (SE)	(7) *Overlapping Component* (O_i) (SE)
Subsamples A–E	0.55	0.55	35,220	0.500	0.2817 (.0031)	1.0004 (.0025)
Subsample F	0.45	0.45	35,427	0.500	0.2839 (.0026)	0.9997 (.0026)
Total	1.00	1.00	35,313	0.500	0.2827 (.0023)	—
Between-groups G_b	0.0000 (.0000) 0.00%					
Within group	0.2827 100.00%					
Between groups/max. between groups (G_b/G_{bp})	0.0014					
$G_b - G_{bp}$ (see equation (9))	– 0.0015					

SOURCE: Authors' calculation from SOEP 2000.
NOTE: Standard errors are given in parentheses.

values for the screener is zero. Note that the trend for Subsamples A through E is more or less stable for the share of missing values as well as for the share of imputed values. The trend for Subsample F for the two income concepts is also rectified but different from Subsamples A through E. Item nonresponse in the monthly income, as well as the share of imputed values in the annual income, is higher in the first year and draws near to the level of Subsamples A through E. The small increase from Wave 2 to Wave 3 (i.e., from 2001 to 2002) may be linked to the introduction of the Euro on January 1, 2002, which complicated answering due to lack of familiarization to the "new" currency.

Unsurprisingly, the Gini coefficients for the two income concepts are different in terms of magnitude,[24] but they are identical in terms of trends and changing patterns. The increase in the Gini for 2002 appears to be very distinct. However, this fits the development of increasing income inequality in Germany since the second half of the 1990s.

Performing the same analysis for the *balanced* panel should control for panel attrition and provide an estimate for the degree of selectivity (see Table 8). The effect observed in the cross-sectional analysis is also present in the longitudinal population. We see differences for the first year, with overlapping indices clearly different from 1 and a rather quick convergence in the results over time. This holds not only for Gini, mean, and overlapping index but also, slightly less distinct, for the share of item nonresponse and the mass of imputed income.

In conclusion, the comparison of Tables 7 and 8 indicates strong evidence that neither panel attrition nor imputation of item nonresponse cause the differences between the results for the first wave of Subsample F and the longer running Subsamples A through E. If learning and confidence building are important within empirical surveys, this phenomenon should be found in other variables as well. However, if learning effects are not relevant, results should remain stable over time.

- According to the literature, it is well known that the "response styles" for *satisfaction* questions change over time (see, e.g., Schräpler 2004). Based on SOEP data, Landua (1991) has shown that respondents of questions about satisfaction change their answering behavior over the first four years. Within the first years, the respondents tend to overstate their satisfaction more often than in later waves by ticking the highest two categories on an 11-point scale running from 0 (*completely*

TABLE 7: Comparison of Annual and Monthly Income (Cross-Sectional)

	Annual Income			*Monthly Income (Screener)*		
Year	*A–E*	*F*	*Total*	*A–E*	*F*	*Total*
			Mean (in DM)			
2000	33,601	32,285	33,010	2,543	2,486	2,518
2001	34,332	33,461	33,941	2,594	2,527	2,565
2002	35,220	35,427	35,313	2,691	2,635	2,666
			Gini* 100 (Gi) (SE)			
2000	26.49 (.28)	28.08 (.26)	27.22 (.21)	24.61 (.31)	25.97 (.22)	25.22 (.23)
2001	26.57 (.37)	26.67 (.21)	26.62 (.22)	24.60 (.31)	25.15 (.22)	24.85 (.21)
2002	28.17 (.31)	28.39 (.26)	28.27 (.23)	25.56 (.27)	26.46 (.26)	25.96 (.22)
			Overlapping Component (O_i) (SE)			
2000	0.9831 (.0022)	1.0182 (.0022)	—	0.9872 (.0023)	1.0153 (.0022)	—
2001	0.9970 (.0023)	1.0029 (.0023)	—	0.9932 (.0023)	1.0076 (.0024)	—
2002	1.0004 (.0025)	0.9997 (.0026)	—	0.9941 (.0027)	1.0069 (.0025)	—

(continued)

TABLE 7 (continued)

	Annual Income			*Monthly Income (Screener)*		
Year	*A–E*	*F*	*Total*	*A–E*	*F*	*Total*
	Percentage at Least One Missing Income Component			Percentage Missing		
2000	20.42	27.68	23.69	6.58	9.78	8.02
2001	18.73	24.29	21.23	5.10	8.20	6.49
2002	20.55	23.91	22.06	5.91	8.86	7.24
	Between-Groups Inequality G_b (*SE*) (Percentage Share of Total Inequality)					
2000	0.0006	(0.0002)	(0.22%)	0.0003	(0.0002)	(0.11%)
2001	0.0002	(0.0001)	(0.07%)	0.0003	(0.0002)	(0.10%)
2002	0.0000	(0.0000)	(0.00%)	0.0002	(0.0001)	(0.07%)

SOURCE: Authors' calculation from SOEP 2000–2002.
NOTE: Standard errors are given in parentheses.

TABLE 8: Comparison of Annual and Monthly Income (Three-Year Balanced Panel Design)

	Annual Income			*Monthly Income (Screener)*		
Year	*A–E*	*F*	*Total*	*A–E*	*F*	*Total*
			Mean (in DM)			
2000	33,588	31,855	32,805	2,539	2,473	2,510
2001	34,440	33,344	33,944	2,591	2,529	2,563
2002	35,650	35,675	35,661	2,702	2,654	2,680
			Gini * 100 (G_i) (SE)			
2000	26.19 (.36)	27.63 (.31)	26.87 (.25)	24.41 (.30)	25.41 (.24)	24.86 (.23)
2001	26.22 (.32)	26.47 (.24)	26.35 (.22)	24.37 (.30)	24.98 (.28)	24.65 (.22)
2002	27.67 (.31)	28.38 (.34)	27.99 (.24)	25.60 (.35)	26.40 (.27)	25.97 (.23)
			Overlapping Component (O_i) (SE)			
2000	0.9827 (.0026)	1.0171 (.0025)	—	0.9901 (.0027)	1.0115 (.0025)	—
2001	0.9953 (.0026)	1.0045 (.0026)	—	0.9923 (.0026)	1.0087 (.0026)	—
2002	0.9976 (.0026)	1.0030 (.0027)	—	0.9968 (.0028)	1.0038 (.0025)	—

(continued)

TABLE 8 (continued)

	Annual Income			*Monthly Income (Screener)*		
Year	*A–E*	*F*	*Total*	*A–E*	*F*	*Total*
	Percentage at Least One Missing Income Component			Percentage Missing		
2000	19.93	26.97	23.11	5.62	10.02	7.61
2001	18.03	23.42	20.47	5.28	7.84	6.44
2002	20.18	23.51	21.69	5.94	8.11	6.92
	Between-Groups Inequality G_b (SE) (Percentage Share of Total Inequality)					
2000	0.0009	(.0003)	(0.35%)	0.0003	(.0002)	(0.13%)
2001	0.0003	(.0002)	(0.12%)	0.0002	(.0002)	(0.09%)
2002	0.0000	(.0000)	(0.00%)	0.0001	(.0001)	(0.05%)

SOURCE: Authors' calculation from SOEP 2000–2002.
NOTE: Standard errors are given in parentheses.

dissatisfied) to 10 (*completely satisfied*). On the basis of this finding, we may expect, with respect to life satisfaction, differences for all three years, but with a declining trend.

- On the other hand, the results for *educational attainment*, being objective and not intimate information to respond to, should not be different between the two subsamples, not even in the very first wave of subsample.

To differentiate attrition from learning effects, all analyses are carried out for the cross-sectional population, as well as for the balanced panel design. In fact, the analysis of educational attainment shows no significant differences for means and Gini coefficients between the two subsamples for all three years in cross-sectional and longitudinal design (see left and right panels of Table 9). Especially in the starting year, as hypothesized, the overlapping indices are also similar. As can be expected (at least for the balanced panel population), we find that means increase in both subsamples, and inequality does follow the same trend in both samples as well (though it is not a priori clear whether to expect an increase or a decrease in inequality of educational attainment).

As expected, the results for life satisfaction draw a very different picture (see Table 10). The Gini indices[25] differ more between the two subsamples, and the overlapping indices remain significantly different from 1 for both groups throughout the entire period—however, there is an indication of convergence. These processes are similar for both cross-sectional and longitudinal populations.

With respect to the mean as well as to the marginal distribution of the variable "life satisfaction" (see Appendix C), our results clearly confirm the finding by Landua (1991), which states that first-time users of such a scale tend to tick the highest categories more often. In 2000, more than twice as many respondents in Subsample F than in Subsamples A through E indicated that they are "completely satisfied": 9.7 percent and 4.2 percent, respectively. Until 2002, this gap decreases most remarkably to only 2.5 percentage points. Again, this picture does not change when moving from a purely cross-sectional to a longitudinal design, which we interpret as an indication for learning effects among the short panel members.

TABLE 9: Comparison of Educational Attainment (Years of Education)

	Cross-Sectional Design			*Three-Year Balanced Panel Design*		
Year	*A–E*	*F*	*Total*	*A–E*	*F*	*Total*
			Mean (in Years)			
2000	12.51	12.50	12.50	12.52	12.51	12.51
2001	12.54	12.61	12.57	12.53	12.58	12.55
2002	12.54	12.60	12.57	12.56	12.60	12.58
			Gini* 100 (G_i) (SE)			
2000	11.09 (.16)	10.75 (.09)	10.94 (.11)	11.01 (.19)	10.58 (.11)	10.82 (.13)
2001	11.12 (.17)	10.73 (.09)	10.94 (.11)	11.07 (.19)	10.65 (.10)	10.88 (.13)
2002	11.04 (.18)	10.71 (.11)	10.89 (.12)	11.15 (.19)	10.69 (.10)	10.95 (.13)

(continued)

TABLE 9 (continued)

	Cross-Sectional Design			*Three-Year Balanced Panel Design*		
Year	*A–E*	*F*	*Total*	*A–E*	*F*	*Total*
			Overlapping Component (O_i) (SE)			
2000	1.0044 (.0033)	0.9950 (.0030)	—	1.0065 (.0041)	0.9925 (.0033)	—
2001	1.0090 (.0038)	0.9897 (.0030)	—	1.0093 (.0041)	0.9889 (.0033)	—
2002	1.0071 (.0041)	0.9920 (.0032)	—	1.0103 (.0042)	0.9876 (.0032)	—
			Between-Groups Inequality G_b (SE) (Percentage Share of Total Inequality)			
2000	0.0000	(.0000)	(0.00%)	0.0000	(.0000)	(0.00%)
2001	0.0000	(.0000)	(0.02%)	0.0000	(.0000)	(0.01%)
2002	0.0000	(.0000)	(0.01%)	0.0000	(.0000)	(0.01%)

SOURCE: Authors' calculation from SOEP 2000–2002.
NOTE: Standard errors are given in parentheses.

TABLE 10: Comparison of Life Satisfaction

	Cross-Sectional Design			*Three-Year Balanced Panel Design*		
Year	*A–E*	*F*	*Total*	*A–E*	*F*	*Total*
			Mean			
2000	6.89	7.28	7.07	6.91	7.33	7.10
2001	6.93	7.26	7.08	6.96	7.28	7.10
2002	6.77	7.07	6.91	6.74	7.06	6.88
			Gini* 100 (G_i) (SE)			
2000	14.06 (.14)	13.61 (.10)	13.94 (.13)	13.68 (.15)	13.22 (.12)	13.58 (.15)
2001	14.12 (.14)	12.90 (.11)	13.63 (.14)	13.80 (.15)	12.76 (.14)	13.39 (.15)
2002	14.51 (.16)	13.39 (.15)	14.05 (.15)	14.68 (.17)	13.34 (.16)	14.13 (.16)
			Overlapping Component (O_i) (SE)			
2000	0.9672 (.0024)	1.0265 (.0024)	—	0.9632 (.0028)	1.0282 (.0029)	—
2001	0.9852 (.0025)	1.0087 (.0026)	—	0.9840 (.0028)	1.0098 (.0031)	—
2002	0.9878 (.0026)	1.0067 (.0030)	—	0.9894 (.0029)	1.0040 (.0032)	—
			Between-Groups Inequality G_b (SE) (Percentage Share of Total Inequality)			
2000	0.0017	(.0003)	(1.24%)	0.0021	(.0004)	(1.55%)
2001	0.0012	(.0002)	(0.89%)	0.0013	(.0003)	(0.94%)
2002	0.0011	(.0003)	(0.77%)	0.0012	(.0003)	(0.84%)

SOURCE: Authors' calculation from SOEP 2000–2002.
NOTE: Standard errors are given in parentheses.

Tables 7 to 10 also give information about between-groups inequality (G_b) and its contribution to total inequality. In line with the group-specific results, between-groups inequality for annual and monthly income measures (in cross-sectional as well as in longitudinal perspective) converges over the three-year period. In the case of education, we hardly find any between-groups inequality even in the first year, whereas in the case of satisfaction—even after three years—G_b is significantly different from zero. In other words, for all considered indicators, the contribution of between-groups inequality to total inequality is irrelevant after three years, except for satisfaction, where it still amounts to almost 1 percent in 2002.

6. CONCLUDING SUMMARY

The main aim of this article is to study the "representativeness" of different subsamples of the German SOEP in the field of income distribution. This issue was chosen for analysis because an unbiased measurement of household incomes is (a) a real challenge for survey research (see Canberra Group 2001), and (b) the analysis of income distribution and mobility is one of the main tasks of household panel surveys such as the SOEP.

However, it appears that the inclusion of a new (independently drawn) representative subsample into an existing, longer running panel survey may yield slightly deviating results, which may be caused by panel attrition or by differences in the answering behavior of respondents.

The methodology used in this article is based on ANOGI, which differs from the analysis of variance because it includes an additional term that reflects the overlapping between the distributions of the different subsamples. This is the first time that this methodology has been empirically applied, and we believe that this article demonstrates its usefulness.

Concluding from our empirical results and a discussion of survey methodology issues employed in the setup of the considered subsamples, we cannot reject the hypothesis that both represent the same universe. Recapitulating from our analyses on objective and subjective indicators for income, education and satisfaction within a cross-sectional and a longitudinal framework, we conclude that there is convincing evidence within the SOEP for changing respondent behavior due to learning effects with respect to the applied instruments and questioning. However, we find the convergence process, in which empirical results based on a new subsample approach those of a longer running subsample, to be of different lengths for the various indicators under investigation. This may be driven by the different degree of complexity of the underlying constructs, especially in the case of subjective indicators such as "satisfaction."

With respect to the originally motivating question on income inequality, we would especially reject the hypothesis that due to panel attrition, a new subsample after two waves is as selective as a longer running panel. Instead of arguing that results from a cross-sectional survey (as is the first wave of any panel study) yield more reliable estimates than those stemming from a panel that may be affected by attrition, we would like to reverse this argument and state that a reliable measurement of complex issues, such as the construction of an annual income measure or a satisfaction measure, clearly profits from repeated surveying, as is the case in panel studies.

APPENDIX A

The German SOEP—Details

The main random Subsample A of SOEP included around 4,500 households. To allow separate analyses of the five groups of labor migrants most strongly represented in the Federal Republic of Germany in 1984, they were oversampled in the study with a total of 1,400 households in a disproportional random sample approach. This random Subsample B was itself subdivided into five subgroups.

To observe the massive social and economic changes in East Germany, along with their respective impacts, the first wave of the East German subsample was collected in June 1990, *before* the currency, economic, and social union in Germany occurred on July 1. This sample, Subsample C, consists of about 2,200 households.

Since the start of SOEP in 1984, Western Europe (and especially Germany) has experienced immigration on a large scale, which cannot be covered by any ongoing longitudinal survey. To correct for this bias, an explicit supplement for immigrants was necessary. For this reason, Subsample D was collected in 1994-1995 for about 500 households with immigrants who had arrived since 1984.

In 1998, a "supplementary random sample" had been started as a test. This subsample fulfilled a number of aims: (1) stabilization of the number of observations in the SOEP for cross-sectional and longitudinal data, (2) allowing for analysis of "panel effects," and (3) allowing for analysis of representativeness.

It was proven that a supplementary sample such as this could be integrated in a user-friendly manner into the ongoing "old subsamples" (see Spiess and Rendtel 2000 for solving the problem of setting up an integrated weighting scheme). Thus, the methodological basis was established for significantly increasing the sample size, which would boost the value of the study for policy analysis by allowing the changes for relatively small groups of the population to be analyzed on the basis of sufficiently large numbers of cases. An enlargement such as this took place in the year 2000. The first wave of Subsample F consists of 10,890 adult respondents and 2,993 children who live in 6,052 households.

APPENDIX B

SOEP Sampling Procedures—Details

In Germany, as in many other countries, sampling of foreigners is a practical problem. Although a *major improvement* of random-route samples, which are conducted in Germany, was implemented with Subsample F, a difference from Subsamples A and B remains. In Subsample B, foreigners were sampled on the basis of local address registers. This expensive method was not possible for Subsample F, where the local polling registers (*Wählerverzeichnisse*) are the basis for drawing sample points, sample points with different shares of non-German citizens are not drawn by probabilities that mirror their correct weight for the population living within the German territory (*Wohnbevölkerung*). Thus, all standard random-route samples (according to the rules of the German Association of Fieldwork Companies [Arbeits Kreis Deutscher Marktforschungs-institute (ADM)]) underestimate the share of foreigners in Germany. To reduce the impact of this shortcoming, the fieldwork organization Infratest introduced an overrepresentation of foreigners in the random walk of SOEP Subsample F. The number of addresses to be collected during a random walk was doubled, but within the so-called "excess addresses," only households with foreigners were selected for interviews. Through this procedure, the share of foreigners in the sample mirrors the true share in the underlying population quite well. However, the structure of the foreigners is eventually biased because sample points with a high share of foreigners still have a downward biased probability of being included in the sample. In principle, this bias can be corrected for by weighting procedures (to be applied).

It is a common problem for population surveys around the world to adequately cover households living in institutions. Unfortunately, it was not possible to include institutionalized households in the first waves of the SOEP in a representative manner. However, by following respondents after a residential move, a panel takes into consideration those who left private households for institutionalized households, whereby over the course of time, the institutionalized population is included in the SOEP. However, on the other hand, any new subsample starts with this problem, which may produce an artificial difference between old and new subsamples. Nevertheless, the population in the starting wave of Subsample F (year 2000) in fact includes 47 institutionalized households (approximately 0.8 percent of all household interviews in this sample) as compared to 85 institutionalized households in Subsamples A through E (approximately 1.2 percent).

One may conclude from the discussion of these various sampling procedures applied to SOEP that the respective universe or population to be represented by the different subsamples (A through E vs. F) differs only marginally.

APPENDIX C

Distribution of Life Satisfaction—Details

TABLE C1: Comparison of Life Satisfaction Distributions (in Percentages)

	Cross-Sectional Design			*Three-Year Balanced Panel Design*		
Year	*A–E*	*F*	*Total*	*A–E*	*F*	*Total*
2000						
0 = low	0.5	0.5	0.5	0.5	0.5	0.5
1	0.4	0.3	0.4	0.4	0.3	0.4
2	1.6	1.0	1.3	1.2	0.9	1.0
3	2.3	2.1	2.2	2.4	1.8	2.1
4	3.9	2.7	3.4	3.7	2.6	3.2
5	13.6	11.6	12.7	13.5	11.4	12.5
6	11.3	9.3	10.4	11.4	9.0	10.3
7	22.3	18.7	20.7	23.1	18.4	21.0
8	29.8	30.7	30.2	29.9	31.6	30.7
9	10.1	13.4	11.6	9.8	13.9	11.6
10 = high	4.2	9.7	6.7	4.1	9.7	6.6
2001						
0 = low	0.5	0.4	0.5	0.4	0.3	0.4
1	0.6	0.2	0.4	0.4	0.3	0.3
2	1.4	0.9	1.2	1.3	1.1	1.2
3	2.5	2.0	2.3	2.4	1.8	2.1
4	4.0	2.4	3.3	4.1	2.2	3.3
5	12.4	11.2	11.8	12.6	10.9	11.8
6	11.1	9.4	10.4	11.0	9.2	10.2
7	22.4	19.9	21.3	22.5	20.5	21.6
8	29.6	32.6	31.0	30.0	32.4	31.1
9	10.8	13.6	12.0	10.6	14.0	12.1
10 = high	4.7	7.4	5.9	4.7	7.4	5.9
2002						
0 = low	0.5	0.4	0.5	0.5	0.4	0.5
1	0.5	0.5	0.5	0.4	0.5	0.4
2	1.6	1.0	1.3	1.8	1.0	1.4
3	2.9	2.4	2.7	3.0	2.4	2.7
4	4.6	3.1	4.0	4.7	3.1	4.0
5	13.8	11.7	12.9	14.1	11.5	12.9
6	11.8	11.2	11.5	12.1	11.1	11.7
7	23.3	21.9	22.7	23.5	22.4	23.0
8	28.1	30.6	29.2	27.2	30.9	28.9
9	9.5	11.5	10.4	9.3	11.4	10.2
10 = high	3.2	5.7	4.3	3.2	5.3	4.2

SOURCE: Authors' calculation from SOEP 2000–2002.

NOTES

1. The German Socio-Economic Panel Study (SOEP) data are made available in user-friendly form ("scientific use file") worldwide to all independent research institutions. Analysis of the data is supported by an extensive online service (www.diw.de/gsoep). Today, the SOEP is also widely used by international organizations such as the Organization for Economic Cooperation and Development (OECD), especially for the analysis of income distributions.

2. For more detailed information, see http://panel.gsoep.de/soepinfo2002/info/persons.html.

3. See Schräpler and Wagner (2001) for details.

4. For an overview on this line of discussion, see Couper and Nicholls (1998), de Leeuw (2002), and Fuchs, Couper, and Hansen (2000).

5. In the sample, the cumulative distribution is estimated by the rank of the observation, normalized to be between 0 and 1.

6. Note that Y_u represents the entire population only if there is no attrition. Otherwise, it represents a biased population of the "entire population," with the bias being a function of the patterns of the attritions.

7. Actually, p_i is not a true population parameter. It is added here for generality and for handling samples of different sizes.

8. Lerman and Yitzhaki (1984) derive the covariance formula; Lerman and Yitzhaki (1989) adjust it to handle weighted samples.

9. Note that the relative version of Gini is used, which is as if one uses the coefficient of variation to perform analysis of variance (ANOVA). Relative measure is chosen since it is the common parameter used in the income distribution literature.

10. The proofs of all statements in this section are given in Yitzhaki (1994).

11. An alternative use is to search for stratification. For example, Heller and Yitzhaki (2003) argue that a perfect classification into groups is achieved if members of each group are similar among themselves (low intragroup variability) and different from others (stratified). This property of the decomposition of the Gini enables them to use overlapping as an indicator of the quality of classification of snails into groups, according to different observed variables.

12. Ranking observations of one variable according to the distribution of another variable is a rare concept in statistics. However, it is common in sports, where each athlete is frequently ranked in his or her country and according to other scales (world, continent, gender, age group, etc.).

13. It is worth noting that the O_i is a kind of a Gini correlation. See Schechtman and Yitzhaki (1987, 1999) for the properties of Gini correlations.

14. See Schechtman and Yitzhaki (1987) for details.

15. This information is derived from various variables on formal qualification levels for schooling and vocational training.

16. The process of deriving annual income figures in the SOEP and the tax simulation procedures are described in Grabka and Frick (2003) and Schwarze (1995).

17. Both income measures are adjusted—without having an impact on the methodological research question—for different household needs by the modified OECD equivalence scale. This scale is used to assign the appropriate weight to each household member in the sample. This scale gives the first adult a weight of 1.0, additional adults (older than 14 years of age) a weight of 0.5, and children (up to age 14) a weight of 0.3.

18. The imputation of item non-response-related missing income data in the SOEP follows a two-step procedure: The general principle is to employ the "row and column imputation technique," as developed by Little and Su (1989), which takes advantage of information on the very same individual over time by combining row (unit) and column (period/trend) information.

However, given that the empirical implementation of this method fails in all those cases where a given income component is not observed in any other panel wave, purely cross-sectional imputation techniques have to be used, which are based on data observed from other units in the very same wave. See Grabka and Frick (2003) for a complete overview of the techniques applied for the various SOEP income variables.

19. The question reads as follows: "If you take a look at the total income from all members of the household: how high is the monthly household income today? *Please state the net monthly income, which means after deductions for taxes and social security. Please include regular income such as pensions, housing allowance, child allowance, grants for higher education support payments, etc. If you do not know the exact amount, please estimate the amount per month*."

20. Institutionalized households are included in all empirical analyses presented here. Sensitivity analyses focusing on the impact of this subpopulation on income distribution measures show the expected result: Income of such nonprivate households is below average, and inequality decreases when excluding these households from the analysis. Nevertheless, the substantive finding in Table 4 concerning the significant deviation of the Gini coefficients for the two SOEP subsamples persists: 0.264 for Subsamples A through E versus 0.279 for Subsample F instead of 0.265 versus 0.281, respectively. Further details about coverage of institutionalized households in the SOEP are given in Appendix B.

21. It should be noted that households consisting solely of adult respondents who recently immigrated to Germany (i.e., after 1998) had a positive sampling probability in the new Subsample F. However, this was not the case in the ones that already existed (A-E), where the most recent subsample was drawn in 1998. However, in our data, this phenomenon appears to be of minor relevance, given that there are only six such households in Subsample F.

22. For a detailed description of missing data within surveys, see, for example, Little and Rubin (2002) or Schafer (1997).

23. In principle, each year the same interviewer consults the very same interviewees in the SOEP.

24. Note that the annual income concept used here clearly differs from the one of monthly income, which is observed from regular (normally monthly) income flows. Following the recommendations of the Canberra Group (2001), our measure of annual income explicitly considers capital income, irregular cash income components such as Christmas bonuses or gratifications, and a major noncash income component (namely, imputed rent from owner-occupied housing). Due to the rather unequal distribution of these income components, inequality for annual income is higher than for monthly income.

25. To restrict the upper bound of the Gini to be 1, we transformed the original 11-point scale in the following way: Values 1 through 10 have been multiplied by 10, and the value 0 was coded into 0.1.

REFERENCES

Arvesen, James N. 1969. "Jackknifing U-Statistics." *Annals of Mathematical Statistics* 40:2076-100.

Canberra Group. 2001. *Expert Group on Household Income Statistics: Final Report and Recommendations*. Ottawa, Canada: Canberra Group.

Couper, Mick P. and William L. Nicholls. 1998. "The History and Development of Computer Assisted Survey Information Collection." Pp. 1-22 in *Computer Assisted Survey Information*

Collection, edited by M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls, and J. M. O'Reilly. New York: John Wiley.

de Leeuw, Edith D. 2002. "The Effect of Computer Assisted Interviewing on Data Quality: A Review of the Evidence" [CD-ROM]. In *Social Science Methodology in the New Millennium*, edited by J. Blasius, J. Hox, E. de Leeuw, and P. Schmidt. Opladen: Leske + Budrich.

Frick, Joachim R. and Markus M. Grabka. 2005. "Item-Non-Response on Income Questions in Panel Surveys: Incidence, Imputation and the Impact on Inequality and Mobility." *Allgemeines Statistisches Archiv* 89:49-60.

Fuchs, Marek, Mick Couper, and Sue E. Hansen. 2000. "Technology Effects: Interview Duration in CAPI and Paper and Pencil Surveys." Pp. 149-66 in *Developments in Survey Methodology*, edited by A. Ferligoj and A. Mrvar. Ljubljana: FDV.

Grabka, Markus M. and Joachim R. Frick. 2003. "Imputation of Item-Non-Response on Income Questions in the SOEP 1984–2002." DIW Research Notes No. 29, DIW Berlin.

Haisken-DeNew, John P. and Joachim R. Frick. 2003. *Desktop Companion to the German Socio-Economic Panel Study (GSOEP), Version 7.0—Update to Wave 19*. Berlin: DIW Berlin.

Heller, Joseph and Shlomo Yitzhaki. 2003. "Classifying Fossils: When the Distribution Is Not Normal." Unpublished manuscript.

Hoeffding, Wassily. 1948. "A Class of Statistics With Asymptotic Normal Distribution." *Annals of Mathematical Statistics* 19:293-325.

Landua, Detlef. 1991. "An Attempt to Classify Satisfaction Changes: Methodological and Content Aspects of a Longitudinal Problem." *Social Indicators Research* 26:221-41.

Lasswell, Thomas E. 1965. *Class and Stratum*. Boston: Houghton Mifflin.

Laurie, Heather. 2000. "From PAPI to CAPI: Consequences for Data Quality on the British Household Panel Survey." Paper presented at the Fifth International Conference on Logic and Methodology, October 3-6, University of Cologne, Cologne, Germany.

Lerman, Robert and Shlomo Yitzhaki. 1984. "A Note on the Calculation and Interpretation of the Gini Index." *Economics Letters* 15:363-8.

———. 1989. "Improving the Accuracy of Estimates of Gini Coefficient." *Journal of Econometrics* 42 (1): 43-7.

Little, Roderick J. A. and Donald B. Rubin. 2002. *Statistical Analysis With Missing Data*. New York: John Wiley.

Little, Roderick J. A. and Hong-Lin Su. 1989. "Item Non-Response in Panel Surveys." Pp. 400-25 in *Panel Surveys*, edited by D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh. New York: John Wiley.

Mookherjee, Dilip and Anthony F. Shorrocks. 1982. "A Decomposition Analysis of the Trend in U.K. Income Inequality." *Economic Journal* 92(368): 886-902.

Pyatt, Graham. 1976. "On the Interpretation and Disaggregation of Gini Coefficient." *Economic Journal* 86(342): 243-55.

Randles, Ronald H. and Douglas A. Wolfe. 1979. *Introduction to the Theory of Nonparametric Statistics*. New York: John Wiley.

Rendtel, Ulrich. 1995. *Panelausfälle und Panelrepräsentativität*. Frankfurt: Campus.

Rendtel, Ulrich, Gert G. Wagner, and Joachim R. Frick. 1995. "Eine Strategie zur Kontrolle von Längsschnittgewichtungen in Panelerhebungen—Das Beispiel des Soziooekonomischen Panels (SOEP)." *Allgemeines Statistisches Archiv* 79(3): 252-77.

Runciman, Walter Garrison. 1966. *Relative Deprivation and Social Justice*. London: Routledge Kegan Paul.

Schafer, Joseph L. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Schechtman, Edna. 2005. "Stratification: Measuring and Inference." *Communications in Statistics, Theory and Methods* 34(11): 2133–2145.

Schechtman, Edna and Soujin Wang. 2004. "Jackknifing Two-Sample Statistics." *Journal of Statistical Planning and Inference* 119(2): 329-40.

Schechtman, Edna and Shlomo Yitzhaki. 1987. "A Measure of Association Based on Gini's Mean Difference." *Communications in Statistics: Theory and Methods* 16(1): 207-31.

———. 1999. "On the Proper Bounds of the Gini Correlation." *Economics Letters* 63:133-8.

Schräpler, Jörg-Peter. 2004. "Respondent Behavior in Panel Studies: A Case Study for Income-Nonresponse by Means of the German Socio-Economic Panel (SOEP)." *Sociological Methods & Research* 33(1): 118-56.

Schräpler, Jörg-Peter, Jürgen Schupp, and Gert G. Wagner. 2005. "Changing From PAPI to CAPI: A Longitudinal Study Dealing with Mode-Effects in the German Socio-Economic Panel (SOEP) Using an Experimental Design." Unpublished manuscript, DIW Berlin.

Schräpler, Jörg-Peter and Gert G. Wagner. 2001. "Das Verhalten von Interviewern—Darstellung und ausgewählte Analysen am Beispiel des 'Interviewerpanels' des Soziooekonomischen Panels." *Allgemeines Statistisches Archiv* 85:45-66.

Schwarze, Johannes. 1995. "Simulating German Income and Social Security Tax Payments Using the GSOEP." Cross-National Studies in Ageing Project Paper No. 19, Syracuse University.

Shao, Jun and Dong Sheng Tu. 1995. *The Jackknife and Bootstrap*. New York: Springer.

Shorrocks, Anthony F. 1984. "Inequality Decomposition by Population Subgroups." *Econometrica* 52(6): 1369-85.

Silber, Jacques. 1989. "Factor Components, Population Subgroups, and the Computation of Gini Index of Inequality." *Review of Economics and Statistics* 71(2): 107-15.

Spiess, Martin and Jan Goebel. 2004. "A Comparison of Different Imputation Rules." Pp. 293-316 in *Harmonisation of Panel Surveys and Data Quality*, edited by M. Ehling and U. Rendtel. Wiesbaden: Federal Statistical Office.

Spiess, Martin and Ulrich Rendtel. 2000. "Combining an Ongoing Panel With a New Cross-Sectional Sample." DIW Discussion Paper No. 198, DIW Berlin.

Wagner, Gert G., Richard V. Burkhauser, and Friederike Behringer. 1993. "The English Language Public Use File of the German Socio-Economic Panel Study." *Journal of Human Resources* 28(2): 429-33.

Yitzhaki, Shlomo. 1982. "Relative Deprivation and Economic Welfare." *European Economic Review* 17:99-113.

———. 1994. "Economic Distance and Overlapping of Distributions." *Journal of Econometrics* 61:147-59.

Yitzhaki, Shlomo and Robert Lerman. 1991. "Income Stratification and Income Inequality." *Review of Income and Wealth* 37(3): 313-29.

Joachim R. Frick is a senior research economist at DIW Berlin and a deputy director of the German Socio-Economic Panel Study (SOEP). His current research interests are in the field of population economics and social policy, especially with respect to immigration and income distribution as well as on survey methodology. Recent journal articles are included in Review of Income and Wealth, International Migration Review, Journal of Comparative Economics, Journal of Population Economics, *and* Allgemeines Statistisches Archiv.

Jan Goebel is a research associate at DIW Berlin, and his main research interest covers income distribution analysis, income dynamics, and item nonresponse and imputation. Recent publications are included in the Journal of European Social Policy *and* Allgemeines Statistisches Archiv.

Edna Schechtman is an associate professor of statistics at Ben Gurion University, Israel. Her current research interests are in the field of measures based on the Gini index and applied statistics. Recent journal articles are included in Journal of Quality Technology, Ophthalmology, Statistics & Probability Letters, Human Factors, Value in Health, Journal of Economic Inequality, Drugs and Aging, The Economic Quarterly, Metron, *and the* Journal of Statistical Planning and Inference.

Gert G. Wagner is a professor of economics at TUB (Berlin University of Technology) and director of the German Socio-Economic Panel Study (SOEP) at DIW Berlin. His current research interests are in the fields of population economics and survey methodology. Recent journal articles are included in Australian Social Monitor, Early Childhood Research Quarterly, Geneva Papers on Risk and Insurance, International Migration Review, *and the* Journal of Public Economics.

Shlomo Yitzhaki is a professor of economics at Hebrew University and the government statistician at the Central Bureau of Statistics, Jerusalem, Israel, and is affiliated with the National Bureau of Economic Research (United States). His research interests are in the fields of income distribution, taxation, portfolio theory, and econometrics. Recent journal articles are included in Review of Income and Wealth, American Economic Review, *and* Metron.