Panel Conditioning and Self-reported Satisfaction: Evidence from International Panel Data

Bert Van Landeghem
University of Sheffield, Maastricht University, and IZA
b.vanlandeghem@sheffield.ac.uk

23 December 2013

Abstract

Data derived from survey questions that ask respondents about their overall life satisfaction, are increasingly being used in multidimensional analyses of trends and fluctuations in a countrys prosperity. However, researchers often disagree on time trends in these data, and on their long-term relationship with other macroeconomic indicators. This paper aims to contribute to this debate, and shows that, ceteris paribus, respondents rate their well-being lower the more experience they have in answering such questions. The analysis uses panel datasets that contain refreshment samples. These serve as natural experiments which are crucial to overcome identification problems as will be discussed. The results can help to resolve at least part of the discrepancies found in the literature on time trends, but can be of importance for the design of empirical strategies in other settings such as programme evaluations.

keywords: Life Satisfaction, Subjective Data, Cognitive Biases, Measurement

1 Introduction

During the past several decades, economists have paid increasing attention to 'stated utility' measures, derived from simple survey questions such as:

> On a scale from 0 (very unsatisfied) to 10 (very satisfied), how satisfied are you with your life, all things considered?

Increasing evidence underpins the internal validity of such data (Krueger and Schkade, 2008; Sgroi et al., 2010), as well as the external validity (Oswald and Wu, 2010). Researchers have examined many relationships between happiness[1] data and socioeconomic variables, and whether these should be interpreted as causal. Happiness has been shown to be strongly correlated with relational goods and social capital (Becchetti et al., 2008; Powdthavee, 2008), with major life events such as changes in marital status, unemployment, bereavement, and disability (Clark et al., 2007; Gardner and Oswald, 2006; Oswald and Powdthavee, 2008a,b), with social status (Blanchflower et al., 2009; Luttmer, 2005; Ravallion and Lokshin, 2010; Senik, 2004, 2008a, 2009), and with expectations and aspirations (de Grip et al., 2012; McBride, 2010; Senik, 2008b). Others studies have analyzed geographical differences in life satisfaction (Oswald and Wu, 2011) or to what extent happiness is determined by genetics (De Neve et al., 2012). Very recently, economists are changing their focus from the causes of happiness to what happiness itself can cause. De Neve and Oswald (2012), Oswald et al. (2012) and Proto et al. (2012), find a causal impact of happiness on productivity, while Ifcher and Zarghamee (2011) find a negative causal link between happiness and time preference (i.e. preference of current over future utility).

In fact, the analysis of subjective well-being data has gained so much importance that even in mainstream economics journals, methodological issues regarding such data are repeatedly being discussed. Oswald (2008) argues that one's self-reported life satisfaction might be a concave function of *true* life satisfaction. This would imply that the log-linear relationship, which is often found between income and reported well-being, does not necessarily reflect the functional relationship between income and true well-being. Ferrer-i-Carbonell and Frijters (2004) find that treating happiness data as cardinal rather than as ordinal has no substantial impact on regression results, but accounting for fixed individual effects has a dramatic impact on the results. The mode of interview

[1] Terms such as happiness, life satisfaction, and subjective well-being are often used interchangeably in the literature.

might also be of importance. Dolan and Kavetsos (2012), for example, find that, ceteris paribus, individuals interviewed over the telephone report, on average, higher life satisfaction scores than those being interviewed face-to-face. Others have been concerned about anchoring effects, that is, people do not report an absolute level of happiness but rather their happiness relative to a reference point. Recently, these concerns have been addressed using vignettes, describing hypothetical individuals or households. The ranking of these vignettes on a happiness ladder by respondents can then give us some information about heterogeneity in anchors across individuals or countries (see Beegle et al., 2012; Bonsang and van Soest, 2012, and Kapteyn et al., 2011 for discussions).

It is clear that taking into account methodological concerns is not just a matter of measuring levels and relationships with more accuracy: they can sometimes completely change conclusions, and the revised results can hence potentially support a theory that was rejected before. A recent salient example is provided by Heffetz and Rabin (2013). Making use of survey data which record the number of calls before the successful attempt, their analysis shows how survey nonresponse can be of such a substantial size that the sign of between-group differences in mean happiness are reversed. In their raw data, women are happier than men, but after correcting the results for the fact that hard-to-reach men are happier than hard-to-reach women, they come to the opposite conclusion.

There are several open-ended debates going on about happiness patterns and their relationship with other socioeconomic variables, and it is hence likely that methodological investigations remain indispensable. One of the most lively and legendary of these debates is about trends of well-being over time, and the relationship between economic growth and subjective well-being.

For example, in a series of papers, Richard Easterlin has presented what is now called the "Easterlin Paradox": while countries have grown dramatically richer over time, people did not get happier (Easterlin 1974, 1995, 2001; Easterlin et al., 2010). A number of researchers have, however, challenged this paradox and find a positive log-linear relationship between subjective well-being and GDP over time (Deaton, 2008; Sacks et al., 2010; Stevenson and Wolfers, 2008, 2013). In some instances, papers that contradict each other rely on the same datasets but apply different empirical strategies, such that differences in results are driven by subtle differences in (technical) identification assumptions. In some instances, however, results contradict when very similar empirical strategies are used, on different datasets but drawn from the same population during the same time span. For example, Stevenson and Wolfers (2009) find that female happi-

ness has been declining since the 1970s in industrialized countries, both absolutely and relative to men. They use several datasets but rely on General Social Survey (GSS) data for the United States. Herbst (2011) replicates the U.S. analysis with the DDB Needham Life Style Survey for a similar time span. He finds a decline in male happiness as well and that, if anything, men's happiness declines at a quicker pace than female happiness. Di Tella and MacCulloch (2006) illustrate for West Germany with panel data from the German Socioeconomic Panel (SOEP) a decline in happiness in Germany in the period 1985-2000 despite steady economic growth. Others, however, using repeated cross- sectional data from the Eurobarometer survey such as Easterlin et al. (2010) and Stevenson and Wolfers (2008) find a flat or increasing trend in happiness over that period. Comparisons between trends in panel data and repeated cross-sections for similar regions and time spans are discussed in the web appendix. However, as to provide the motivation for the investigation described in the remainder of the paper, Figure 1 shows trends in standardized life satisfaction for West Germany, where one line is constructed from Eurobarometer data and the other line from German SOEP data. The graph shows an overall difference into trends: the German SOEP data suggest an overall decline in happiness, especially in the eighties, while the Eurobarometer data show a slight increase. The web appendix of this paper will document similar discrepancies in other countries. Obviously such graphs leave room for interpretation. The difference in trends might be due to panel attrition, changing survey designs, but maybe also due to the fact that trained respondents answer subjective well-being questions differently than unexperienced respondents.

Given the existing interdisciplinary literature on the measurement of subjective well-being, it is not obvious to a priori exclude this latter possibility. Indeed, a "life satisfaction" or "happiness question" is a retrospective, and not especially easy question to answer, and limitations to people's cognitive capacities are likely to add noise to the data. Researchers such as Deaton (2012) and Conti and Pudney (2011), argue that subtle changes in questionnaire design (such as with respect to question ordering or response options) can influence results, not only the average scores in the raw data but also the distribution of happiness in the population and the correlations with socioeconomic variables. Kahneman and Krueger (2006) give an overview of several lab and field experiments investigating measurement issues of subjective well-being. For example, when people are being subject to a stream of pleasant and less pleasant stimuli for a certain period (e.g. having their hands immersed in water with varying temperature), the last stimulus will have a disproportionately high weight in their overall retrospective report

on well-being for that period (Kahneman et al., 1983). Kahneman and Krueger (2006) also mention evidence that answers to such retrospective questions are influenced by circumstances of the moment, e.g. the weather of the day (Schwarz and Clore, 1983). It is not implausible that such cognitive biases may change the more trained a respondent becomes in reflecting about his general life satisfaction.

The phenomenon that trained respondents answer differently to survey questions than unexperienced respondents is often referred to in the (interdisciplinary) literature as a "panel effect" or "panel conditioning". Das et al. (2011) and Toepoel et al. (2009) offer a cross-disciplinary overview of the literature on panel conditioning, and by comparing two samples, one fresh and the other, more experienced, they find that a panel effect is especially strong for knowledge questions but not for attitudinal questions. When they compare answers from trained respondents with answers from fresh respondents, they find that trained respondents score much higher on questions such as:

> Do you know what Campylobacter is? Do you know what Cross infection is?

while there does not seem to be a notable difference between the two groups for questions such as:

> Do you think pensions will be higher about ten years from now? Do you think people will be more satisfied with their pensions about ten years from now?

The authors note, however, that research on panel conditioning is still rather limited and that generally no attempt is made to distinguish panel conditioning from panel attrition.

In the context of life satisfaction measures, the steep decline in average well-being scores in the first rounds of a panel have not been unnoticed. Both Frijters and Beatton (2012), and Kassenboehmer and Haisken-Denew (2012), who endeavour to explain the frequently found U-shaped pattern of subjective well-being over the life cycle, find a negative coefficient on a variable measuring the number of times having participated into the panel. However, apart from potential attrition biases, in a model that includes individual fixed effects, age, and time effects, it might be difficult to interpret a variable measuring the length of staying in the panel. Indeed, multicollinearity problems force researchers (often unconsciously) to make arbitrary identifying assumptions; these can make a dramatic impact on results and hence explain why recent life-course studies using the same data sometimes report different outcomes (see Van Landeghem, 2012a

Appendix One for an algebraic illustration). Sharpe and Gilbert (1998) find in a lab experiment that testing individuals for depression twice with a one-week interval leads to a decrease in self-reported negative emotions and does not seem to have an effect on self-reported positive emotions. However, the very specific small sample of college undergraduates, the nature of the questions which aim to detect depression rather than overall life satisfaction, as well as the very short time period that has elapsed between the two sessions, make it difficult to extrapolate the conclusions to subjective well-being data from nation-wide panel surveys repeated with yearly intervals.

As the phenomenon might be of great importance for interpreting existing subjective well-being models and designing empirical frameworks for future studies, this paper will investigate whether cognitive biases for subjective well-being data change the more experience people have in answering such questions. The analysis will take seriously the possibility that results are artificial due to model misspecification, changing survey design or panel attrition. Substantial refreshment samples in different calendar years in panel data will be exploited as quasi-natural experiments to help to better identify the effect. Methodologically, it is hence inspired though not identical to the approach of Das et al; (2011) due to the different context of this paper's datasets, which bring along different limitations but also different opportunities.

The remainder of the paper is as follows. Section Two documents the data that will be used. Section Three points out why a quasi-experiment is so crucial to obtain identification by presenting an extended age-period-cohort model, and it outlines the core of the empirical strategy. Section Four presents and discusses the results, and Section Five concludes.

2 Data

The study draws upon two panel datasets, the German Socioeconomic Panel (SOEP), and the Swiss Household Panel (SHP). In both panels, by definition, the aim is to re-interview the same individuals in successive rounds. New individuals can enter, however, for several reasons. For example, new samples can be added to the survey to oversample some minorities,to sample individuals who were excluded from the main panel (e.g. migrants) or to refresh the panel. Apart from adding new samples, each year, the datasets will contain a small number of new respondents. First, members from interviewed house-

holds will reach the eligible age to enter the panel. Second, if a new member eligible for the survey moves into the household, the enumerator will attempt to interview this new member as well. Third, if a household member leaves the household, the aim is to follow the respondent and, at the same time, try to interview other members eligible for the survey in that person's new household. Generally, new respondents not stemming from a refreshment sample account for around 3% of the respondents in the second round of a (sub)sample, and this percentage of this type of first-year respondents will increase over the lifetime of the (sub)sample. They ensure that each age group is represented in each survey round.[2]

Among the datasets analysed, the German SOEP is probably the one that allows the most extensive analysis and that is the most common panel dataset in happiness research. The German SOEP is provided by DIW Berlin and is repeated at yearly intervals, running from 1984 for West Germany and 1990 for East Germany (see Wagner et al., 2007).

The observations can be categorized into eight[3] different samples:

- Sample A (started in 1984) represents the West German population, while sample C (started in 1990) represents the East German population.
- Samples E, F, and H (started in 1998, 2000 and 2006 respectively) are refreshment samples.
- Samples B and D (started in 1984 and 1994 respectively) are immigrant subsamples, and sample G (started in 2002) comprises high-income households. The latter three samples are excluded from the analysis since they are drawn from a different population than the Original West- or East-German sample.

This results in a sample of around 31,000 individuals: 21,000 in West Germany and 10,000 in East Germany, or 190,500 and 92,500 person-year observations, respectively.

The following question on subjective well-being, asked at the end of a face-to-face questionnaire, has been in the survey from 1984 onwards and ever since has reappeared in every survey round.

[2]To enhance readability, the refreshers stemming from these three sources will be referred to as 'natural refreshers'.

[3]In 2009, a ninth sample was started, but the analysis will not make use of this sample as these individuals are currently observed in only two rounds of the panel.

> On a scale from 0 (completely dissatisfied) to 10 (completely satisfied): How satisfied are you with your life, all things considered?

Across all rounds, the average life satisfaction score in West Germany equals 7.15 with a standard error of 1.82. In East Germany, life satisfaction is, on average, considerably lower than in West Germany, with an average score of 6.41 and standard error of 1.80.

The SHP is a panel repeated at yearly intervals that started in 1999 and is run by the FORS, the Swiss Centre of Expertise in the Social Sciences. Currently, there are 12 rounds available (up to 2010), with more than 95,000 person-year observations and 18,300 individuals. In 2004, a refreshment sample was started. Respondents are interviewed by telephone (See Voorpostel et al., 2011 for more details). A general life satisfaction question is asked from the year 2000 onwards, which proceeds as follows:

> In general, how satisfied are you with your life if 0 means 'not at all satisfied' and 10 means 'completely satisfied'?

The average score across all person-year observations equals 8.01 with a standard error of 1.48. Of those who answered, 16.5% consider themselves completely satisfied. Swiss respondents thus seem to report, on average, a considerably higher happiness score than German respondents. However, insights from Dolan and Kavetsos (2012) suggest that this difference might likely be, at least partly, due to a different mode of interview rather than by a difference in true happiness between the two populations.

3 Identification and the Empirical Strategy

3.1 Identification Caveats

Before outlining the empirical strategy, it may be useful to point out more in detail why deviating from the standard happiness regressions is crucial. As many panel datasets now include happiness data for several survey rounds, the standard "happiness equation" generally includes individual-fixed effects and a set of other controls, in particular age and time dummies. Hence, it might at a first glance seem logical to study the effect of being a trained respondent by extending the model with the number of survey participations as an additional explanatory characteristic. The latter characteristic could be added either as a linear term, a higher-order polynomial or a set of dummies.

Such a set-up, however, suffers from an identification problem by construction, of which the nature and consequences are discussed below using an extended age-period-cohort model.[4]

If we consider the course of an individual in a panel over time, somebody's age, the calender year of the interview, the number of times participated into the panel, are all identical arithmetic series of the form $X_{t+1} = X_t + 1$: only the starting value is different. An individual aged 18 in calendar year 2000 participating at the panel for the first time, will be 19 in calendar year 2001 when he participates into the panel for the second time. This causes severe multicollinearity problems and hence identification problems in models that include all three variables as well as an individual fixed effect.

To see this, denote $X_{i,t}$, $Y_{i,t}$ and $Z_{i,t}$ the number of survey participations, age, and the calendar year for individual $i = 1 \dots N$ at time $t = 1 \dots T$. Since these variables are linearly increasing for an individual by one over each time period, the three variables are interrelated as follows: $Y_{i,t} = X_{i,t} + \lambda_1^i$ $Z_{i,t} = X_{i,t} + \lambda_2^i$ where λ_1 and λ_2 are two integers which can vary *across* though not *within* individuals.

A regression with a linear term of X, Y, and Z simultaneously as right-hand-side variables in an individual-fixed effects regression suffers from multicollinearity problems since we can think of linear combinations that equal an individual-specific constant.

Indeed:

$$Y_{i,t} - X_{i,t} = \lambda_1^i \tag{1}$$

and

$$Z_{i,t} - X_{i,t} = \lambda_2^i \tag{2}$$

Hence, we will need to leave out two out of the three linear terms from the regression in order to resolve the multicollinearity problem, and the estimated coefficient on the remaining term will be a mixture of the linear trend of the independent variable across the three explanatory variables (number of survey participations, age and calendar time).

[4]The identification problem at hand is very similar to the classical additive age-period-cohort model, in which one's age and the calendar time are collinear with year-of-birth or an individual-fixed effect (see e.g. Heckman and Robb, 1985; Kapteyn et al., 2005). Since Wooden and Li (forthcoming) seem to have overlooked the issue in a response to an earlier version of this manuscript (Van landeghem, 2012b), this version wants to draw some more attention to the identification issue.

There are no multicollinearities between the individual fixed effects and higher-order terms of X, Y, and Z.

Hence, it is clear that the extended version of the age-period-cohort model leads to very similar identification problems. Researchers such as Attanasio (1993) and McKenzie (2006) point out, all in their own style, that the relationship between the latter characteristics and the dependent variable can only be known up to deviations from a linear trend if one does not want to make any further assumptions. With other words, the functional relationship between the dependent variable on the one hand, and the independent variables on the other hand, is only known from the second derivative onwards. This implies, for example, if we do not want to make any identification assumptions, that we cannot identify the path of well-being over the life cycle, but that we can still identify structural breaks, and exclude certain patterns (see also Van Landeghem, 2012a). The same goes for the extended version with the number of survey participations, except that now two linear terms will need to be dropped from the regression framework. One can disguise though not solve the problem by changing the functional specification. The most common practice would be to replace the linear terms and higher-order polynomials by sets of dummies, a strategy which is attractive if there is enough data to allow such a flexible estimation. However, statistical software packages will automatically drop additional dummies in order to obtain identification, which leads to arbitrary normalization assumptions and consequently huge biases if these assumptions are not entirely correct (see Van Landeghem, 2012a for a numerical illustration).

Figures 2 and 3 aim to graphically illustrate the enormous impact of different normalization assumptions on results. Using the German SOEP for West Germany from 1984 to 2010 and for individuals aged 17 to 85, subjective well-being is regressed on individual-fixed effects, a set of time dummies, a set of age dummies, and a set of survey participation dummies. The regression is run twice omitting a different combination of dummies, which means that both regressions are based on different normalization assumptions and that they allocate the linear trend in the data differently.

In both regressions, the first dummy of each set is left out, which is the usual practice when dummy variables are included in a regression. This means that the dummies for the first time of participation, age 17 and calendar year 1984 are left out. Due to the identification issue discussed above, we need to leave out additional dummies and make assumptions about the slope. In the first regression (Normalization one), the dummies

for age 18 and calendar year 1985 are left out. In the second regression (Normalization two), the dummies for age 53 and calendar year 1994 are left out.

Figure 2 shows that under Normalization One, the path of well-being over time is increasing very strongly, and well-being is almost 2.2points higher in 2010 than in 1986. Under Normalization Two, however, the path is rather flat. Figure 3 shows that under Normalization One, the path of well-being across the number of survey participations is very steep: the top of the curve is around 2.6 points higher than the bottom. Under Normalization Two, the curve is still steep but much less, with approximately 0.9 points between the top and the bottom of the curve.

These graphical illustrations should make it clear that the identification issues discussed above are nontrivial, and that a substantially different strategy is required to investigate the research question.

3.2 Empirical strategy

Both datasets contain refreshment samples in the strict sense, that is, new samples drawn from the same population from which the original sample was drawn. Refreshment samples are often added in the course of a panel study to increase the sample size, allowing for a wider possibility of analyses, or at least to mitigate any decrease in sample size due to panel attrition.

Such refreshment samples can serve as a natural experiment and help us identify whether answers on (subjective well-being) survey questions are prone to a panel effect. Indeed, they allow us to compare the average responses between a group of first-time respondents on the one hand and more experienced respondents on the other hand. When comparing a more experienced sample with refreshers within the same panel study, one need not worry about differences in survey design. Of course, the more comparable the samples are in observed and unobserved characteristics, the less biased will be the estimated panel effect by a simple comparison of means and vice versa, and several robustness checks should help to get a clearer picture of the reliability of the results.

As a start, one can compare the means of well-being between fresh respondents and more experienced respondents,conditional on some characteristics which are likely not prone to a panel effect themselves. This would correspond to the following cross-

sectional regression specification, that can be estimated for each calendar year in which a refreshment sample was introduced:

$$WB_{it} = \alpha_t + \beta_t R_{it} + X'_{it}\gamma + e_{it} \tag{3}$$

Where the subscript i indexes the individual and t the calendar year, WB denotes subjective well-being and R is an indicator taking the value of one when an individual i either belongs to a refreshment sample started in year t or happens to be a natural refresher belonging to the more experienced sample in that year. e_{it} is an individual-specific error term, and α_t and β_t parameters to be estimated. The coefficient β_t can be interpreted as the difference in well-being between the fresh and more experienced respondents in year t. X_{it} is a vector representing control variables. Throughout the analysis, these will include age dummies, marital status dummies, labour force status dummies and education dummies. Finally, the vector includes a dummy which equals 1 when the individual happened to enter the panel at some point due to the process of natural refreshment.

One may of course be sceptical as to whether the two groups of new and experienced respondents are comparable and if the results are clouded by panel attrition. Indeed, especially in the German SOEP, refreshment samples are introduced rather late in the panel and this might compromise comparability. One might consider the strategy of Das et al. (2011), who extend the framework of Keisuke et al. (2001) by recognizing that differences between an experienced sample and a refreshment sample can be due to not only panel attrition but also panel conditioning. Das et al. (2011) point out if no assumptions regarding the attrition process are to be made, one can still identify upper and lower bounds of the panel effect. They illustrate this with dichotomous responses and calculate bounds by assuming that all attritors would have chosen either 0 or 1. In this context, however, calculating bounds seems less useful. The range of possible ordinal responses is much larger than just two options, and moreover, the attrition rate in the considered datasets is relatively high (see below). However, while Das et al. (2011) only have two rounds of data, this paper uses datasets in which refreshment samples are all introduced well before the last calendar year for which data are available. Hence, one can rerun regressions as described in Equation (3), but now restricting the refreshment sample to those who will respond to the subjective well-being question in at least N consecutive survey rounds. Such an exercise should give us a clue of the direction of the bias that panel attrition is causing.

Finally, an alternative strategy will be followed to overcome the issue of panel attrition, and that will moreover help us to get an idea about the dynamics of a panel effect. Is a panel effect of importance only between the first and second interview, or does it cumulate across several interview rounds? After defining $\bar{N}_i$ to be the total number of consecutive rounds in which the individual i answered the subjective well-being question, one can define the following pooled cross-sectional regression model:

$$WB_{it}|\bar{N}_i \geq N = \sum_{n=1}^{N} (\beta_n I_n) + X'_{it}\Delta + \sum_{t=2}^{T} (\gamma_t D_t) + \alpha + e_{it} \tag{4}$$

Again, WB denotes subjective well-being. α and e_{it} are a constant and an error term, respectively. The subscripts i, t, and n are to index the individual, the calendar year and the number of subsequent interviews an individual has participated into the panel. I_1, I_2 $\ldots I_N$ are indicators for the first, second, $\ldots$, Nth subsequent interview. The baseline category consists of those who are being interviewed more than N times. X is the vector of covariates as outlined under Equation (3). Some calendar years will have many more new respondents than average years, in particular those years in which a new sample was started. Hence, time dummies (D_2 to D_T) are included as controls to address the problem of nonrandom distribution of newcomers across calendar years. Since the regressions are run on a subsample of individuals who are in the sample for at least N consecutive survey rounds, we need not worry that the path of the panel effect over the different survey rounds (from the first to the Nth) is clouded by panel attrition.[5] Panel ageing might be an issue, since an individual answering the survey for the nth time can never be younger than the minimum age of a respondent plus $n-1$. This should be a minor concern, though, when N is taken relatively small. Indeed, even though the literature does not entirely agree on how life satisfaction evolves through life, it seems that researchers do at least agree that life cycle happiness evolves rather smoothly and not with big jumps (Blanchflower and Oswald, 2008; Easterlin, 2006; Fischer, 2009; Gwozdz and Sousa-Poza, 2010). Hence, a higher-order polynomial of age (rather than a full set of age dummies) will be included in the regression.[6]

[5]It is obvious that those individuals that stay in the panel for at least five consecutive years are not entirely representative for the populations from which the datasets are drawn. However, we can, at least for this subpopulation, obtain a clearer view on a panel effect's existence and dynamics.

[6]As a robustness check, one could also compare time trends in repeated cross-sections with time trends in panel data, although such an analysis does not allow us to convincingly disentangle panel conditioning from panel attrition. Results from this exercise can be found in an older version, and will be made available in a web appendix.

4 Results

The German SOEP contains three refreshment samples in the years 1998, 2000, and 2006, while the SHP introduced a refreshment sample in 2004. For West Germany, the three refreshment samples (plus the natural refreshers who entered in the same years) offer us 1,747, 8,891, and 2,171 new respondents, respectively, while the figures are 449, 2,245, and 591 for East Germany. In the SHP, the 2004 panel refreshment (together with the natural refreshers in that year) bring us data on new respondents for 5,371 individuals.

The differences in average scores, conditional on a set of covariates, for the new respondents on the one hand and the experienced sample in the corresponding calendar year on the other hand are obtained by running regression specifications of the form as in Equation (3), and are presented in Table 1 for West Germany, East Germany, and Switzerland.

In all cases, scores in the calendar year in which a refreshment sample is started are higher for the refreshment sample than for the more experienced sample. The results are statistically significant at any significance level, and the magnitude is substantial, varying between 0.16 and 0.64 on a 0-10 scale. For completeness, Table 2 and Table 3 report results separately for men and women, respectively and show the results hold for both sexes.

There might be concern the refreshment samples are nevertheless substantially different from the more experienced samples. For example, Table 4 shows us probit equations for each survey year in which a refreshment sample was started, and these predict whether or not an individual is a new respondent. The results from the regressions suggest that socioeconomic characteristics are not always completely randomly distributed across the refreshment and the experienced samples due to sampling error or attrition biases, or maybe because the explanatory variables are subject to panel conditioning themselves. The signs and size of the coefficients on the variables predicting the likelihood of belonging to a refreshment sample are, however, not very consistent across the different regressions. One common trend across the different regressions seems to be that singles are more likely to be in the refreshment sample than in the more experienced sample. This is not completely illogical, since singles are more mobile and thus more prone to quit the panel. It also seems that, across the different regressions,

individuals with lower household income are more likely to be first-time respondents. We have a closer look at the distribution of this latter variable in Table 5, which shows the difference between real household income in the refreshment subsample on the one hand and in the experienced sample on the other hand.

The results suggest that differences in average income between the refreshment samples and experienced samples are rather small (see Table 5). Real household income in the refreshment samples for Germany are, on average, 0 to 17% lower than in the more experienced sample. As for the SHP, real household income is around 6% lower in the refreshment sample than in the experienced sample. These results thus seem to offer reassurance that the differences between well-being in refreshment and experienced samples shown in Table 1 indeed reflect a panel effect, since it is well-known that life satisfaction scores are correlated positively with income. The observation that individuals in the refreshment sample have, on average, a lower income than individuals in the more experienced sample is in line with the observation in the literature that people with lower income are more likely to exit the panel (Kroh, 2011).

The data offer us, however, more possibilities to investigate the bias of panel attrition. As pointed out in Section 2, the refreshment samples are started well before the last round of data. In order to take a next step to make experienced respondents more comparable with fresh respondents, Table 6 shows results from regression specifications described in Equation (3), restricting the refreshment samples to those individuals who will be in the panel for at least three more years. The results of this exercise seem to suggest that the discrepancy is even slightly higher, and they hence offer additional evidence that a selection effect would bias a panel effect towards zero rather than the reverse. Indeed, low life satisfaction scores are good predictors for future attrition (Kroh, 2011; Lipps, 2007).

Of course, in order for the latter robustness check to be convincing, the course of attrition over a sample's age should be similar for both the experienced and the refreshment samples. Attrition rates could be influenced by time-varying socioeconomic factors, or by factors related to the data collection (e.g. different interviewers, different management, etc.). Large differences in attrition rates might then be an indication that the characteristics of the attritors across the experienced and the refreshment samples are not comparable. Figures 4, 5, and 6 show the attrition rate over the life cycle for the different samples for West Germany, East Germany, and Switzerland, respectively. An individual is regarded as an attritor at age n of a sample if the individual was a

respondent at age one, and if the individual answered the subjective well-being question strictly less than n times. The evolution of the attrition rates over a sample's life cycle within a panel are rather similar, concavely increasing. Attrition in the SHP is much higher than in the German SOEP, but this goes for both the 1999 sample as well as for the 2004 refreshment sample.

In the Swiss panel, the refreshment sample is introduced five years after the start of the panel, and four years after the first time the subjective well-being question has been asked. This fairly short time span between the original and refreshment sample might help to claim a convincing quasi-experiment, especially if we can condition on some additional controls and correct for panel attrition. One might be more skeptical about the results from the German SOEP, however, especially for West Germany, since the first refreshment sample has only been introduced 14 years after the start of the survey. Hence, in the exploratory regression analysis in Table 1 Specification 1, one is comparing refreshers from 1998 with a mixture of individuals who mostly entered the panel in 1984, and individuals who entered the panel between 1985 and 1997 as natural refreshers. Hence, one might have remaining worries about the comparability of the sample, even after including a few covariates (such as an indicator for having entered the panel through natural refreshment) and applying a correction for panel attrition which however leads to an even more pronounced effect. Therefore, Table 7 shows regressions of the form as in Equation (3), but now the subjective well-being of a refreshment sample is compared in its year of introduction with only these individuals who stem from the previous substantial refreshment. Thus, 2000 refreshers are compared with individuals who were newcomers in 1998, and 2006 refreshers are compared with individuals who were fresh respondents in 2000.

When comparing the 2000 and 1998 refreshers, the panel effect turns out to be rather small, both for West and East Germany, and is only significant at conventional significance levels after applying the correction for panel attrition. The comparison of the 2006 with the 2000 sample, however, again shows us a very substantial panel effect. The subjective well-being of the refreshers is, on average, 0.39 points higher than that of the more experienced respondents in West Germany, and 0.56 in East Germany. The differences are statistically different from zero at any significance level, and are again more pronounced after correcting for panel attrition.

As a final approach to investigate the phenomenon, and in order to investigate the size and dynamics of panel conditioning, regressions as of the form displayed in Equation

(4) are run with N equal to five. Hence, a subsample of individuals is selected who are re-interviewed in four or more consecutive years after the initial interview in which the question measuring subjective well-being was asked.[7] Regression results are displayed in Table 8 for West Germany, East Germany, and Switzerland, respectively. The regression results suggest that the panel effect is not entirely established between being interviewed for the first and second time but that it accumulates over the different survey rounds. In West Germany, there is a substantial panel effect from the second to the fifth interview, which is also statistically significant at any conventional significance level.[8] There is an estimated panel effect of -0.13 and -0.11 during the second and third interview, respectively. There seems to be a panel effect of -0.09 during the fourth interview, and of -0.04 during the fifth interview. For East Germany, we see a similar pattern of negative panel effects recurring from interview to interview, although the null hypothesis that coefficients on subsequent interview indicators are equal can only be rejected with slightly higher P-values than for West Germany, ranging from 0.00 to 0.08. This is not surprising, since the sample sizes are much smaller.

The pattern for Switzerland is in line with that of Germany, but no panel effect is measured any longer after the third interview, and a panel effect is only statistically significant (be it at any conventional level) for the second interview.

The coefficient on the dummy for being interviewed for the fifth time can be interpreted as the negative of a residual panel effect, that is, which will be established over the interviews after the fifth has taken place. The coefficient on this dummy for Switzerland is nearly equal to 0, while for West and East Germany, this coefficient still has a substantial magnitude of 0.18 and 0.29, respectively. The reason why the accumulation path of panel effects for Switzerland is much shorter than for Germany is rather speculative, but one should note that in the SHP, 90% of people answering the well-being question for the first time were asked in either 2000 or 2004. Moreover, one should keep in mind that, contrary to the path of panel effects from the first to the fifth interviews, the estimate for a residual panel effect might be slightly clouded by panel attrition. Indeed, some respondents will no longer be interviewed after the fifth interview, while others will remain in the panel for many years.

[7] In the SHP, the first interview for an individual does not necessarily equal the interview in which he was first asked about his life satisfaction, as a question on subjective well-being was only introduced in the second round of the panel.

[8] To be clear, a panel effect for interview n is calculated as $\beta_n - \beta_{n-1}$.

5 Conclusion

A question asking to give a retrospective report on one's overall well-being is not easy to answer, and the literature shows that such answers might suffer from several cognitive biases. Fortunately, these biases often just add some random noise to the data in many empirical frameworks. In some cases, however, they can lead us to wrong conclusions. A discrepancy in time trends of subjective well-being derived from panel data and repeated cross-sections triggered the question whether subjective well-being questions are subject to a panel effect, or, in other words, whether trained respondents answer these questions differently than new respondents.

This paper's aim was hence to systematically investigate the latter question. The analysis has used substantial refreshments in two panel datasets, the German Socioeconomic Panel and the Swiss Household panel, as quasi-experiments for the identification strategy. Subjective well-being scores are substantially higher in the refreshment sample than in the more experienced sample in the corresponding calendar year. Most importantly, these results do not seem to be driven by panel attrition, and it seems that panel effects accumulate across subsequent interviews. Methodologically, the paper stresses the importance of the quasi-natural experiments for clean identification, illustrated with an extended age-period-cohort model.

Throughout the presentation of the analysis, it will have become obvious that, although the apparent presence of a panel effect, there is quite some variation in the estimated magnitudes across the different datasets and subsamples. This might be due to some remaining sampling error, or due to the fact that the panel effects interact with other time-varying variables. Moreover, one might think of alternative explanations for the panel effect other than changing cognitive biases, e.g. that asking people about their subjective well-being does change well-being itself.[9] Nevertheless, despite the variation in magnitude and the open question about the exact source which the panel effect stems from, the analysis seems to offer evidence that trained respondents answer subjective well-being questions differently than new respondents, and that the results might have important implications for subjective well-being research.

First, they might be useful to help to reinterpret or understand results of past studies. Since the main protagonists of the debate on the long-term association between

[9] Zwane et al. (2011) find in a development context that surveying people about their health and finances changes their behaviour in these domains.

economic growth and subjective well-being are using repeated cross-sectional data and time series data, this paper's findings will not cause this debate to end. However, it might nevertheless help to make significant progress in resolving parts of the puzzle and to explain some economically counterintuitive results. For example, it can help to explain the negative trend in balanced panel data from the German SOEP as discussed in Di Tella and MacCulloch, while GDP is increasing. It will be harder to clarify the discrepancies between the findings of Stevenson and Wolfers (2009) and Herbst (2011) for gender-specific well-being trends in the United States, both based on different repeated cross-sectional datasets. Although many routes should be investigated to resolve this discrepancy (such as differences in nonresponse), one might even here need to consider the role of a panel effect. Indeed, Herbst (2011) uses data that are sampled from a large pool of individuals who have previously signalled their interest in participating to surveys. When time goes by and after having participated to several different surveys, they are likely to become professional survey respondents.

Second, results can be useful for the design of the empirical framework of future studies. Obviously they can guide the still expanding research on trends in well-being, as they suggest that repeated cross-sections and pseudo panel data might be favoured over genuine panel data in studying the latter phenomenon. Moreover, they might help fine tune policy evaluation design. For example, oversampling a minority in a certain year of a household panel to compare its well-being scores with those of the experienced subsample might lead us to wrong conclusions, even if attrition in the older sample is negligible. Third, the algebraic and graphical illustration of identification issues in the standard fixed-effects regressions should remind researchers to look for alternative strategies (such as exploiting natural experiments) to further gain insights into the phenomenon, to make explicit the identification assumptions and the sensitivity of the results when these assumptions are not entirely met in reality. This goes for further studying the phenomenon of panel conditioning, but it is also of importance if one endeavours to contribute to the debate on life-cycle patterns in subjective well-being or in any other variable. Finally, the identification of a substantial panel effect in subjective well-being data might help draw attention to this phenomenon in other areas of research, where survey data based on cognitive questions are being used.

References

Attanasio, O. (1993) A Cohort Analysis of Saving Behavior by U.S. Households. *NBER Working Papers* No. 4454, National Bureau of Economic Research.

Becchetti, L., A. Pelloni and F. Rossetti (2008) Relational Goods, Sociability, and Happiness. *Kyklos* 61, 343-363.

Beegle, K., H. Himelein and M. Ravallion (2012) Frame-of-reference Bias in Subjective Welfare. *Journal of Economic Behavior and Organization* 81, 556-570.

Blanchflower, D. and A. Oswald (2008) Is Well-being U-shaped over the Life Cycle? *Social Science and Medicine* 66, 1733-1749.

Blanchflower, D., A. Oswald and B. Van Landeghem (2009) Imitative Obesity and Relative Utility. *Journal of the European Economic Association* 7, 528-538.

Bonsang, E. and A. van Soest (2012) Satisfaction with Job and Income among Older Individuals across European Countries. *Social Indicators Research* 105, 227-254.

Clark, A., E. Diener, Y. Georgellis and R. Lucas (2007) Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis. *Economic Journal* 118, 222-243.

Clark, A. and C. Senik (2010) Will GDP Growth Increase Happiness in Developing Countries? *PSE Working Papers* 2010-43.

Conti, G. and S. Pudney (2011) Survey Design and the Analysis of Satisfaction. *Review of Economics and Statistics* 93, 1087-1093.

Das, J., A. van Soest and V. Toepoel (2011) Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys. *Sociological Methods and Research* 40, 32-56.

Deaton, A. (2008) Income, Health, and Well-being around the World: Evidence from the Gallup World Poll. *Journal of Economic Perspectives* 22, 53-72.

Deaton, A. (2012) The Financial Crisis and the Well-being of Americans. *Oxford Economic Papers* 64, 1-26.

de Grip, A., M. Lindeboom and R. Montizaan (2012) Shattered Dreams: The Effects of Changing the Pension System Late in the Game. *The Economic Journal* 122, 1-25.

De Neve, J.-M., N. Christakis, J. Fowler and B. Frey (2012) Genes, Economics, and Happiness. *Journal of Neuroscience, Psychology, and Economics*, Forthcoming.

De Neve, J.-M. and A. Oswald (2012) Estimating the Influence of Life Satisfaction and Positive Affect on Later Income Using Sibling Fixed-effects. *Proceedings of the National Academy of Sciences*, Forthcoming.

Di Tella, R. and R. MacCulloch (2008) Gross National Happiness as an Answer to the Easterlin Paradox? *Journal of Development Economics* 86, 22-42.

Di Tella, R., R. MacCulloch and A. Oswald (2001) Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness. *American Economic Review* 91, 335-341.

Di Tella, R., R. MacCulloch and A. Oswald (2003) The Macroeconomics of Happiness. *Review of Economics and Statistics* 85, 809-827.

Dolan, P. and G. Kavetsos (2012) Happy Talk: Mode of Administration Effects on Subjective Well-being. *CEP Discussion Paper* No. 1159, Centre for Economic Performance.

Easterlin, R. (1974) Does Economic Growth Improve the Human Lot? In Paul A. David and Melvin W. Reder, eds., *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, New York: Academic Press, Inc.

Easterlin, R. (1995) Will Raising the Incomes of All Increase the Happiness of All? *Journal of Economic Behavior and Organization* 27, 35-48.

Easterlin, R. (2001) Income and Happiness: Towards a Unified Theory. *Economic Journal* 111, 465-484.

Easterlin, R. (2006) Life Cycle Happiness and Its Sources: Intersections of Psychology, Economics, and Demography. *Journal of Economic Psychology* 27, 463-482.

Easterlin, R., L. Angelescu-McVey, M. Switek, O. Sawangfa and J. Smith-Zweig (2010) The Happiness-Income Paradox Revisited. *Proceedings of the National Academy of Sciences* 107, 22463-22468.

Ferrer-i-Carbonell, A. and P. Frijters (2004) How Important Is Methodology for the Estimates of the Determinants of Happiness? *Economic Journal* 114, 641-659.

Fischer, J. (2009) Happiness and Age Cycles - Return to Start.... *MPRA Paper* No. 15249, University of Munich.

Frijters, P. and T. Beatton (2012) The Mystery of the U-shaped Relationship between Happiness and Age. *Journal of Economic Behavior and Organization* 82, 525-542.

Gardner, J. and A. Oswald (2006) Do Divorcing Couples Become Happier by Breaking Up? *Journal of the Royal Statistical Society Series A* 169, 319-336.

Gwozdz, W. and A. Sousa-Poza (2010) Ageing, Health and Life Satisfaction of the Oldest Old: An Analysis for Germany. *Social Indicators Research* 97, 397-417.

Heckman, J. and R. Robb (1985) Using longitudinal data to estimate age, period, and cohort effects in earnings equations. in Mason and Fienberg, eds. *Cohort Analysis in Social Research*, New York: Springer-Verlag.

Heffetz, O. and M. Rabin. (2013) Conclusions Regarding Cross-Group Differences in Happiness Depend on Difficulty of Reaching Respondents. *American Economic Review*, forthcoming.

Herbst, C. (2011)"Paradoxical" Decline? Another Look at the Relative Reduction in Female Happiness. *Journal of Economic Psychology* 32, 773-788. Ifcher, J. and O. Zarghamee (2011) Happiness and Time Preference: The Effect of Positive Affect in a Random-Assignment Experiment. *American Economic Review* 10, 31093129.

Kahneman, d., B. Fredrickson, A. Schreiber and D. Redelmeier (1993) When More Pain Is Preferred to Less: Adding a Better End. *psychological Science* 4, 401-405.

Kahneman, D. and A. Krueger (2006) Developments in the Measurement of Subjective Well-Being. *Journal of Economic Perspectives* 20, 3-24.

Kapteyn, A., R. Alessie and A. Lusardi (2005) Explaining the Wealth Holdings of Different Cohorts: Productivity Growth and Social Security. *European Economic Review* 49, 1361-1391.

Kapteyn, A., J. Smith and A. van Soest (2011) Are Americans Really Less Happy with Their Incomes? *Discussion Paper* 2011-059, Tilburg University, Center for Economic Research.

Kassenboehmer, S. and J. Haisken-Denew (2012) Heresy or Enlightenment? The Well-being Age U-shape Effect Is Flat. *Economics Letters* 117, 235-238.

Keisuke, H., G. Imbens, G. Ridder and D. Rubin (2001) Combining Panel Data Sets with Attrition and Refreshment Samples. *Econometrica* 69, 1645-1659.

Kroh, M. (2011) Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2010). *Data Documentation* 59, Deutsches Institut fur Wirtschaftsforschung.

Krueger, A. and D. Schkade (2008) The Reliability of Subjective Well-being Measures. *Journal of Public Economics* 92, 1833-1845.

Lipps, O. (2007) Attrition in the Swiss Household Panel. *Methoden - Daten - Analysen* 1, 45-68.

Luttmer, E. (2005) Neighbours as Negatives: Relative Earnings and Well-being. *Quarterly Journal of Economics* 120, 963-1002.

McBride, M. (2010) Money, Happiness, and Aspirations: An Experimental Study. *Journal of Economic Behavior and Organization* 74, 262-276.

McKenzie, D. (2006) Disentangling Age, Cohort and Time Effects in the Additive Model. *Oxford Bulletin of Economics and Statistics* 68, 473-495.

Oswald, A. (2008) On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings. *Economics Letters* 100, 369-72.

Oswald, A. and N. Powdthavee (2008a) Death, Happiness, and the Calculation of Compensatory Damages. *Journal of Legal Studies* 37, S217-S252.

Oswald, A. and N. Powdthavee (2008b) Does Happiness Adapt? A Longitudinal Study of Disability with Implications for Economists and Judges. *Journal of Public Economics* 92, 1061-1077.

Oswald, a., E. Proto and D. Sgroi (2012) Happiness and Productivity. *Working Paper*, University of Warwick.

Oswald, A. and S. Wu (2010) Objective Confirmation of Subjective Measures of Human Well-being: Evidence from the U.S.A. *Science* 327, 576-580.

Oswald, A. and S. Wu (2011) Well-being across America. *review of Economics and Statistics* 93, 1118-1134.

Powdthavee, N. (2008) Putting a Price Tag on Friends, Relatives, and Neighbours: Using Surveys of Life Satisfaction to Value Social Relationships. *Journal of Socio-Economics* 37, 1459-1480.

Proto, E., D. Sgroi and A. Oswald (2012) Are Happiness and Productivity Lower among University Students with Newly-Divorced Parents? An Experimental Approach. *Experimental Economics* 15, 1-23.

Ravallion, M. and M. Lokshin (2010) Who Cares About Relative Deprivation? *Journal of Economic Behavior and Organization* 73, 171-185.

Sacks, D., B. Stevenson and J. Wolfers (2010) Subjective Well-being, Income, Economic Development, and Growth. *NBER Working Paper* No. 16441, National Bureau of Economic Research.

Schwarz, N. and G. Clore (1983) Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States. *Journal of Personality and Social Psychology* 45, 513-523.

Senik, C. (2004) When Information Dominates Comparison: Learning From Russian Subjective Panel Data. *Journal of Public Economics* 88, 2099-2123.

Senik, C. (2008a) Ambition and Jealousy: Income Interaction in the "Old" (Europe Versus the "New" Europe and the United States. *Economica* 75, 495-513.

Senik, C. (2008b) Is Man Doomed to Progress? *Journal of Economic Behavior and Organization* 68, 140-152.

Senik, C. (2009) Direct Evidence on Income Comparisons and Their Welfare Effects. *Journal of Economic Behavior and Organization* 72, 408-424.

Sgroi, d., E. Proto, A. Oswald and A. Dobson (2010) Priming and the Reliability of Subjective Well-being Measures. *Warwick Economic Research Papers* No. 935.

Sharpe, J. and D. Gilbert (1998) Effects of Repeated Administration of the Beck Depression Inventory and Other Measures of Negative Mood States. *Personal Individual Differences* 24, 457-463.

Stevenson, B. and J. Wolfers (2008) Economic Growth and Subjective Well-being: Reassessing the Easterlin Paradox. *NBER Working Papers* No. 14282, National Bureau of Economic Research.

Stevenson, B. and J. Wolfers (2009) The Paradox of Declining Female Happiness. *American Economic Journal: Economic Policy* 1, 190-225.

Stevenson, B. and J. Wolfers (2013) Subjective Well-Being and Income: Is There Any Evidence of Satiation? *American Economic Review* 103, 598-604.

Toepoel, V., J. Das and A. van Soest (2009) Relating Question Type to Panel Conditioning: A Comparison between Trained and Fresh Respondents. *Survey Research Methods* 3, 73-80.

Van Landeghem, B. (2012a) A Test for the Convexity of Human Well-being over the Life Cycle: Evidence from a 20-Year Panel. *Journal of Economic Behavior and Organization* 81, 571-582.

Van Landeghem, B. (2012b) Panel Conditioning and Self-Reported Satisfaction: Evidence from International Panel Data and Repeated Cross-Sections. *SOEPpapers on Multidisciplinary Panel Data Research* 484, DIW Berlin, The German Socio-Economic Panel (SOEP).

Voorpostel, M., R. Tillmann, F., Lebert, B. Weaver, U. Kuhn, O., Lipps, V.-A. Ryser, F. Schmid, M. Rothenbuehler and B. Wernli (2011) *Swiss Household Panel Userguide (1999-2010), Wave 12, October 2011.* Lausanne: FORS.

Wagner, G., J. Frick and J. Schupp (2007) The German Socio-economic Panel Study (SOEP) - Scope, Evolution, and Enhancements. *Schmollers Jahrbuch* 127, 139-169.

Wolfers, J. (2003) Is Business Cycle Volatility Costly? Evidence from Surveys of Subjective Well-being. *International Finance* 6, 1-26.

Wooden, M. and N. Li (2013) Panel Conditioning and Subjective Well-being. *Social Indicators Research*, Forthcoming.

Zwane, A., J. Zinman, E. Van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. Karlan, R. Hornbeck, X. giné, E. Duflo, F. Devoto, B. Crepon and A. Banerjee (2011) Being Surveyed Can Change Later Behavior and Related Parameter Estimates. *Proceedings of the National Academy of Sciences* 108, 1821-1826.

Table 1: Exploratory Regressions: Conditional Differences in Subjective Well-Being Between New Respondents and More Experienced Respondents

	West Germany			East Germany			Switzerland
new 1998	0.505***			0.525***			
	(0.051)			(0.090)			
new 2000		0.339***			0.399***		
		(0.031)			(0.051)		
new 2004							0.184***
							(0.035)
new 2006			0.491***			0.742***	
			(0.041)			(0.075)	
married	0.372***	0.321***	0.355***	0.339***	0.306***	0.407***	0.412***
	(0.064)	(0.046)	(0.050)	(0.101)	(0.084)	(0.090)	(0.057)
widowed	0.113	0.088	0.139	0.264	0.346**	0.270*	-0.006
	(0.119)	(0.084)	(0.088)	(0.173)	(0.140)	(0.142)	(0.121)
divorced	-0.173	-0.130*	-0.171**	-0.276*	-0.024	-0.011	-0.390***
	(0.106)	(0.073)	(0.077)	(0.162)	(0.118)	(0.125)	(0.087)
education = low	-0.330***	-0.523***	-0.549***	-0.118	-0.463***	-0.391***	-0.086
	(0.062)	(0.046)	(0.051)	(0.112)	(0.081)	(0.108)	(0.053)
education = medium	-0.124***	-0.163***	-0.279***	-0.181**	-0.192***	-0.137**	0.047
	(0.048)	(0.033)	(0.034)	(0.078)	(0.058)	(0.057)	(0.054)
natural refresher	0.006	0.038	-0.010	0.123	0.328***	0.131*	0.189
	(0.060)	(0.045)	(0.042)	(0.090)	(0.076)	(0.070)	(0.142)
Constant	7.416***	6.613***	7.339***	6.373***	5.751***	5.288***	8.581***
	(0.414)	(0.330)	(0.294)	(0.817)	(0.514)	(0.609)	(0.338)
Observations	7,836	15,884	13,313	3,838	5,810	5,045	7,599
R-squared	0.072	0.053	0.070	0.119	0.095	0.123	0.063

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: German Socioeconomic Panel and Swiss Household Panel.
All regressions include age and labour force status dummies.

Table 2: Exploratory Regressions: Conditional Differences in Subjective Well-Being Between New Respondents and More Experienced Respondents: Male Subsample

	West Germany			East Germany			Switzerland
new 1998	0.477***			0.496***			
	(0.071)			(0.122)			
new 2000		0.315***			0.449***		
		(0.044)			(0.072)		
new 2004							0.140***
							(0.052)
new 2006			0.445***			0.761***	
			(0.058)			(0.109)	
married	0.282***	0.237***	0.341***	0.415***	0.215*	0.378***	0.281***
	(0.089)	(0.063)	(0.072)	(0.147)	(0.118)	(0.122)	(0.081)
widowed	0.343	-0.296*	-0.016	0.541*	0.257	0.124	-0.023
	(0.239)	(0.173)	(0.162)	(0.318)	(0.251)	(0.246)	(0.269)
divorced	-0.269	-0.022	-0.118	-0.400	-0.048	0.089	-0.473***
	(0.166)	(0.102)	(0.115)	(0.258)	(0.181)	(0.175)	(0.138)
education = low	-0.430***	-0.560***		0.156	-0.396***	-0.417***	-0.196***
	(0.098)	(0.071)		(0.194)	(0.124)	(0.157)	(0.070)
education = medium	-0.135**	-0.236***		-0.040	-0.223***	-0.209***	-0.023
	(0.065)	(0.045)		(0.119)	(0.086)	(0.080)	(0.069)
natural refresher	0.002	0.004	0.041	0.061	0.378***	0.162	0.492**
	(0.085)	(0.064)	(0.060)	(0.121)	(0.095)	(0.099)	(0.192)
Constant	6.700***	6.043***	6.855***	6.792***	6.291***	6.744***	8.657***
	(0.683)	(0.557)	(0.425)	(0.449)	(0.868)	(0.466)	(0.326)
Observations	3,740	7,620	6,295	1,859	2,787	2,424	3,380
R-squared	0.103	0.080	0.083	0.141	0.107	0.160	0.075

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: German Socioeconomic Panel and Swiss Household Panel. All regressions include age and labour force status dummies.

Table 3: Exploratory Regressions: Conditional Differences in Subjective Well-Being Between New Respondents and More Experienced Respondents: Female Subsample

	West Germany			East Germany			Switzerland
new 1998	0.560***			0.575***			
	(0.073)			(0.135)			
new 2000		0.367***			0.295***		
		(0.044)			(0.069)		
new 2004							0.227***
							(0.049)
new 2006			0.541***			0.675***	
			(0.057)			(0.104)	
married	0.408***	0.332***	0.311***	0.209	0.352***	0.403***	0.510***
	(0.093)	(0.069)	(0.072)	(0.148)	(0.126)	(0.133)	(0.079)
widowed	0.014	0.152	0.117	0.111	0.398**	0.269	-0.010
	(0.146)	(0.104)	(0.112)	(0.216)	(0.182)	(0.186)	(0.144)
divorced	-0.140	-0.253**	-0.265**	-0.283	0.024	-0.072	-0.340***
	(0.141)	(0.103)	(0.104)	(0.216)	(0.163)	(0.180)	(0.114)
education = low	-0.316***	-0.513***	-0.534***	-0.241*	-0.505***		-0.048
	(0.084)	(0.063)	(0.069)	(0.141)	(0.111)		(0.086)
education = medium	-0.116	-0.096*	-0.237***	-0.250**	-0.159**		0.102
	(0.072)	(0.051)	(0.053)	(0.104)	(0.081)		(0.088)
natural refresher	0.028	0.088	-0.055	0.225			-0.102
	(0.085)	(0.064)	(0.060)	(0.143)			(0.205)
Constant	7.754***	6.688***	7.351***	6.013***	5.265***	4.110***	8.699***
	(0.531)	(0.402)	(0.409)	(1.225)	(0.562)	(0.859)	(0.524)
Observations	4,096	8,264	7,018	1,979	3,023	2,621	4,219
R-squared	0.072	0.051	0.065	0.132	0.108	0.118	0.079

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: German Socioeconomic Panel and Swiss Household Panel. All regressions include age and labour force status dummies.

Table 4: Exploring the Determinants of the Propensity to Be a Refresher: Results for West Germany, East Germany, and Switzerland

	1998		2000		2006		2004
	West	East	West	East	West	East	Switzerland
log real hh. income	-0.043***	-0.027**	-0.094***	-0.011	-0.002	-0.010	-0.051***
	(0.012)	(0.011)	(0.012)	(0.019)	(0.011)	(0.011)	(0.012)
disabled	-0.056***	0.016	-0.070***	-0.005	-0.008	-0.014	0.020
	(0.015)	(0.018)	(0.017)	(0.026)	(0.012)	(0.015)	(0.013)
household size	-0.000	-0.015**	0.033***	0.006	0.009*	-0.008	-0.030***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.005)	(0.006)	(0.006)
low education	-0.008	0.167***	-0.074***	0.094***	-0.033*	0.007	0.053***
	(0.017)	(0.040)	(0.018)	(0.029)	(0.019)	(0.027)	(0.020)
medium education	-0.046***	0.055***	0.017	0.234***	-0.056***	-0.050***	0.006
	(0.015)	(0.011)	(0.014)	(0.017)	(0.014)	(0.017)	(0.021)
unemployed	-0.066***	0.002	-0.018	-0.016	0.006	0.006	0.043
	(0.024)	(0.013)	(0.032)	(0.022)	(0.020)	(0.015)	(0.043)
single	0.013	0.040***	0.021	-0.062***	-0.017	0.027*	-0.047**
	(0.017)	(0.013)	(0.017)	(0.024)	(0.015)	(0.015)	(0.019)
widowed	-0.045**	-0.037***	-0.046**	-0.020	0.033	-0.004	-0.012
	(0.019)	(0.013)	(0.022)	(0.031)	(0.021)	(0.021)	(0.032)
divorced	-0.043**	-0.008	-0.047**	0.048*	-0.002	-0.012	-0.037*
	(0.020)	(0.015)	(0.022)	(0.027)	(0.016)	(0.015)	(0.022)
female	-0.019*	-0.003	-0.004	0.024*	-0.023**	0.001	-0.040***
	(0.012)	(0.009)	(0.011)	(0.014)	(0.011)	(0.011)	(0.012)
Observations	5,049	3,569	9,520	5,169	8,439	4,581	6,707

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The table shows marginal probabilities, computed at the average value of independent variables, derived from probit equations. Regressions are run for each survey year in which a refreshment sample was started. The dependent variable is dichotomous and takes the value 1 whenever an individual belongs to the refreshment sample started in that particular year.

Table 5: Differences in Real Household Income between Refreshers and More Experienced Respondents (Expressed in %)

	West Germany	East Germany	Switzerland
$E[Y^{re}_{1998}] - E[Y^{exp}_{1998}]$	-9.6***	-16.7***	
$E[Y^{re}_{2000}] - E[Y^{exp}_{2000}]$	-6.8***	-6.5***	
$E[Y^{re}_{2004}] - E[Y^{exp}_{2004}]$			-6.2***
$E[Y^{re}_{2006}] - E[Y^{exp}_{2006}]$	-0.2	-0.4	

Source: German Socioeconomic Panel and Swiss Household Panel.
Y = real household income, subscripts denote the survey year, superscripts RE and EXP denote 'refreshment sample' and 'experienced sample', respectively. One to three asterisks denote significance at the 10%, 5%, and 1% significance levels, respectively.

Table 6: Conditional Differences in Subjective Well-Being Between New Respondents and More Experienced Respondents: Taking Into Account Attrition Biases

	West Germany			East Germany			Switzerland
new 1998	0.595***			0.636***			
	(0.059)			(0.107)			
new 2000		0.421***					
		(0.034)					
new 2004							0.248***
							(0.061)
new 2006			0.567***			0.760***	
			(0.052)			(0.092)	
married	0.356***	0.349***	0.322***	0.337***	0.263***	0.349***	0.498***
	(0.066)	(0.051)	(0.051)	(0.102)	(0.089)	(0.089)	(0.079)
widowed	0.057	0.123	0.054	0.218	0.213	0.249*	-0.019
	(0.112)	(0.084)	(0.084)	(0.162)	(0.139)	(0.135)	(0.144)
divorced	-0.175*	-0.079	-0.222***	-0.295**	-0.104	-0.060	-0.349***
	(0.100)	(0.076)	(0.074)	(0.146)	(0.120)	(0.118)	(0.114)
education = low	-0.356***	-0.532***				-0.422***	-0.034
	(0.064)	(0.049)				(0.106)	(0.085)
education = medium	-0.155***	-0.171***				-0.155**	0.106
	(0.053)	(0.039)				(0.061)	(0.088)
natural refresher	0.005	0.014	-0.010	0.154	0.348***	0.126*	-0.057
	(0.062)	(0.049)	(0.044)	(0.103)	(0.085)	(0.075)	(0.203)
Constant	7.245***	6.773***	6.910***	6.224***	5.032***	5.223***	8.753***
	(0.445)	(0.404)	(0.283)	(1.027)	(0.889)	(0.523)	(0.523)
Observations	7,241	12,473	12,391	3,711	5,141	4,838	4,219
R-squared	0.073	0.061	0.062	0.118	0.097	0.118	0.078

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: German Socioeconomic Panel and Swiss Household Panel. All regressions include age and labour force status dummies. The newcomers are restricted to those who will stay in the panel during at least four consecutive rounds.

Table 7: Conditional Differences in Subjective Well-Being Between New Respondents and Respondents Stemming from the Previous Substantial Refreshment: Results for West and East Germany

	Exploratory Analysis				Correcting for Attrition			
	West		East		West		East	
new 2000	0.021		0.071		0.178***		0.136*	
	(0.052)		(0.105)		(0.035)		(0.074)	
new 2006		0.410***		0.571***		0.418***		0.503***
		(0.044)		(0.085)		(0.054)		(0.098)
married	0.351***	0.385***	0.498***	0.601***	0.346***	0.391***	0.495***	0.627***
	(0.059)	(0.067)	(0.137)	(0.141)	(0.059)	(0.067)	(0.137)	(0.142)
widowed	0.164*	0.249**	0.523**	0.755***	0.157	0.257**	0.516**	0.782***
	(0.098)	(0.109)	(0.206)	(0.211)	(0.097)	(0.109)	(0.206)	(0.212)
divorced	-0.087	-0.183*	0.278	0.328*	-0.095	-0.181*	0.275	0.334*
	(0.091)	(0.096)	(0.174)	(0.182)	(0.091)	(0.096)	(0.174)	(0.182)
educ_low	-0.489***	-0.569***	-0.510***	-0.604***	-0.478***	-0.563***	-0.509***	-0.592***
	(0.056)	(0.066)	(0.137)	(0.152)	(0.056)	(0.066)	(0.137)	(0.153)
educ_med	-0.162***	-0.288***	-0.141	-0.251**	-0.166***	-0.288***	-0.158	-0.273**
	(0.045)	(0.049)	(0.110)	(0.106)	(0.045)	(0.049)	(0.110)	(0.106)
natural refresher	-0.054	-0.191	0.138	-0.071	-0.032	-0.200	0.136	-0.034
	(0.104)	(0.138)	(0.151)	(0.192)	(0.102)	(0.139)	(0.148)	(0.193)
Constant	6.883***	7.165***	5.519***	5.953***	6.892***	7.219***	5.584***	6.087***
	(0.371)	(0.368)	(0.780)	(0.786)	(0.368)	(0.368)	(0.775)	(0.790)
Observations	10,247	7,554	2,621	2,129	10,247	7,554	2,621	2,129
R-squared	0.044	0.070	0.097	0.165	0.047	0.066	0.099	0.157

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: German Socioeconomic Panel All regressions include age and labour force status dummies. The newcomers are restricted to those who will stay in the panel during at least four consecutive rounds.

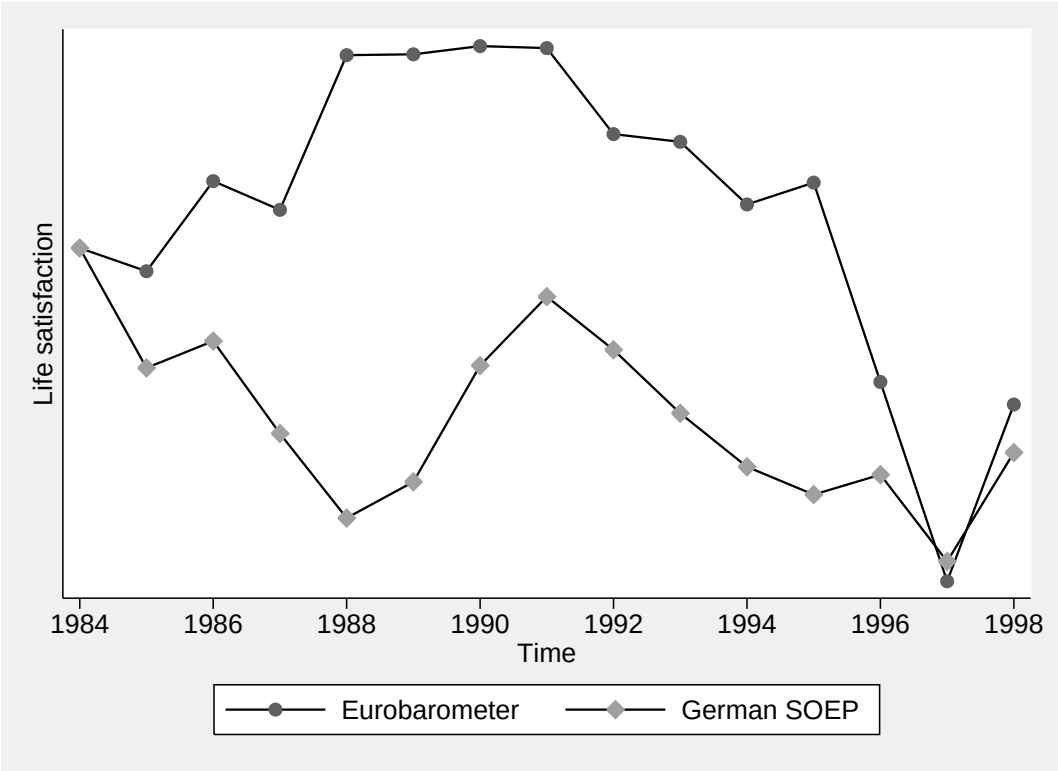
Table 8: The Path of Panel Effects over Interviews: Results for West Germany, East Germany, and Switzerland

	West Germany	East Germany	Switzerland
1st interview	0.488***	0.463***	0.231***
	(0.023)	(0.040)	(0.059)
2nd interview	0.379***	0.345***	0.105**
	(0.022)	(0.037)	(0.043)
3rd interview	0.288***	0.317***	0.072*
	(0.021)	(0.035)	(0.039)
4th interview	0.205***	0.261***	0.061*
	(0.020)	(0.033)	(0.036)
5th interview	0.181***	0.241***	0.050
	(0.020)	(0.032)	(0.032)
Constant	9.415***	8.177***	10.116***
	(0.177)	(0.301)	(0.277)
Observations	211,847	74,727	50,206
R-squared	0.047	0.075	0.050

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

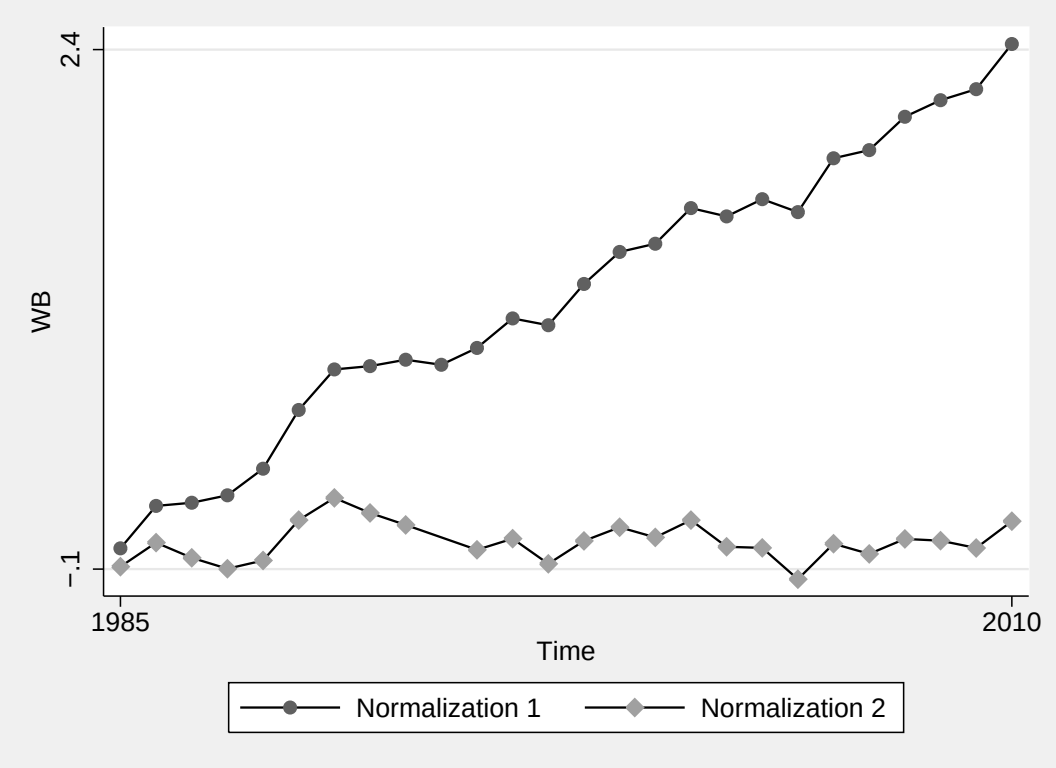
Source: German Socioeconomic Panel and Swiss Household Panel.
Regressions are run on individuals who have answered at least the subjective questions in the four survey rounds following the interview in which they were first asked the subjective well-being question. All regressions include marital status dummies, labour force status dummies, education dummies, time dummies, a dummy for having entered the panel as natural refresher, and a third-order age polynomial as controls.

Figure 1: Comparison between Trends in Standardized Life Satisfaction in the German SOEP and the Eurobarometer Survey: West Germany



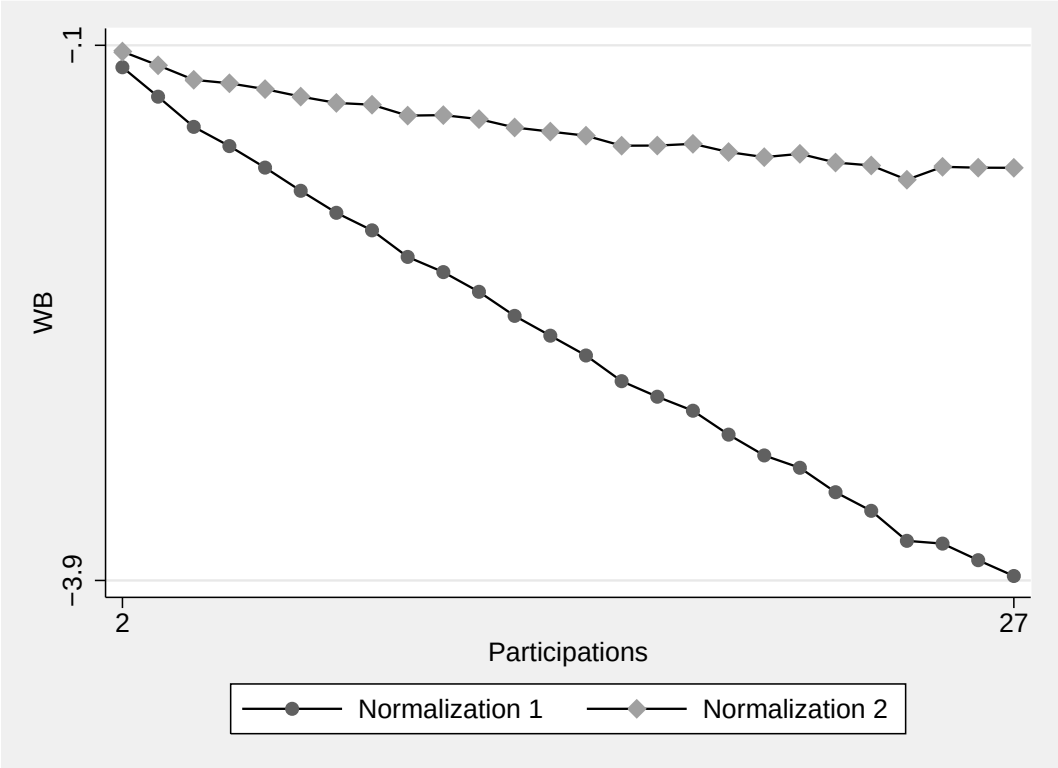
Source: German Socioeconomic Panel and Eurobarometer Survey.
For both data sources, life satisfaction has been standardized with mean 0 and standard error 1. Hence, the graphs are only meant for illustrative purposes, and they should not be used to quantify the differences in trends between the datasets.

Figure 2: The Impact of Different Normalization Assumptions on the Estimated Path of Well-Being over Time: Results from Two Extended Age-Time-Cohort Models



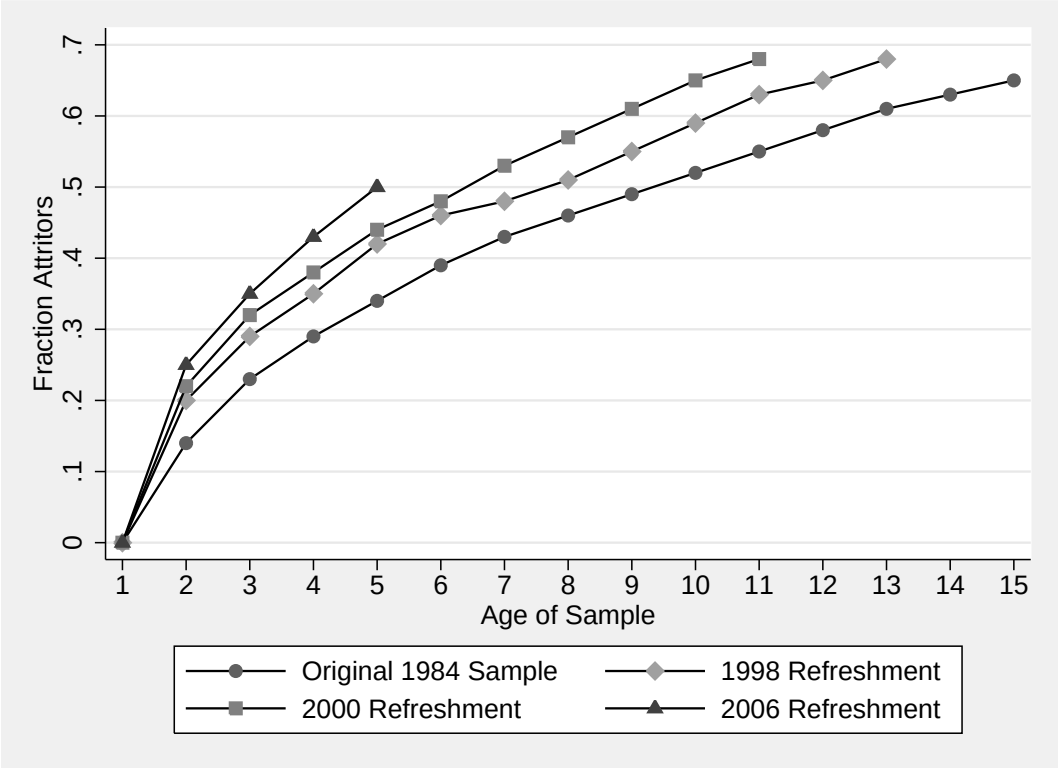
Source: German Socioeconomic Panel

Figure 3: The Impact of Different Normalization Assumptions on the Estimated Path of Well-Being over the Number of Survey Participations: Results from Two Extended Age-Time-Cohort Models



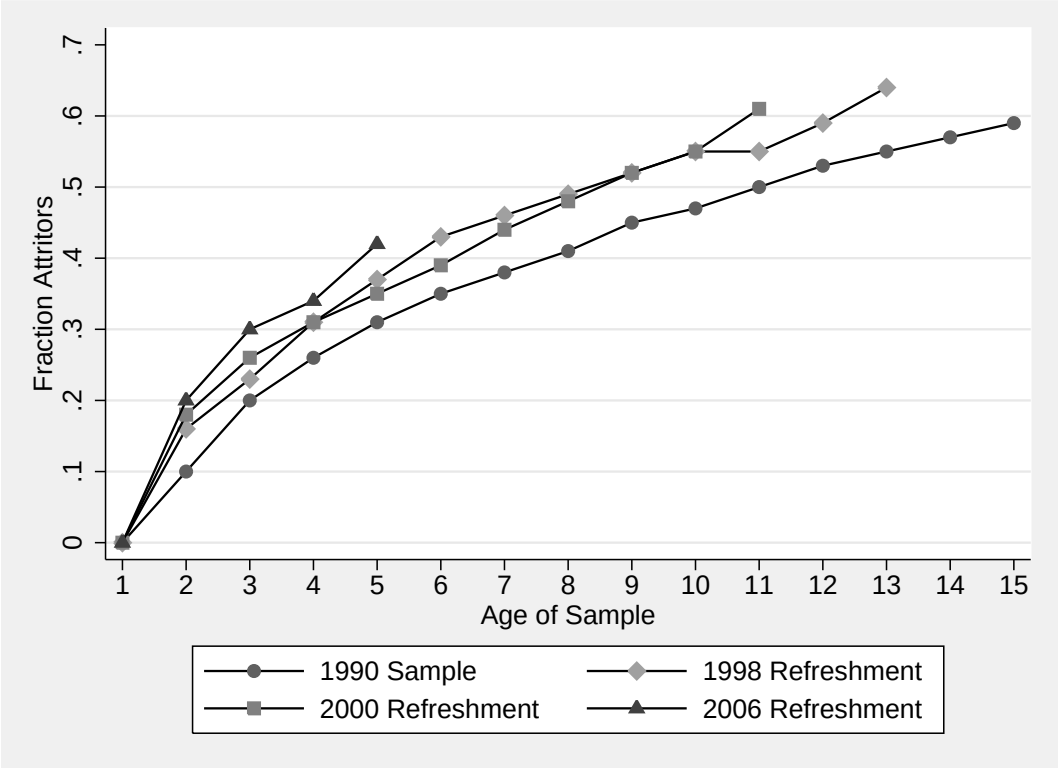
Source: German Socioeconomic Panel

Figure 4: The Attrition Rate over a Sample's Life Cycle: West Germany



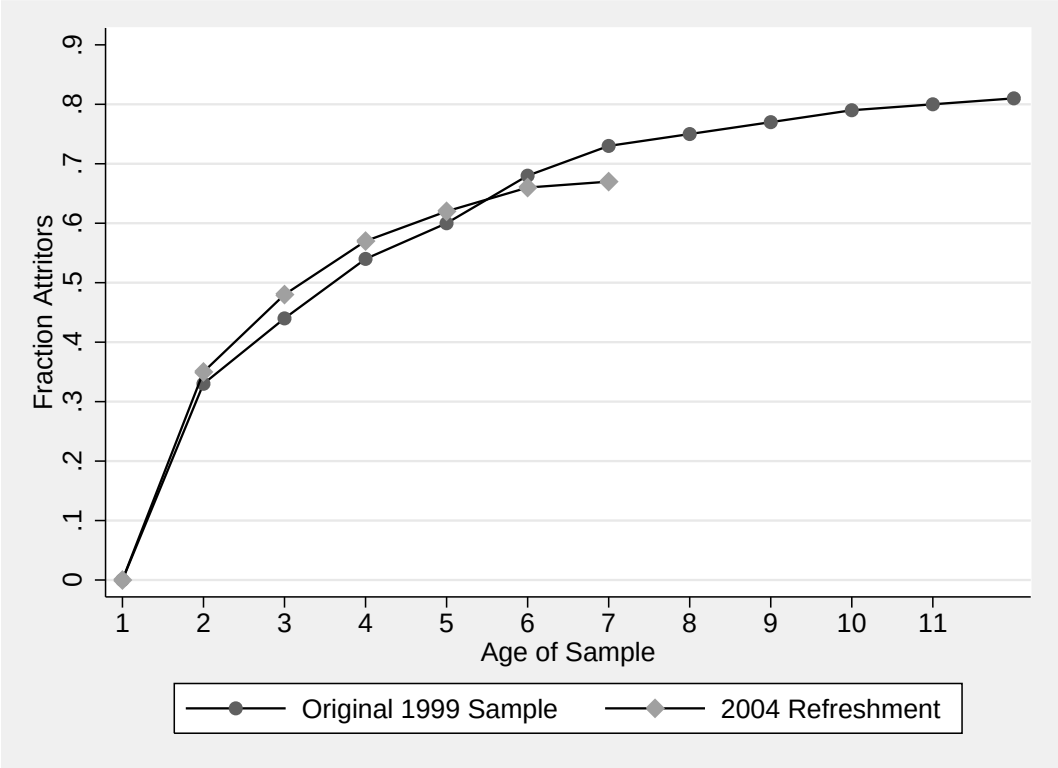
Source: German Socioeconomic Panel.
An individual is regarded as an attritor at age n of the panel if the individual was a respondent at age one and if the number of times the individual responded to a subjective well-being question is strictly less than n.

Figure 5: The Attrition Rate over a Sample's Life Cycle: East Germany



Source: German Socioeconomic Panel.
An individual is regarded as an attritor at age n of the panel if the individual was a respondent at age one and if the number of times the individual responded to a subjective well-being question is strictly less than n.

Figure 6: The Attrition Rate over a Sample's Life Cycle: Switzerland



Source: Swiss Household Panel.
An individual is regarded as an attritor at age n of the panel if the individual was a respondent at age one and if the number of times the individual responded to a subjective well-being question is strictly less than n.