Which Decision Theory Describes Life Satisfaction Best? Evidence From Annual Panel Data

Fırat Yaman* Patricia Cubí-Mollá† Sergiu Ungureanu‡

Abstract

We use an annual household panel to test which features of prospect theory can be supported by measures of life satisfaction. We also test whether recalled or expected life satisfaction is anchored at current life satisfaction and adjusted in the direction of the recall or expectation. Using a fixed effects estimator we find that life satisfaction contains features of both classic expected utility and prospect theory. Life satisfaction depends positively on levels of income, good health, and on employment. It also depends positively on income and employment improvements, however the reverse is true for health increases. Life satisfaction is concave in income gains and convex in income losses, and it exhibits loss aversion in income and employment status, but not in health. Moreover, we find that current levels of life satisfaction are better predictors of recalled (expected) life satisfaction than past (future) life satisfaction. The results support viewing life satisfaction as representing a mixture of the classic decision utility of expected utility theory, and the value function of prospect theory. Subjects seem to use an anchoring and adjustment heuristic when answering questions about past and expected life satisfaction.

JEL classification: I31, D81

Keywords: life satisfaction, prospect theory, loss aversion, recall bias, anchoring and adjustment

*Department of Economics; City, University of London. Email: Firat.Yaman.1@city.ac.uk.

†Office of Health Economics. Email: Pcubi-molla@ohe.org.

‡Department of Economics; City, University of London. Email: Sergiu.Ungureanu.1@city.ac.uk.

1 Introduction

In this paper we investigate which decision theory best explains life satisfaction (henceforth LS) as reported in an annual household panel. We consider the two most prominent decision theories in economics: expected utility theory (EUT), and prospect theory (PT). We analyse their predictions by looking in the German Socio-Economic Panel (SOEP) at three outcomes which have consistently been found to improve LS: income, health, and employment (compared to unemployment).

We consider the following properties of PT: First, whether individuals evaluate their LS against a reference point. Specifically, we test whether changes in income, health and employment can help explain LS levels. An affirmative answer would be consistent with individuals using their past income, health and employment as a reference point in evaluating their current LS. Second, whether individuals exhibit loss aversion in income, health and employment, that is, whether the effect of a decrease in one of those variables reduces LS more than an equivalent increase improves it. Finally — and only for the case of income — whether individuals exhibit the reflection effect, that is, we test whether LS is concave in income gains and convex in income losses. In PT, these three features — a reference point, loss aversion and the reflection effect — are identified as the distinguishing elements of the value function. PT is a theory of "decision under risk" but these properties are relevant even in choices not involving chance (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991).[1]

In addition, we also look for evidence for the use of the anchoring and adjustment (AA) heuristic in the individuals' evaluation of past and expected LS. This test is motivated by the psychological literature showing that individuals' perceptions of past feelings or experiences are affected by conditions in the present (O'Brien et al., 2012; Wilson et al., 2003; Gilbert, 2006).

Our empirical model allows for absolute effects, relevant to EUT, and relative effects, relevant to PT. We find that LS is best described by an empirical model that incorporates features of both EUT and PT. In line with the literature on subjective well-being, we find that LS is increasing in income levels and in good health. Being employed implies higher LS than being unemployed, and LS is concave in income. In addi-

[1] We do not test for the presence of probability weighting in this paper, another important assumption in PT, because our data do not provide a good source of identification of probability weighting. Probability weighting appears in the original article (Kahneman and Tversky, 1979) and in many if not all subsequent versions of PT.

tion, many — but not all — of the features of PT are also supported. The marginal effect of income changes is asymmetric and supports loss aversion. The positive effect of income gains on LS is non-significant, but the negative effect of losses is highly significant. Similarly, LS concavity in income gains is not significant, but LS convexity in income losses is strong and highly significant — consistent with the reflection effect of PT. As expected, losing a job decreases LS while finding a job increases it. Maybe our most striking finding is that LS is *decreasing* in health improvements and *increasing* in health losses controlling for health status. We also find a significant loss aversion effect in employment changes. The only domain for which we do not find evidence of loss aversion is health — instead we find that individuals have a loss preference. Finally, we find strong evidence that respondents use the AA heuristic when trying to determine past and future LS.

Our paper is related to two areas of research in the literature: the (quasi-)experimental research on PT and related decision theories, and the LS research, which mainly builds on survey-type data. What sets our paper apart is that we estimate a model which can nest elements of both EUT and PT in one empirical model. In contrast, the literature has typically focused on one aspect of a decision theory and sought to find evidence for or reject it. For example, Boyce et al. (2013) test for loss aversion in income changes, but do not include income levels as an explanatory variable. While they support the notion of loss aversion, it is not clear whether income changes contribute more to life satisfaction than income levels, or indeed whether loss aversion would still hold once income levels are controlled for. A further contribution is that we consider life domains other than income. Since health is a good, and employment is by definition preferred to unemployment, we should expect life satisfaction to display characteristics of a value function with respect to those goods, too.

Section 2 describes related literature and the contribution of our work. The rest of the paper is organised as follows. Section 3 describes the dataset, section 4 presents the main econometric model and states the hypotheses tested and section 5 analyses the results. Section 6 has the discussion, section 7 contains extensions and alternative specifications, and finally section 8 concludes.

2 Literature Review

Since the publication of "Prospect theory: An analysis of decision under risk" by Kahneman and Tversky (1979), non-expected utility theory has developed into a very active literature. Many authors have proposed alternatives or extensions to the classical expected-utility model,[2] and another strand of the literature has tested the assumptions of, and hypotheses derived from, PT and its subsequent extensions, by means of laboratory or field experiments.[3] The strongest support for PT, and other reference-dependent value models, comes from such laboratory and field experiments (Abdellaoui et al., 2007; List, 2004).

In contrast, the literature on LS and subjective well-being in general has relied heavily on survey data. Dolan et al. (2008) and Clark (2018) provide good surveys of the literature. Researchers on subjective well-being have long advocated to use LS as a measure to be taken into account for policy decisions equating it to utility explicitly (Layard et al., 2008), or implicitly when "pricing" intangible or non-marketed goods (Levinson, 2012). LS has been consistently found to correlate with variables such as income, good health, and other socio-economic variables in the expected directions, and these correlations are relied on as justifications for using LS as a proxy for utility.

Methodologically, we follow the LS literature in using survey data. Experimental methods have the important advantage that confounding factors can be controlled. However, lab results come with their own set of limitations: lab experiments use weak incentives, lack realism, and there are questions regarding the external validity of the results (Harrison and List, 2004).[4] Experiments are also restricted to short time spans between the presentation of the experiment and the observation of the participant's response, and to the participant's evaluation of a controlled experience. But decision theories and PT in particular have intuitive appeal when considering longer time spans as well. By contrast, the advantage of using large representative datasets to verify and calibrate models is that the datasets directly address policy relevant measures, like income or LS. Therefore, there are fewer concerns regarding external validity. The disadvantage is that the analysis requires greater care in handling confounders and alternative explanations. While not an ideal

[2] There are many examples. Kőszegi and Rabin (2006) propose a model with reference-dependent utility where the reference-point is defined to be expected consumption shortly before consumption occurs. Bénabou (2012), Brunnermeier and Parker (2005), Gollier (2011), and Gottlieb (2014) all build models in which agents can - to some extent - choose what to believe in order to enjoy anticipatory utility.

[3] See DellaVigna (2009) for an overview of this burgeoning literature.

[4] For example, experimental measures of loss aversion are greatly influenced by context (Novemsky and Kahneman, 2005), so the applicability of lab measures of loss aversion outside of the lab is under question.

solution, we mitigate these concerns by allowing for individual fixed effects in our panel data estimation, and by carrying out a number of extensions.

The work closest to ours is Boyce et al. (2013), Di Tella et al. (2010), Vendrik and Woltjer (2007), Ferrer-i Carbonell (2005), who also use the SOEP, Fang and Niimi (2017), who use Japanese panel data, and De Neve et al. (2018), who use large nationally representative surveys. The listed papers however differ in important respects from ours, chiefly because of fundamentally different modelling choices.

Boyce et al. (2013) consider the effect of income changes on well-being and find evidence for loss aversion. They rely on the SOEP and the British Household Panel Survey. Their paper is closest in spirit to ours. While informative, the empirical model used is, we believe, improper as a test of discriminating between PT and EUT, because it does not map to the value functions of either of those theories. In particular, they use previous-period LS as a control variable (with a large, highly significant effect). This makes it impossible to identify LS with any value function known in decision theory. EUT as an explanation for LS is ruled out ex-ante, as income levels are not included as an explanatory variable. Another concern is the unexplained use of a dummy variable for whether income decreased or not. This dummy has a large and significant effect, and makes the estimated LS as a function of income changes glaringly discontinuous — a feature which is not supported by any theory of utility that we are aware of.

Di Tella et al. (2010) are mainly concerned with the long-run adaptation to income and status of individuals. This work is best placed in the strand of literature concerned with finding an explanation to the Easterlin paradox (Easterlin, 1974). They use the SOEP like us, and find that there is adaptation to income but not to status. The results are reversed for sub-samples defined as "politically right-leaning", as "men", and as "self-employed", compared to the complementary sub-samples. In a section of the paper, they consider a model with changes in income as independent variables. They also allow asymmetric effects from gains and losses — loss aversion — but these gains and losses are from the present to the *next* period (this also assumes that the future income is closely related to anticipated income). The model will only accommodate a symmetric effect of the change in income from the previous period. Then the asymmetric effect on LS comes from the asymmetric anticipation of future utility. By contrast, our choice of reference point can be interpreted as looking for the asymmetric effect of already-realised income changes on currently experienced utility. In addition, we also allow for diminishing marginal effects and look for differences in curvature over gain and losses.

Fang and Niimi (2017) look for loss aversion in a Japanese household panel. Their contribution is the application to a new dataset and the use of quantile regression techniques, which allows them to look for distributional effects. For loss aversion in income, they select to use a very similar specification to the one of Di Tella et al. (2010), discussed above. A similar discussion regarding the reference point can be made, but they enjoy an advantage over Di Tella et al. (2010) in that they have data for expected future income changes. This means they do not have to proxy expected future income by future income in the panel, which brings their empirical reference point closer to its theoretical counterpart. Like in the other study, they do not consider possible diminishing marginal effects of income changes. The quantile regression approach shows that the loss aversion comes from the bottom quantiles, and no effects can be seen at the mean.

Vendrik and Woltjer (2007) have the same goal as us of testing the predictions of PT on the SOEP, but make a radically different choice for the reference point, namely the income of the social reference group of an individual. This group is defined from the panel to be a cell of individuals with similar education, similar age, from the same region and of the same gender. This specification has advantages and disadvantages over ours. Since PT has no clearly defined theory of reference point formation, one can see these results as being complementary to ours. With this choice of reference point, they are able to use a general specification to look for all features of PT, like loss aversion and asymmetries in the diminishing marginal effects of relative income. In addition, they look carefully at the robustness of the concavity results to distortions of the LS scale. One result is significant concavity for both positive and negative relative income, so no convexity over losses as in PT. The concavity is stronger for losses, so not easy to dismiss, and robust to distortions in the LS scale. Their proposed explanation for loss concavity is that there are two competing effects: there are declining marginal effects of social comparison, but stronger increasing marginal effects of "social participation", which dominate.

In a similar vein, Ferrer-i Carbonell (2005) also looks at the importance of "comparison income" — the income relative to a social group — with an ordered probit model. She concludes that the income relative to the reference group is about as important as the absolute income level for individual happiness. Restricting

Table 1: Summary of most relevant literature

Study	Variables	Features	Reference point
Boyce et al. (2013)	Household income	Loss aversion	Past self
Di Tella et al. (2010)	Household income	Loss aversion	Future (realised) self
Fang and Niimi (2017)	Household income, Living standards	Loss aversion	Future (expected) self, Social peers
Vendrik and Woltjer (2007)	Household income	Loss aversion, Reflection effect	Social peers
Ferrer-i Carbonell (2005)	Household income	Loss aversion	Social peers
De Neve et al. (2018)	National income	Loss aversion	Past national income
Yaman, Cubí-Mollá, and Ungureanu (2019)	Household income, Health, Employment	Loss aversion, Reflection effect	Past self

the analysis to West Germany, she can also find support for loss aversion. McBride (2001) also considers comparison income with ordered probit techniques, but does not allow for asymmetric effects of positive and negative differences. He finds stronger effects at lower level of income, but does not consider features for prospect theory.

Using large survey datasets, De Neve et al. (2018) find that measures of subjective well-being are linked to growth, but individuals are more than twice as sensitive to negative as compared to positive economic growth — a sign of loss aversion. This has implications for the long term effect of economic growth with volatility, and is one of the proposed explanations to the Easterlin paradox. They use the Gallup World Poll for 150 countries, Behavioral Risk Factor Surveillance System data on the US, and Eurobarometer data. They test the fit of a piecewise linear relationship between growth and SWB. The effect of growth changes on subjective well-being subsumes the effects of income changes, unemployment changes, and perhaps inflation and inequality. When the authors introduce controls for these other macro variables, it weakens their main result.

Table 1 summarises the most relevant literature and highlights our contribution. While most papers have considered loss aversion (and by extension reference points), only Vendrik and Woltjer (2007) is a comprehensive test of PT on LS. Our paper differs from theirs in two important respects: We consider income, but also health and employment, and we use individuals' past values of those variables as reference points, instead of social peer groups.

More widely speaking, there is an active literature on the Easterlin paradox — the apparent contradiction between the strong short-term association of GDP and subjective well-being and the lack of long-term association seen in developed countries. Proto and Rustichini (2013), Di Tella et al. (2010), Fang and Niimi (2017), Stevenson and Wolfers (2013), Clark et al. (2008b) can be seen as contributions on this topic. Sacks et al. (2012) also discuss the effect of absolute and relative income on subjective well-being, in a comparison across countries. De Neve and Oswald (2012) look for the inverse effect of LS on later income in a US panel, and Binder and Coad (2015) look for the inverse effect of LS on later unemployment or mental well-being measures in the British Household Panel Survey. Using the SOEP, Lucas et al. (2004) look for short and long-term effects on subjective well-being, Clark et al. (2008a) look for evidence of habituation after life and labour market events, Frijters et al. (2004) find a strong long-term effect of income growth on LS in East Germany after reunification, and Clark et al. (2016) find that there is no adaptation to poverty in terms of LS. Luttmer (2005) uses information on the earnings of neighbours to see if there are any comparison income effects.

An independent contribution of this paper is finding evidence for the use of the AA heuristic. AA was originally proposed in Tversky and Kahneman (1974) as a descriptive model of how answers to difficult questions are generated. In the model, an initial starting value serves as the *anchor*, and this value is improved (*adjusted*) until it becomes a satisfactory answer. The anchor is often the answer to an easier question, a hint suggested by the question itself, or even unrelated information in the question. The model lacks a specific description of the cost benefit analysis of the search process, but its qualitative features are very well supported by evidence. Epley and Gilovich (2006) provide an up-to-date discussion, and Odermatt and Stutzer (2018) discuss anchoring and other heuristics in the context of the SOEP and individuals' prediction of future LS after life events. In the analysis of the dataset, it was readily apparent that the current LS values can be interpreted as anchors for the subsequent questions of what future expected and past recalled LS were. To further test this theory, we looked for evidence of the adjustment process, relative to the anchor. The results are that current LS is a much better predictor of recalled LS than past LS, and a much better predictor of expected LS than actual future LS — suggesting that the current LS value is the anchor used in determining the answers. Moreover, the true values of future and past LS go in the same direction as the predicted values relative to the anchor value, which is evidence for adjustment. This implies that views of the past and expectations of the future are considerably influenced by present levels of LS.

In short, our novel contribution is to use a well known panel dataset to conduct a comprehensive test of the predictions of PT with a different, clearly discussed, assumption for the reference point. We consider that using future changes in income, or the expectations of such changes, is an unnatural and perhaps hard to justify choice. It is however straightforward to justify why income changes that are currently experienced have an effect on LS, and we want to know if this effect shows the asymmetries described by PT. Our paper fills this gap. We test for the presence of all prominent features of EUT and PT in a model which nests both theories. Moreover, we do not restrict our analysis to income, but also look at health and employment status. Section 7 considers for completion other reference point specifications similar to Vendrik and Woltjer (2007). We also confirm the use of the AA heuristic in the recall of past LS and evaluation of future LS.

3 Data

We use the German Socio-Economic Panel (SOEP), a dataset that has been widely used in LS research, covering the survey years 1992 – 2015, and 1984 – 1987. The health variable that we use is only available from 1992 onwards, and questions relating to recalled and expected LS are available only for the survey years 1984 – 1987. We consider individuals aged 18–85.

3.1 Outcomes

The measurement of current LS is done with the question: (1) "How satisfied are you at present with your life as a whole?" The respondents can answer with an integer number between 0 and 10, with 0 being the lowest and 10 the highest level of LS.[5] For the AA model (described below) we also use as dependent variables the answers to the following two questions: (2) "How satisfied with your life were you a year ago?" and (3) "And what do you think will it [LS] be in a year?". We refer to the answer to question (2) as recalled LS, and to the answer to question (3) as expected LS. Both questions used the same scale as the question about current LS. The order of the questions in the questionnaire is the same as presented here: (1) current LS, (2) LS last year, and (3) LS next year. We treat LS as a cardinal variable. This is

[5] The question clarifies that "0 means completely dissatisfied and 10 means completely satisfied", but no labels (such as "excellent", "good", etc.) are attached to the values.

a choice of convenience. We have also conducted all our estimations and tests based on the Blow-up and Cluster estimator in Baetschmann et al. (2015). The estimator treats the dependent variable as ordinal but not necessarily cardinal. It is consistent and based on the dichotomisation of the dependent variable at every threshold and the application of the conditional logit model in Chamberlain (1980). The estimator allows for individual fixed effects and makes no assumption about their correlation with the independent variables. The results were only marginally different from our main results. Ferrer-i Carbonell and Frijters (2004) also support the robustness of LS regression results to treating LS as cardinal or ordinal. The results for the Blow-up and Cluster estimator can be found in table 8 in the appendix.

3.2 Main independent variables

We are mainly interested in the effect of three variables on LS. For *income* we use equivalised net monthly real household income. Household income is typically used in LS research to account for the fact that resources are often pooled and distributed at the household level (Proto and Rustichini, 2015). We use the OECD equivalence scale: Total net monthly household income is divided by a weighted sum of household members, where the first adult household member is counted fully, any other person above the age of 13 as 0.7, and all younger household members as 0.5. To eliminate the effect of outliers we restrict the estimation sample to observations whose equivalised net monthly household income does not change by more than 500 Euros from one year to another, which comprise 88.8% of all person-year observations, with a sample mean of 1,414 Euros. Our estimation results are not sensitive to this restriction.

Health is captured by the individuals' self-assessment. Survey participants are asked: "How would you describe your current health?" and can choose between the answer boxes "Bad", "Poor", "Satisfactory", "Good" and "Very good", which we code from 1 to 5 in the same order. We create dummy variables for each category and use "Bad" as the omitted category in the regressions.

Labour force status is captured by three dummy variables for being employed (E), being unemployed (U), and not being in the labour force (N, the omitted category). Among the three categories only two can be ranked in terms of which is preferred by the respondent. For an employed person, unemployment is an available option, so employment is preferred to unemployment. For an unemployed person, employment is preferred by definition. We therefore treat employment as a status better than or preferred to unemployment.

Table 2: Summary statistics

	Mean	S.d.
Outcomes		
Life satisfaction (0-10)	6.98	1.76
Main independent variables		
Income (equivalised in 1,000 Euros)	1.31	0.67
Income gain (equivalised in 1,000 Euros)	0.07	0.11
Income loss (equivalised in 1,000 Euros)	-0.06	0.11
Self-assessed health (1-5)	3.35	0.95
Employed (E)	0.60	0.49
Unemployed (U)	0.06	0.23
Control variables		
Not in labor force (N)	0.34	0.47
Male	0.47	0.50
Has partner	0.74	0.44
Has children	0.38	0.38
Years of education	11.90	2.61
Age	48.4	16.3

Survey years 1992-2015. Observations: 320,053. Persons: 51,079. Income is net real income, equivalised by the OECD scale. The health categories range from bad (1) to very good (5).

For a pair-wise comparison between employment and non-participation as well as between unemployment and non-participation we cannot make any general assumptions in terms of their preference ranking.

3.3 Control variables

The choice of the remaining explanatory variables is informed by the literature on happiness (Dolan et al., 2008). We include the following individual characteristics: A dummy for males, a dummy for living with a partner, a dummy for having children, years of education, and six age categories (18-29, 30-39, 40-49, 50-59, 60-69, 70-85). Table 2 presents the summary statistics of our sample. Figure 1 depicts the distribution of LS and figure 2 depicts the distribution of self-assessed health for the pooled sample. Both LS and health are highly skewed, with LS having a mode at 8, and health at "good". These modes and the skewness of the two variables holds across survey years (not reported).

4 Model and estimation

We test whether LS exhibits the properties of reference dependence, diminishing marginal utility, and loss aversion, all of which are discussed further below. To this end we estimate an equation which can accommo-

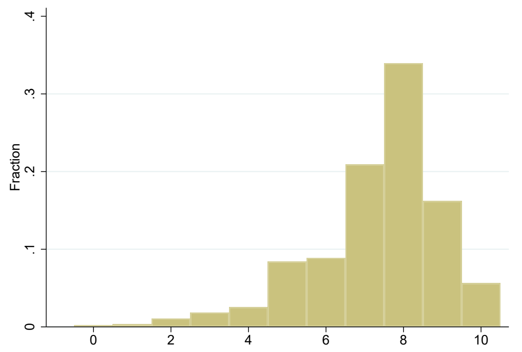

Figure 1: Distribution of LS. Pooled sample.

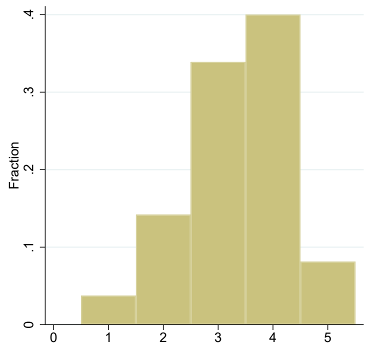

Figure 2: Distribution of self-assessed health. Pooled sample.

date and potentially reject all these properties. We then analyse whether an AA model can explain subjects' responses to recalled and expected LS. We discuss these models and hypotheses in turn.

To test for reference dependence, diminishing marginal utility, and loss aversion we estimate the following equation:

$$ls_{it} = \beta_0 + \beta_1 y_{it} + \beta_2 y_{it}^2 + \mathbb{1}_{\Delta y_{it}>0}\left(\gamma_1 \Delta y_{it} + \gamma_2 (\Delta y_{it})^2\right) + \mathbb{1}_{\Delta y_{it}\leq 0}\left(\delta_1 \Delta y_{it} + \delta_2 (\Delta y_{it})^2\right) \tag{1}$$

$$+ \alpha_2 H2 + \cdots + \alpha_5 H5 + \sum_{j\in J} \sum_{k\in(J\setminus j)} \alpha_{jk} TH_{it}^{jk} \tag{2}$$

$$+ \rho_E E + \rho_U U + \sum_{l\in L} \sum_{m\in(L\setminus l)} \rho_{lm} TL_{it}^{lm} \tag{3}$$

$$+ \eta \mathbf{X_{it}} + u_i + v_t + \varepsilon_{it} \tag{4}$$

The first line (1) includes the income variables, the second line (2) the health variables, the third line (3) the employment status variables, and the fourth line (4) includes other control variables, $\mathbf{X_{it}}$, person fixed effects, u_i, year fixed effects, v_t, and the classical error term, ε_{it}.

In (1), $\Delta y_{it} := y_{it} - y_{i,t-1}$, and $\mathbb{1}_A$ is an indicator variable which evaluates to 1 if the statement A is true, and to 0 if A is false. The income specification thus allows for *level* effects (through β_1 and β_2), and for *change* effects, differentiated by gain and loss effects (through γ_1, γ_2, δ_1 and δ_2). In the former case the amount of income available to the individual affects her LS. In the latter the change in income affects her LS. Using a quadratic function allows for concavity and convexity. One objection is that the quadratic function imposes an (inverted) U-shape on the LS-income relationship. However this will not be a problem as long as most of the observed incomes fall into the domain for which LS is only increasing. The decreasing part of the function would be empirically irrelevant.[6]

In (2) we include dummy variables for all but one of the different health categories, $H2$ to $H5$, as well as all possible transitions TH from one health state to another, where J is the set of integers from one to five (one for each health state). For example, TH^{23} is 1 if an individual reported the second health category in the previous year, and reports the third category in the current year. Finally, for labour force status, we also include dummies for being employed and unemployed, E, U, as well as all possible transitions TL from one labour force status to another, where L is the set $L = \{E, U, N\}$ and N stands for not being in the labor force. For example, TL^{EU} is 1 if an individual was employed in the previous year, but is unemployed in the current year.

In the following subsections we discuss our hypotheses and how we are going to test them. Table 3 summarises the main hypotheses along with their results. To control the power of our tests we have formulated our hypotheses such that the features we are looking for are stated as alternative hypotheses.

4.1 Income, health and employment are goods

We first establish that the variables of interest are considered "goods" in terms of LS. The literature has firmly established that income, good health, and not being unemployed exert a positive influence on LS. We re-affirm these results for completeness. It is sensible to show that the variables of interest are desirable,

[6] A separate effect of past income is not identified in the linear model. However under the assumption that past income has a positive independent effect on LS, its omission leads to an overestimation of the effect of current income and to an underestimation of the effect of income changes. To see this for example for income gains, consider that

$$\beta_1 y_t + \gamma_1 \Delta y + \eta y_{t-1} = (\beta_1 + \eta) y_t + (\gamma_1 - \eta) \Delta y$$

Our estimate for current income thus combines the effect of current and past income, while the estimate for income change is biased downward by the value of η.

before proceeding to tests regarding reference points or loss aversion.

For income to be a good, it needs to increase LS. Holding fixed the change in income Δy (and omitting the individual subscript) we have the following condition for income to be a good:

$$\frac{\partial ls}{\partial y} = \beta_1 + 2\beta_2 y_t > 0, \tag{5}$$

leading to the following hypothesis:

Hypotheses: (Y1) *Income is a good,*

$$\mathbf{H_{a,y_1}}: \quad \beta_1 + 2\beta_2 y_t > 0.$$

For health and labour force status we formulate the following hypotheses, assuming that a person has been in the same state in the previous year – that is all transition dummies are zero:

Hypotheses: (H1) *Better health increases LS,*

$$\mathbf{H_{a,h}}: \quad \alpha_j - \alpha_k > 0 \quad \forall\, j > k.$$

(L1) *Being employed is better than being unemployed,*

$$\mathbf{H_{a,l}}: \quad \rho_E - \rho_U > 0.$$

4.2 Reference point

PT postulates that utility is derived from the value of a variable compared against a reference value. What that reference value in a given context should be is not always clear. Kőszegi and Rabin (2006) argue that "a person's reference point is her probabilistic beliefs about the relevant consumption outcome held between the time she first focused on the decision determining the outcome and shortly before consumption occurs", thus proposing expected consumption as reference point. Others have argued that in evaluating their LS people compare themselves to a peer group (see Vendrik and Woltjer (2007) and the papers cited there). In that case the reference point is usually constructed as the average of the variable of interest within a sub-sample which share the demographic characteristics of the individual for whom the reference point is being calculated. In panel data a natural reference point is the lagged value of the variable of interest, assuming that the reference point is the person herself in her near past. This approach has been taken by many papers discussed above which have — sometimes implicitly — used past income values as reference point.

Whether people use their past self or a social comparison group as reference depends on a number of factors (Schwarz and Strack, 1999; Wilson and Ross, 2000): their goal in engaging in the comparison (an accurate self-assessment or increasing their self-esteem), what their attention is focused on (depending on the previous question, the surroundings, or other people in the room), and their recent experiences or concerns (whether they have recently transitioned from one life stage to another, e.g. retired, married, etc). While social comparisons received more attention in the psychological literature, Wilson and Ross (2000) demonstrate that comparisons to one's own self are at least as common as social comparisons. There are two more arguments to support the past self as a reference point in our context. First, the fact that any respondent included in the estimation sample must have answered the survey in the previous year, and probably has done so over a number of years. Second, the LS question is the last question being asked during the interview, after the respondent has been engaged in a lengthy interview about their life circumstances, opinions and attitudes. Questions about satisfaction with certain life domains are asked at the beginning of the interview, so that an attention focus on any particular aspect of their lives induced by the interview itself is unlikely. Finally, the variables characterising the state of the past self are readily available. Using a social reference point would require a number of ad-hoc or poorly justified choices: which social group to choose, over which spatial and temporal dimension, and which statistic to use as reference point. Hence, we use the past self as reference point, but we return to social comparisons in the robustness and extensions section.

The reference point test in the context of PT is straightforward. If LS is not evaluated against a reference point (or, more conservatively, if the past value of a variable is not a reference point) then the coefficients on changes and transitions should be zero. Conversely, if LS is evaluated *only* against reference points, then the coefficients on the current levels should be zero. We test whether income changes and health and employment status transitions have an influence on LS once income levels and health and employment states are controlled for. PT postulates a positive effect of "increases" of those goods. We therefore have for income gains:

$$\frac{\partial ls}{\partial \Delta y} = \gamma_1 + 2\gamma_2(\Delta y) > 0. \tag{6}$$

Suppose two persons have the same income. If the first has had the same income in the previous year, while the second person has arrived at their current income from a lower level, then 6 implies that the second person enjoys higher levels of LS.

For income increases to be a good in the loss domain – a reduction of an income loss – we have

$$\frac{\partial ls}{\partial \Delta y} = \delta_1 + 2\delta_2(\Delta y) > 0. \tag{7}$$

Suppose two persons have the same income. If the first has had the same income in the previous year, while the second person has arrived at their current income from a higher level, then 7 implies that the second person enjoys lower levels of LS.

Similarly, for two people with the same health status, we should expect higher LS for the person who arrived at this status from a poorer health state, and lower LS for the person who arrived at this status from a better health state. For employment, a person who has lost his job should be less satisfied than someone who is and was unemployed. A person who found a job should be more satisfied than someone who is employed but was also employed in the previous period.

For small gains in income (Δy approaching zero) we test:

Hypotheses: (RP1) *LS is increasing in changes in the gains domain,*

$$\mathbf{H_{a,y}}: \quad \gamma_1 > 0,$$

$$\mathbf{H_{a,h}}: \quad \alpha_{jk} > 0 \quad \forall j,k \quad \text{such that } j < k,$$

$$\mathbf{H_{a,l}}: \quad \rho_{UE} > 0,$$

Hypotheses: (RP2) *LS is increasing in changes in the loss domain,*

$$\mathbf{H_{a,y}}: \quad \delta_1 > 0,$$

$$\mathbf{H_{a,h}}: \quad \alpha_{jk} < 0 \quad \forall j,k \quad \text{such that } j > k,$$

$$\mathbf{H_{a,l}}: \quad \rho_{EU} < 0,$$

In the hypotheses above, only transitions between the employed and unemployed statuses are considered, since only these are unambiguously ranked in relation to each other.

4.3 Diminishing marginal utility

Testing for the presence of diminishing marginal utility in LS requires certain assumptions on the variables. If both LS and the independent variable are cardinal, testing for diminishing marginal utility is straightforward. If we relax the cardinality assumption for LS but retain its ordinal property, we can still apply a latent variable framework such as ordered probit or logit and assume that the latent variable has cardinal properties (though not the reported LS). We have estimated our model under both assumptions. We used the fixed effects OLS estimator for the cardinal, and the fixed effects ordinal logit (Blow-up and cluster, Baetschmann et al., 2015) for the ordinal specification. The two specifications yielded almost identical results. We produce a table of results for the ordinal model in the appendix, and proceed here with the cardinal model.

Cardinality in the independent variable however cannot be dispensed with. To see this, consider a person who reports the same increase in LS when going from satisfactory to good and from good to very good health. If these two changes in health categories reflect an equivalent change in the person's underlying "true" health we would conclude that marginal utility is constant. But if the incremental gain in health between satisfactory and good is smaller than the gain in health between good and very good, the person would still exhibit diminishing marginal utility with respect to health. Of our explanatory variables, income is the only cardinal variable, therefore it is the only variable for which we can test whether it exhibits diminishing marginal LS. While health categories can be ordered, we do not assume that gains in health have ratio properties.

We test for diminishing marginal utility in *levels* and in *changes*. The two channels are not mutually exclusive but EUT supports the levels effect, while PT supports the changes effect. For income levels to have diminishing marginal effects on LS, the sufficient condition is $\beta_2 < 0$.

For income changes we can use equations 6 and 7. Taking derivatives with respect to Δy we obtain

$$\frac{\partial^2 ls}{\partial \Delta y^2} = \gamma_2, \qquad \frac{\partial^2 ls}{\partial \Delta y^2} = \delta_2.$$

Thus for LS to be concave in gains and convex in losses we require

$$\gamma_2 < 0, \qquad \delta_2 > 0.$$

The hypotheses are the following:

Hypotheses: (DMU1) *LS is concave in income levels,*

$$\mathbf{H_a}: \quad \beta_2 < 0.$$

(DMU2) *LS is concave in income gains,*

$$\mathbf{H_a}: \quad \gamma_2 < 0.$$

(DMU3) *LS is convex in income losses,*

$$\mathbf{H_a}: \quad \delta_2 > 0.$$

4.4 Loss Aversion

Loss aversion means that the decrease in utility due to a loss (of income, health, employment) is greater than the increase in utility due to the corresponding gain. To classify anything as a loss or a gain, a reference point must be defined. While marginal effects for health and employment status could not be estimated, the presence of loss aversion can, as individuals can go from good to bad health and vice versa, or from employment to unemployment and vice versa. For loss aversion in income, we need to compare $\frac{\partial ls}{\partial \Delta y}$ in the domain of gains to the same derivative in the domain of losses. Loss aversion requires that the rate of change of LS for a decrease in y be greater than the rate of change in LS for a corresponding increase. From equations (6) and (7),

$$\delta_1 + 2(-\Delta y)\delta_2 > \gamma_1 + 2\Delta y \gamma_2, \quad \forall \Delta y \geq 0.$$

In particular, $\lim_{\Delta y \to 0}$ implies $\delta_1 > \gamma_1$, giving the utility function with loss aversion its characteristic kink at the origin. For labour force status we compare only two states: employment and unemployment, as by definition employment is preferred to unemployment by both the employed and the unemployed. The change in LS for someone who moves from unemployment to employment (assuming that in $t-2$ her labour force status was also unemployed) is $(\rho_E - \rho_U) + \rho_{UE}$, and the change in LS for someone who moves from employment to unemployment (assuming that in $t-2$ her labour force status was also employed) is $(\rho_U - \rho_E) + \rho_{EU}$. The former is expected to be positive, and the latter to be negative. If so, loss aversion will also imply:

$$(\rho_E - \rho_U) + \rho_{UE} < -((\rho_U - \rho_E) + \rho_{EU})$$
$$\Rightarrow \quad \rho_{UE} < -\rho_{EU}.$$

For health, the same argument as in the labour force status case applies. However, as there are 5 (ordered) health categories, there are 10 comparisons that can be made.

The hypothesis on loss aversion is:

Hypothesis: (LA) *LS exhibits loss aversion in income, health and employment,*

$$\mathbf{H_{a,y}}: \quad \gamma_1 - \delta_1 < 0,$$
$$\mathbf{H_{a,h}}: \quad \alpha_{kj} + \alpha_{jk} < 0 \quad \forall\, j > 1 \quad \text{and} \quad \forall\, k < j,$$
$$\mathbf{H_{a,l}}: \quad \rho_{UE} + \rho_{EU} < 0.$$

4.5 Anchoring and adjustment

The final question we address is whether the answer to recalled LS and to expected LS can be described by an anchoring and adjustment (AA) heuristic as in Tversky and Kahneman (1974). According to AA, when people are asked a question which they find difficult to answer, they substitute the question by a different question to which they know the answer, and adjust from this answer towards the conjectured direction of the answer to the initial question. In the case of LS, when the respondent is asked to evaluate her LS in the previous year, but she cannot recall her exact LS, she will use her current LS as an anchor and then adjust up or down depending on whether she feels that she is more or less satisfied with her life now than last year. For example, a person might evaluate her LS to be higher compared to the previous year. Her answer might then be constructed as the value of her current LS minus 1. We estimate:

$$R_{it}(ls_{i,t-1}) = \beta_0 + \beta_1 ls_{it} + \beta_2 ls_{i,t-1} + u_i + \varepsilon_{it}, \tag{8}$$

$$E_{it}(ls_{i,t+1}) = \gamma_0 + \gamma_1 ls_{it} + \gamma_2 ls_{i,t+1} + v_i + \eta_{it}, \tag{9}$$

Here $R_{it}(ls_{i,t-1})$ gives the LS at time $t-1$ recalled by individual i at time t, and $E_{it}(ls_{i,t+1})$ gives the LS at time $t+1$ expected by individual i at time t. These equations can accommodate a range of models about how people answer the questions about past and expected LS. If $\beta_1 = 0$ and $\beta_2 = 1$ then people are perfectly recalling their past LS. Our a priori conjecture is that people follow an AA heuristic with the current LS level as the anchoring point. That is, at time t, the anchor for $R_t(ls_{t-1})$ is ls_t. Therefore, we expect the anchor to serve as a good predictor for recalled LS in equation (8) implying $\beta_1 > 0$. Moreover, if the adjustment process brings the answer closer to past LS starting at the anchor, we should observe that past LS improves the answer to recalled LS, implying $\beta_2 > 0$.

For expected LS, if $\gamma_1 = 0$ and $\gamma_2 = 1$, then people are perfectly predicting their future LS. As for recalled LS, we conjecture that people use their current LS level as the anchoring point. That is, at time t, the anchor for $E_t(ls_{t+1})$ is ls_t. Therefore, we expect the anchor to serve as a good predictor of expected LS in equation (9) implying $\gamma_1 > 0$. If the adjustment process brings the answer closer to future LS starting at the anchor, we should observe that future LS improves the answer to expected LS, implying $\gamma_2 > 0$. That is, if the expected value of LS in a previous year is adjusted towards the true future LS value, starting at the anchor, it must have additional predictive power and should be positively related to current LS. AA is not the only model that implies $\gamma_1 > 0$ and $\gamma_2 > 0$. An alternative is the projection bias model in Loewenstein et al. (2003) according to which people "exaggerate the degree to which their future tastes will resemble their current tastes." While the projection bias model implies the same coefficients as AA for equation (9), it does not make any predictions about equation (8).

Hypotheses: (AA1) *Individuals use their current LS as an anchor for recalled or expected LS. In equations (8) and (9),*

$$\mathbf{H_a}: \quad \beta_1 > 0 \quad \text{and} \quad \gamma_1 > 0.$$

(AA2) *If $\beta_1 > 0$, individuals adjust their recalled or expected LS in the direction of past or future LS. In equations (8) and (9),*

$$\mathbf{H_a}: \quad \beta_2 > 0 \quad \text{and} \quad \gamma_2 > 0.$$

An advantage of AA is that it allows us to make sense of imperfect recall, since the procedure can be applied to all questions where the answer is numeric or ordinal, and where a suitable anchor can be posited.

5 Results

The exact results for our main econometric model from equation (1 – 4) can be found in table 7 in the appendix. The results are generally in line with what is known about LS. Having a partner, having children, not being unemployed, being in good health, and income are associated with higher levels of LS. The differences between the OLS and fixed effects coefficients demonstrate the importance of unobserved individual characteristics. Our preferred specification is therefore the fixed effects estimator.[7] Table 3 summarizes the hypotheses and test results on reference points, diminishing marginal utility, loss aversion, and anchoring and adjustment. Regarding our tests on whether income, health and employment are goods in term of LS (panel A), we find that all of those variables are goods. Hypothesis Y1 is rejected at the 1% level for a wide range of income levels y_t (between 0 and 5,000 Euros – 99.8% of our sample). Hypothesis H1 is strongly rejected for all possible health states. L1 is also strongly rejected.

5.1 Reference point

From panel B of table 3 we see that none of the hypotheses in RP1 can be rejected. For income, there is no statistical evidence that a gain has any effect on LS over and above the level effects. For employment, a transition from unemployment to employment actually carries a significant (for a one-sided test) negative sign.

[7] A Hausman test rejects the equality of coefficients from the random effects and fixed effects estimators.

Table 3: Summary of hypotheses to be tested.

Code	Alternative hypothesis H_a	H_a supports:	p-value
Panel A: Level effects			
Y1	LS is increasing in income levels.	EUT	0.000
H1	LS is increasing in good health.	EUT	0.000
L1	LS is increasing in employment.	EUT	0.000
Panel B: Reference point			
RP1	LS is increasing in income changes in the gains domain.	PT	0.373
	LS is increasing in health gains in the gains domain.	PT	>0.517
	LS is increasing when one becomes employed.	PT	0.988
RP2	LS is increasing in income changes in the loss domain.	PT	0.001
	LS is increasing in health gains in the loss domain.	PT	>0.956
	LS decreases when one becomes unemployed.	PT	0.030
Panel C: Diminishing marginal utility			
DMU1	LS is concave in income levels.	EUT	0.000
DMU2	LS is concave in income changes in the domain of gains.	PT	0.120
DMU3	LS is convex in income changes in the domain of losses.	PT	0.004
Panel D: Loss aversion			
LA	LS exhibits loss aversion in income.	PT	0.044
	LS exhibits loss aversion in health.	PT	n.a.
	LS exhibits loss aversion in (un)employment.	PT	0.002
Panel E: Anchoring and adjustment			
AA1	Recalled LS is anchored at current LS.	AA	0.000
	Expected LS is anchored at current LS.	AA	0.000
AA2	Recalled LS is adjusted towards past LS.	AA	0.000
	Expected LS is adjusted towards future LS.	AA	0.000

*In the case of RP for health, we tested each possible transition between health states separately. The reported p–value is the lowest among the 10 tests.

However, the results for the loss domain are in agreement with the properties of PT. An income increase in the loss domain — that is making the loss smaller — increases LS. The transition from employment to unemployment reduces LS, controlling for employment status. A person who just became unemployed has lower LS than a person who was unemployed in the previous period and is still unemployed. Thus, the evidence is that losses hurt but gains do not help.

For health we get a completely reversed, and unexpected, result. Since there are many different health improvement transitions, we test each of them. The p-value in the table is the smallest out of all p-values for those tests, that is, none of the tests for health gains had a p-value less than 0.517. However, this masks the fact that all but one of the tests supported a *negative* coefficient. That is, someone whose health state is x as a result of a health improvement still enjoys lower LS than someone whose health state is and was x. The reverse is true for health losses. All coefficients here are positive. That is, a person whose health is x as a result of a health decrease still enjoys higher LS than someone whose health is and was x.

Taken together, the results reported in panels A and B do not lend exclusive support for or against EUT or PT. Rather, LS seems to be best described by a model which incorporates both level and change effects, and exhibits asymmetries between gains and losses, albeit with important differences to the value function in PT.

5.2 Diminishing marginal life satisfaction

Panel C presents the results for hypotheses DMU1 to DMU3. LS is concave in income levels. While it is clearly convex in the loss domain of income, the support for concavity in the gains domain is not strong. We cannot reject that it is flat or convex at the 10% significance level.

5.3 Loss aversion

Loss aversion (panel D) is present at the 5% significance level for income, and is strongly present for employment status. Given our results for reference points for health transitions, it makes no sense to test for loss aversion in health. Rather, if we test for "loss preference", that is, the LS increase after a drop in health exceeding the LS decrease after a symmetric health improvement, we find significant support at the 5% level for this phenomenon for six out of the ten possible health transitions.

One explanation is that both LS and self-assessed health are measures that reflect how the respondent sees herself. If the respondent over-values her health gain — compared to her actual health improvement — loss aversion will not be detected using self-assessed health even if it is present in actual health. Another possibility is that actual health gains reflected by reported better health states do not correspond to equivalent health losses reflected by reported worse health states. For example, respondents might report better health states for small improvements in health, while they might report no change in health states for small health deterioration. Loss aversion – even if present – will not be detected in that case. A third possibility is that the crude health reporting scores obscure important distinctions between acute and chronic health conditions. A change in health level to x might be associated with higher LS, as compared to being stable at x, because the second situation is more likely to be associated with a chronic condition, whereas the first situation is more likely to be associated with an acute, more easily curable, condition.

Since these explanations derive mainly from the subjectivity of self-assessed health, we have repeated our estimation with an alternative health variable which is based on less subjective measures of good health. We created a health index which is the sum of a dummy for not having visited a doctor in the previous three months, a dummy for not having any hospital visit in that year, and a variable categorising the extent to which health interferes with daily functions (0: substantially, 1: partially, 2: not at all). Our index ranges from 0 to 4, and we have used the same parametrisation as for our original health variable (dummies for all categories and transitions). The results were unchanged. As in our benchmark model health improvements decrease, and health decreases increase LS in most of the pair-wise health state comparisons.

5.4 Shape of LS function

The LS function for changes in income in a range of -200 to +200 Euros is depicted in figure 3. In the left panel we hold current income levels constant so that income changes are equivalent to varying past income. This corresponds to a pure PT scenario. In the right panel we hold past income constant. A change in income affects LS thus both through Δy but also through y, as an income change also changes y if we hold constant past income. While the figure on the left has some characteristics of the value function in PT, the figure on the right resembles the value function more closely. It has most of the characteristic features of Prospect Theory: the kink at the origin, a slight concavity in gains, convexity in losses, and a general stronger effect

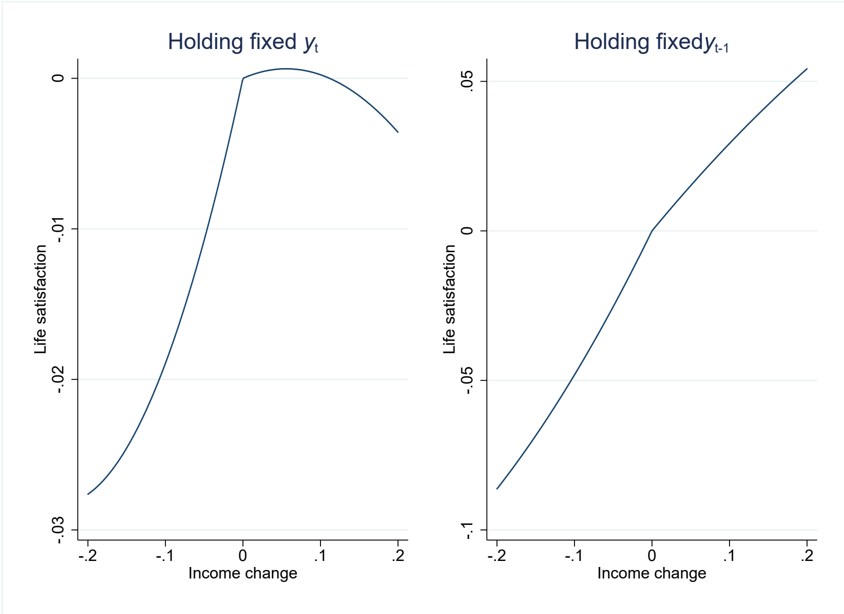

Figure 3: Life satisfaction as a function of income changes. Income changes are in 1,000s of Euro; the vertical axis is LS relative to no income change.

of losses than gains.

5.5 Anchoring and adjustment

Table 4 shows results from the estimation of equations (8) and (9) and panel D of table 3 shows the corresponding results for our hypotheses. The main result is that current LS is a much better predictor of recalled LS than the actual LS that was reported in the past. Similarly, current LS is a much better predictor of expected LS than LS reported a year later. The recall result for equation (8) cannot be reconciled with any models in which the individual can perfectly recall LS. Similarly, models that make strong assumptions on the ability of individuals to predict future LS are also hard to reconcile with the result for equation (9). The results are in strong agreement with the AA model. In AA, the current LS serves as the anchor for determining the answer for both the recalled past LS and the expected future LS. Respondents then adjust in the correct direction. For example, the categories 7 and 8 are the most frequently reported levels of LS (21% and 30% respectively). A person with a current LS level of 7 and past level of 8 will, on average, recall a LS of 7.05. A person with a current LS level of 8 and past level of 7 will recall a LS of 7.61. It is clear that the anchoring part is much more important quantitatively than the adjustment.

The margin by which current LS is a stronger predictor than past/future LS is striking. This result has important consequences for example in the evaluation of the effectiveness of medical treatments. Treatments can be evaluated prospectively (patients are asked to evaluate their health before and after treatment), or retrospectively (patients are asked to evaluate their health after treatment and compare it to their health before treatment). Prospective evaluations can be biased by adaptation by patients, and retrospective evaluations are prone to recall errors. The results on recalled LS here demonstrate how strongly the remembrance of the past can be tainted by the present.

Two more ex-post observations can be made. First, the R^2 for equation (9) is higher than for (8). The size of the coefficient for current LS (the anchor) is higher for expected LS than for recalled LS. Current LS seems to be a much better anchor for the question about expectation than the question about past LS. At the same time the adjustment works better for the past LS question. This is not surprising, in light of the fact that recalling information on LS should be much easier than making a prediction of LS. Furthermore, any anxiety or optimism about future events are likely to feed into current LS, so that current LS becomes a strong predictor for expected LS.

The second observation is that expectations and recalls seem to be systematically biased. Figures 4 and 5 display the histograms for the difference between expected and recalled LS, and the corresponding realised LS, that is $E_{it}(ls_{i,t+1}) - ls_{i,t+1}$ for expectations and $R_{it}(ls_{i,t-1}) - ls_{i,t-1}$ for recalls. Deviations of expectations from realisations are skewed to the right, and deviations of recalls from realisations are skewed to the left. The mean of the former is significantly positive, and the mean of the latter is significantly negative. In general, respondents seem to have an overly optimistic view of the future, and an overly negative recall of the past.

6 Discussion

The purpose of our analysis was to establish which features of the utility function in EUT and PT are displayed by reported current LS, and whether answers to recalled and expected LS can be explained by an AA model. With regard to the first question the findings suggest that LS can be best described as a hybrid measure exhibiting properties of both EUT and PT. LS follows EUT in that higher income and better health levels are associated with higher LS. Individuals enjoy higher LS when employed rather than unemployed. Furthermore, LS is concave in income levels and thus exhibits diminishing marginal LS.

Table 4: Expected and recalled life satisfaction.

	Dependent variable	
	$R_t(ls_{t-1})$	$E_t(ls_{t+1})$
ls_t	0.647***	0.740***
	(0.010)	(0.006)
ls_{t-1}	0.086***	
	(0.008)	
ls_{t+1}		0.023***
		(0.005)
Constant	1.835***	1.896***
	(0.090)	(0.062)
Observations	29,288	38,177
R-squared	0.387	0.555

Robust standard errors in parentheses. The R-squared is the squared correlation between the de-meaned life satisfaction and predicted de-meaned life satisfaction. Stata reports this measure as R-squared within. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

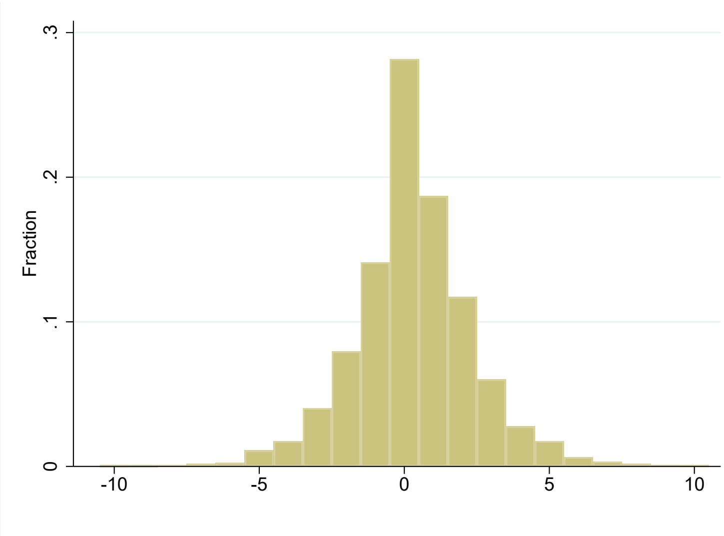

Figure 4: Expected minus realised life satisfaction.

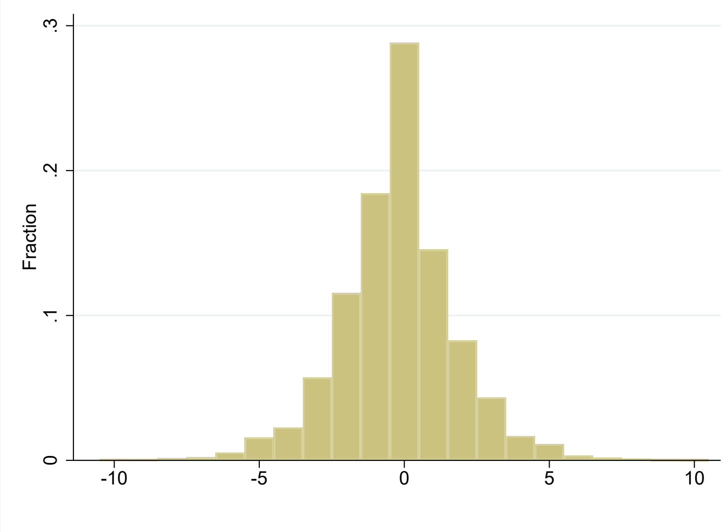

Figure 5: Recalled minus realised life satisfaction.

LS follows PT in that income gains, health improvements, and a transition from unemployment to employment are associated with higher LS. We also find that for income LS is convex in the losses domain. LS also exhibits loss aversion with respect to income and to employment. Health changes on the other hand do not show the gain-loss asymmetric effects on LS, suggested by loss aversion. Finally, we demonstrated that recalled and expected LS are consistent with an AA model where current LS is used as the anchor, and the adjustment is in the direction of past LS (in the case of recalled LS) and future LS (in the case of expected LS).

One can argue that the validity of our findings will be compromised under certain conditions. Vendrik and Woltjer (2007) discuss the possibility that people might under-report LS for high values and over-report for low values. Related to this, they argue that the shape of the "true" LS function might be different from the shape of the reported LS function. The validity of these objections hinge on how one interprets LS. While some studies have used LS and other subjective well-being measures as direct or proxy measures of utility, they have not discussed what exactly they mean by utility. Kahneman et al. (1997) provide a useful taxonomy. They distinguish utility by how it is inferred — decision utility vs. experienced utility — as well as by when and over what time period it is measured — instant, total, and recalled utility —, with important consequences for decision making (Dolan and Kahneman, 2008). Decision utility is the utility concept underlying modern economics, according to which utility is that theoretical construct that individuals maximise with their choices under a set of constraints. If LS is a measure of decision utility, then any distinction between reported and true LS is not meaningful, as long as reported LS can rationalise and in turn explain individuals' choices, just as different utility functions can rationalise observed choices. If however one wishes to use LS as a measure of experienced utility then issues of measurement, inter- and intra-personal comparability, and reporting behaviour become paramount.

The experienced utility interpretation can be supported by the view that LS is seen as a global evaluation of one's life. Schwarz and Strack (1999) argue that subjective well-being measures suffer from a myriad of context effects, however the LS question is asked to survey participants at the end of a lengthy interview and asks about life in general — well after a section which asks about satisfaction with different domains in life. This should minimise the probability that their answers will be biased due to attention focus or context effects. Perhaps a more serious argument against the experienced utility interpretation is the argument that LS is evaluated based on past and current, but also prospective experiences (Dolan and Kahneman, 2008).

Given that EUT and PT are theories of choice, the relevant utility concept in the current context is decision utility. However, LS is unlikely to perfectly overlap with decision utility. Benjamin et al. (2014) show that actual choices do not correspond to the options with the highest anticipated happiness in a context of residency choices of medical students. But even if realised choices would correspond to the highest anticipated LS, phenomena like adaptation (Loewenstein and Ubel, 2008) and projection bias (Loewenstein et al., 2003) will induce systematic differences between anticipated and realised LS.

How LS and other subjective well-being measures exactly relate to the different utility concepts is an exciting and open question but it is not the focus of this paper. We have established that LS — as reported by survey respondents — shows many of the properties of utility functions stipulated by EUT and PT.

7 Extensions

We consider two robustness checks and extensions to the model. First, we consider a number of alternative reference points that have been used and tested in the literature. Second, we explore the possibility that the evidence for loss aversion might be driven by heterogeneity in income sensitivities.

7.1 Alternative reference points

In this section we consider alternative reference points. Our aim is to understand whether reference points other than one's own past have explanatory power, and whether the evaluation of one's own standing relative to the reference point is asymmetric. If the value of one's own state variable is below the reference point, then this can be interpreted as a loss, and the marginal effect of the variable should be greater than when one is above the reference point. The empirical literature has used reference groups which were constructed along geographical, demographical, and occupational dimensions, or based on predicted values from Mincer-type equations. Clark et al. (2008b) provide an exhaustive discussion. Cheung and Lucas (2016), Jorgensen et al. (2010) and Luttmer (2005) use average incomes in geographic areas as reference income. Ferrer-i Carbonell (2005) and Vendrik and Woltjer (2007) — using the same data as we do — create reference groups based on age, education, and whether the respondent resides in East or West Germany. In a field experiment Card et al. (2012) assume that colleagues working in the same university department are the relevant reference group for job satisfaction, and Clark and Oswald (1996) use the predicted income from a Mincer-type regression as reference income. The latter two approaches are arguably more relevant for job than life satisfaction.

We follow the literature and construct reference values for the variables household income, health, and unemployment, in five different specifications. The first specification uses the individuals' past state variables (as in our benchmark model). The next three specifications segment the population and calculate annual cell averages to serve as reference values. Cell averages are based on observations in the SOEP other than the household members for whom the average is being computed. For the second specification, we calculate reference values for each of the 16 German states, and, where applicable, separately for urban and rural locations. This results in 27 geographic areas. For the third specification we segment the population into five age, five education, and East-West cells as in Ferrer-i Carbonell (2005). The fourth specification uses only observations who are in full-time employment and segments this population by the first digit of the four-digit occupation code and by East and West Germany. For the last specification we use full-time employees to regress the natural logarithm of household income and health on age, age squared, a male dummy, years of education, years of education squared, the natural logarithm of annual hours worked, state dummies, year dummies, occupation dummies (based on the first digit of the four-digit occupation code), and household size dummies. We use the predicted values from those regressions as reference values. As in our main model, LS is then modelled as a function of a quadratic polynomial in income, as well as a quadratic polynomial of income relative to reference income, where parameters for relative income can differ below and above 0 (e.g., we replace $y_{i,t-1}$ by y_i^*, the reference of person i).

For ease of comparison between the different specifications, instead of using dummies for all health states and all pair-wise transitions, we use the original five-point health scale and define reference health either as the individual's past health score (specification 1), or the average health score of the corresponding reference group (specifications 2 to 4), or predicted health based on a regression of health on the same variables as the Mincer regression (specification 5). Finally, reference unemployment is either the past unemployment state of the individual, or the unemployment rate among their corresponding reference group. Where applicable, we also create an interaction term between unemployment and reference unemployment.

Table 5 presents results for our benchmark model and alternative specifications of reference groups. The panel for income reports the slopes of relative income when it is close to zero. The panel for health is the effect of relative health on LS, and the third panel summarises the effect of reference unemployment on LS. All models include level effects of income, health, and employment which are similar to the benchmark results, and we therefore omit them from the results table.[8] The benchmark model (column 1) echoes our earlier result. Income changes in the gains domain are neutral, but have positive slope (and are convex) in the loss domain. Current health increases LS, but holding current health fixed, health improvements reduce, and health deteriorations increase LS. A person whose health is stable enjoys higher LS than someone with the same health state, but who arrived at their health from a less healthy state. At the same time, among two equally healthy individuals, the one whose health was better in the previous period enjoys higher LS. The effect of a gain is the same in magnitude as the effect of a loss. Past unemployment also exerts a negative effect on current LS, even if the person is not unemployed in the current period. However, past unemployment does not have any effect over and above the current unemployment effect if the person is still unemployed (the interaction term offsets the effect of past unemployment). This is consistent with the previous finding that people do not adapt to unemployment (Clark et al., 2008a).

For all other specifications we find that LS *decreases* in income gains and is neutral with respect to income losses, thus reversing the finding obtained with the reference point as the past self. As for health,

[8]Results are not exactly the same because we do not dichotomise all possible health states and transitions, but use the original health variable.

Table 5: Results: Alternative reference points. Fixed-effects models

	Reference point				
	(1)	(2)	(3)	(4)	(5)
	Past self	Geographic cells	Demographic cells	Occupation cells	Mincer prediction
Slope income gains	0.005	-0.183***	-0.249***	-0.134***	-0.185***
	(0.072)	(0.051)	(0.032)	(0.042)	(0.056)
Slope income losses	0.279***	0.007	-0.039	-0.050	1.448
	(0.079)	(0.069)	(0.047)	(0.066)	(1.411)
Health gains	-0.101***	-0.256***	-0.125***	-0.324***	-0.133*
	(0.006)	(0.068)	(0.031)	(0.103)	(0.079)
Health losses	0.104***	0.128*	0.203***	-0.031	0.124
	(0.007)	(0.068)	(0.031)	(0.103)	(0.079)
Reference unemployment	-0.098***	-1.131***	-0.502***		
	(0.019)	(0.255)	(0.118)		
Unemployed × Reference unemployment	0.093***	-1.710***	0.232		
	(0.032)	(0.515)	(0.231)		
R-squared	0.092	0.094	0.093	0.076	0.076

Robust standard errors in parentheses. Cell means exclude the value for the observation for whom the mean is calculated. The coefficients in row 2 and 3 are the slopes of income changes in the gains and loss domain respectively. Regressions include the same variables as in table 7 — including income, health state and employment status — except for pairwise health transitions. In column 1, health gains and losses quantify the effect of increasing and decreasing health respectively on LS over and above the level effects. In columns 2 to 5, health gains and losses quantify the slopes of health gains when health is above and below reference health respectively. The R-squared is the squared correlation between the de-meaned life satisfaction and predicted de-meaned life satisfaction. Stata reports this measure as R-squared within. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

we find that, holding health state fixed, health gains above the reference point decrease, and health gains below the reference point increase LS. Since level effects are positive and greater in magnitude, this result implies that health gains increase LS more strongly when initial health is weak than when initial health is strong. Reference unemployment depresses LS in columns 2 and 3, suggesting the presence of negative externalities (for example through poorer public services). Interestingly, the regional negative externality is amplified through one's own unemployment. Both being unemployed, as well as living in a high unemployment region is bad, but being unemployed in a high unemployment region is particularly depressing.

Our benchmark model and the models based on geographic and demographic segmentation perform similarly well, while the specifications based on an occupational segmentation (column 4) and on a Mincer regression (column 5) perform worse in terms of the correlation of observed and predicted LS (R-squared), probably because the sample is conditioned on being employed, and unemployment explains an important part of the variation in LS in a sample containing both employed and unemployed observations.

Overall, our results in this section suggest that there is probably not one "right" reference point. Indeed, given the measurement error due to the strong discrepancy between what probably constitutes the right reference group and our aggregate measures, finding such strong and significant effects in most specifications is surprising. Furthermore, while relative measures and reference points exert important influences on LS, the effects do often not clearly align with predictions from PT, especially in the case of health. LS is a multidimensional construct, and our intuition suggests that it is evaluated against reference points in each of its dimensions, with varying weights.

7.2 Heterogeneity in income sensitivity

In this section we analyse whether the loss aversion in income that we have detected is an artefact of heterogeneous income sensitivities in our sample. The question is motivated by findings in the consumer choice literature. Price-sensitive consumers will purchase cheaper brands and have lower reference prices. Since prices will often be above their reference price, they will be facing "losses" more often than consumers who are less price-sensitive and who have higher reference prices. As a result, in a sample of consumers, a kink around the reference price will be estimated which reflects price-response parameter heterogeneity rather than loss aversion (Bell and Lattin, 2000). In our case, if income-sensitive individuals face income losses relatively more often, we would also find a steeper slope of income on LS in the domain of income losses than income gains. To see if this is a plausible scenario we first tested whether observations with high incomes are more or less likely to experience an income gain or loss. We found that high incomes entail a higher probability of ensuing losses and smaller expected income changes.[9] While this can easily be explained by a regression to the mean, even at an individual level, it still poses a challenge to the identification of loss aversion as explained above (e.g., if individuals are more loss averse when their incomes are high). We then estimated our model given by equations 1 to 4 on two subsamples, one consisting of individuals whose first observed income is below the median, and the other above. We do find that individuals in the lower half of the income distribution are indeed more sensitive to income changes, both upward and downward.

To account for heterogeneity in income-sensitivity we use a finite-mixture model: We assume that there are K segments in the population which each have different sensitivities (in terms of LS) to income gains and losses. The fraction of the segment k in the population is given by

$$P_k = \frac{\exp(\alpha_k)}{\sum_{i=1}^{K} \exp(\alpha_i)},$$

where α_i are parameters to be estimated (and α_1 is normalized to zero). To economise on parameters that need to be estimated, we specify the LS function of individual i in segment k and at time t as

$$ls_{itk} = \beta_k + \beta_{yk} y_{itk} + \beta_{y^+k} y^+_{itk} + \beta_{y^-k} y^-_{itk} + \varepsilon_{itk},$$

where ε_{itk} is assumed to be a normally distributed random variable with mean zero and the same variance across the different segments, y_{itk} is log of income in Euros, and y_t^+ and y_t^- are defined as

$$y_t^+ = \begin{cases} \ln(1+(y_t - y_{t-1})) & \text{if} \quad y_t > y_{t-1} \\ 0 & \text{else} \end{cases}$$

and

[9] Results not reported, but available upon request.

Table 6: Loss aversion heterogeneity: Four segments.

	$\ln(y)$	Slope gains	Slope losses	Fraction (in %)
Segment 1	0.204***	0.092***	-0.023	87.4
	(0.009)	(0.016)	(0.017)	
Segment 2	0.584***	-0.306***	0.378***	7.4
	(0.032)	(0.073)	(0.072)	
Segment 3	0.002	.511***	-0.151	4.5
	(0.035)	(0.084)	(0.083)	
Segment 4	0.703***	-0.225	0.131	0.7
	(0.080)	(0.188)	(0.188)	

$n = 373,385$. Standard errors in parentheses.

$$y_t^- = \begin{cases} \ln(1-(y_{t-1}-y_t)) & \text{if} \quad y_t < y_{t-1} \\ 0 & \text{else} \end{cases}$$

The likelihood contribution of an observation (suppressing subscripts), conditional on belonging to segment k, is thus

$$\phi(\varepsilon|k) = \phi\left(ls - (\beta_k + \beta_{yk}y + \beta_{y^+k}y^+ + \beta_{y^-k}y^-)\right),$$

and the unconditional likelihood contribution is $\sum_k P_k \phi(\varepsilon|k)$. The parameters to be estimated are the β for each segment ($4 \times K$ parameters), the segment probability parameters α ($K-1$ parameters) and the variance of ε (one parameter).

We have estimated this model with up to four segments. The model with only one segment exhibits positive slopes for income changes in both the gains and the loss domain (the coefficient in the gains domain approaches zero when including the remaining control variables). The model with four segments had the lowest AIC (and BIC) value. We therefore only present the results of the model with four segments in table 6. The sample is dominated by observations belonging to the first segment (87.4%) who experience a positive effect of income changes in the gains domain, but do not exhibit any significant effect of income changes in the loss domain.

An individual in segment 2 experiences strong effects of both income levels and income changes. However, the negative slope of income gains is in contrast to PT. For a segment 3 individual there is a strong effect of gains on life satisfaction, but no significant loss effect. Segment 4 resembles segment 2, but has a very low probability mass. The results suggest a high degree of segmentation, with very different sensitivities to income gains and income losses across segments. Still, within segments, we also observe strong degrees of asymmetry between gain and loss sensitivity, though little support for a shape of the value function as stipulated by PT. However, we caution the reader that the finite mixture model is very parsimonious and treats all observations as independent, even within individuals.

8 Conclusion

We have analysed which characteristics of expected utility theory and prospect theory can be found in life satisfaction in an annual household panel. We found that LS exhibits features of both theories, in particular: LS resembles utility in EUT in that it increases in levels of income, health, and employment status, but it also shows features of utility in PT as it also increases in positive *changes* of those variables (except for health). Furthermore, LS exhibits loss aversion in income and in employment, but not in health. The main caveat here is our choice of reference point: the value of the variable of interest for *the same person*, at the time of the *previous interview*. It can well be that different reference points will result in a rejection of EUT in favour of PT or vice versa. However, this is also true for the other models of the reference point, in the literature or analysed herein. Finding the correct model, or the best model for a specific type of data available, is an important goal for future work.

Our finding about recalled and expected LS can also be analysed in a similar way. For example, a less distant past might be remembered more accurately, and therefore individuals might be more responsive to changes over a short time interval than to changes over the course of a year. Similarly, extrapolation of current to future LS might be based on an inability to foresee future events or to misjudge the probabilities of such events, or it might be a – conscious or unconscious – choice in order to enjoy anticipatory utility. In general, finding support for diminishing marginal LS and loss aversion is reassuring, given the popularity that PT has enjoyed.

We have also tested for the use of the anchoring and adjustment heuristic, by looking at the relative importance of current and past (future) LS in predicting recalled (expected) LS. We found that current LS is a much better predictor of recalled as well as expected LS than lagged or leading LS, and that the reported expected and recalled LS values are adjusted towards the true values, relative from the anchor. These observations provide compelling evidence for the use of the heuristic. We can conclude that questions about recalled and expected LS are very difficult to answer meaningfully for the average respondent, so care has to be taken in relying on them for policy decisions such as in the pricing of intangible and public goods.

References

Abdellaoui, Mohammed, Han Bleichrodt, and Corina Paraschiv (2007), "Loss aversion under prospect theory: A parameter-free measurement." *Management Science*, 53, 1659–1674.

Baetschmann, Gregori, Kevin E Staub, and Rainer Winkelmann (2015), "Consistent estimation of the fixed effects ordered logit model." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178, 685–703.

Bell, David R and James M Lattin (2000), "Looking for loss aversion in scanner panel data: The confounding effect of price response heterogeneity." *Marketing Science*, 19, 185–200.

Bénabou, Roland (2012), "Groupthink: Collective delusions in organizations and markets." *The Review of Economic Studies*, 80, 429–462.

Benjamin, Daniel J, Ori Heffetz, Miles S Kimball, and Alex Rees-Jones (2014), "Can marginal rates of substitution be inferred from happiness data? evidence from residency choices." *American Economic Review*, 104, 3498–3528.

Binder, Martin and Alex Coad (2015), "Heterogeneity in the relationship between unemployment and subjective wellbeing: A quantile approach." *Economica*, 82, 865–891.

Boyce, Christopher J, Alex M Wood, James Banks, Andrew E Clark, and Gordon DA Brown (2013), "Money, well-being, and loss aversion does an income loss have a greater effect on well-being than an equivalent income gain?" *Psychological Science*, 2557–2562.

Brunnermeier, Markus K and Jonathan A Parker (2005), "Optimal expectations." *American Economic Review*, 95, 1092–1118.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez (2012), "Inequality at work: The effect of peer salaries on job satisfaction." *American Economic Review*, 102, 2981–3003.

Chamberlain, Gary (1980), "Analysis of covariance with qualitative data." *The Review of Economic Studies*, 47, 225–238.

Cheung, Felix and Richard E Lucas (2016), "Income inequality is associated with stronger social comparison effects: The effect of relative income on life satisfaction." *Journal of Personality and Social Psychology*, 110, 332–341.

Clark, Andrew E (2018), "Four decades of the economics of happiness: Where next?" *Review of Income and Wealth*, 64, 245–269.

Clark, Andrew E, Conchita D'Ambrosio, and Simone Ghislandi (2016), "Adaptation to poverty in long-run panel data." *Review of Economics and Statistics*, 98, 591–600.

Clark, Andrew E, Ed Diener, Yannis Georgellis, and Richard E Lucas (2008a), "Lags and leads in life satisfaction: A test of the baseline hypothesis." *The Economic Journal*, 118, F222–F243.

Clark, Andrew E, Paul Frijters, and Michael A Shields (2008b), "Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles." *Journal of Economic Literature*, 46, 95–144.

Clark, Andrew E and Andrew J Oswald (1996), "Satisfaction and comparison income." *Journal of Public Economics*, 61, 359–381.

De Neve, Jan-Emmanuel and Andrew J 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*, 109, 19953–19958.

De Neve, Jan-Emmanuel, George Ward, Femke De Keulenaer, Bert Van Landeghem, Georgios Kavetsos, and Michael I Norton (2018), "The asymmetric experience of positive and negative economic growth: Global evidence using subjective well-being data." *Review of Economics and Statistics*, 100, 362–375.

DellaVigna, Stefano (2009), "Psychology and economics: Evidence from the field." *Journal of Economic Literature*, 47, 315–372.

Di Tella, Rafael, John Haisken-De New, and Robert MacCulloch (2010), "Happiness adaptation to income and to status in an individual panel." *Journal of Economic Behavior & Organization*, 76, 834–852.

Dolan, Paul and Daniel Kahneman (2008), "Interpretations of utility and their implications for the valuation of health." *The Economic Journal*, 118, 215–234.

Dolan, Paul, Tessa Peasgood, and Mathew White (2008), "Do we really know what makes us happy? a review of the economic literature on the factors associated with subjective well-being." *Journal of Economic Psychology*, 29, 94–122.

Easterlin, Richard A (1974), "Does economic growth improve the human lot? some empirical evidence." In *Nations and Households in Economic Growth*, 89–125, Elsevier.

Epley, Nicholas and Thomas Gilovich (2006), "The anchoring-and-adjustment heuristic: Why the adjustments are insufficient." *Psychological Science*, 17, 311–318.

Fang, Zheng and Yoko Niimi (2017), "Does everyone exhibit loss aversion? evidence from a panel quantile regression analysis of subjective well-being in japan." *Journal of the Japanese and International Economies*, 46, 79–90.

Ferrer-i Carbonell, Ada (2005), "Income and well-being: an empirical analysis of the comparison income effect." *Journal of Public Economics*, 89, 997–1019.

Ferrer-i Carbonell, Ada and Paul Frijters (2004), "How important is methodology for the estimates of the determinants of happiness?" *The Economic Journal*, 114, 641–659.

Frijters, Paul, John P Haisken-DeNew, and Michael A Shields (2004), "Money does matter! evidence from increasing real income and life satisfaction in east germany following reunification." *American Economic Review*, 94, 730–740.

Gilbert, Daniel (2006), *Stumbling on happiness*. Harper Perennial.

Gollier, Christian (2011), "Optimal illusions and the simplification of beliefs." *Working Paper, University of Toulouse*.

Gottlieb, Daniel (2014), "Imperfect memory and choice under risk." *Games and Economic Behavior*, 85, 127–158.

Harrison, Glenn W and John A List (2004), "Field experiments." *Journal of Economic literature*, 42, 1009–1055.

Jorgensen, Bradley S, Robert D Jamieson, and John F Martin (2010), "Income, sense of community and subjective well-being: Combining economic and psychological variables." *Journal of Economic Psychology*, 31, 612–623.

Kahneman, Daniel and Amos Tversky (1979), "Prospect theory: An analysis of decision under risk." *Econometrica*, 47, 263–291.

Kahneman, Daniel, Peter P Wakker, and Rakesh Sarin (1997), "Back to bentham? explorations of experienced utility." *The Quarterly Journal of Economics*, 112, 375–406.

Kőszegi, Botond and Matthew Rabin (2006), "A model of reference-dependent preferences." *The Quarterly Journal of Economics*, 121, 1133–1165.

Layard, Richard, Guy Mayraz, and Stephen Nickell (2008), "The marginal utility of income." *Journal of Public Economics*, 92, 1846–1857.

Levinson, Arik (2012), "Valuing public goods using happiness data: The case of air quality." *Journal of Public Economics*, 96, 869–880.

List, John A (2004), "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica*, 72, 615–625.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin (2003), "Projection bias in predicting future utility." *The Quarterly Journal of Economics*, 118, 1209–1248.

Loewenstein, George and Peter A Ubel (2008), "Hedonic adaptation and the role of decision and experience utility in public policy." *Journal of Public Economics*, 92, 1795–1810.

Lucas, Richard E, Andrew E Clark, Yannis Georgellis, and Ed Diener (2004), "Unemployment alters the set point for life satisfaction." *Psychological science*, 15, 8–13.

Luttmer, Erzo FP (2005), "Neighbors as negatives: Relative earnings and well-being." *The Quarterly Journal of Economics*, 120, 963–1002.

McBride, Michael (2001), "Relative-income effects on subjective well-being in the cross-section." *Journal of Economic Behavior & Organization*, 45, 251–278.

Novemsky, Nathan and Daniel Kahneman (2005), "The boundaries of loss aversion." *Journal of Marketing research*, 42, 119–128.

O'Brien, Ed, Phoebe C Ellsworth, and Norbert Schwarz (2012), "Today's misery and yesterday's happiness:

Differential effects of current life-events on perceptions of past wellbeing." *Journal of Experimental Social Psychology*, 48, 968–972.

Odermatt, Reto and Alois Stutzer (2018), "(Mis-) predicted subjective well-being following life events." *Journal of the European Economic Association*, 17, 245–283.

Proto, Eugenio and Aldo Rustichini (2013), "A reassessment of the relationship between gdp and life satisfaction." *PlOS One*, 8, e79358.

Proto, Eugenio and Aldo Rustichini (2015), "Life satisfaction, income and personality." *Journal of Economic Psychology*, 48, 17–32.

Sacks, Daniel W, Betsey Stevenson, and Justin Wolfers (2012), "The new stylized facts about income and subjective well-being." *Emotion*, 12, 1181.

Schwarz, Norbert and Fritz Strack (1999), "Reports of subjective well-being: Judgmental processes and their methodological implications." *Well-being: The foundations of hedonic psychology*, 7, 61–84.

Stevenson, Betsey and Justin Wolfers (2013), "Subjective well-being and income: Is there any evidence of satiation?" *American Economic Review*, 103, 598–604.

Tversky, Amos and Daniel Kahneman (1974), "Judgment under uncertainty: Heuristics and biases." *Science*, 185, 1124–1131.

Tversky, Amos and Daniel Kahneman (1991), "Loss aversion in riskless choice: A reference-dependent model." *The Quarterly Journal of Economics*, 106, 1039–1061, URL +http://dx.doi.org/10.2307/2937956.

Vendrik, Maarten CM and Geert B Woltjer (2007), "Happiness and loss aversion: Is utility concave or convex in relative income?" *Journal of Public Economics*, 91, 1423–1448.

Wilson, Anne E and Michael Ross (2000), "The frequency of temporal-self and social comparisons in people's personal appraisals." *Journal of personality and social psychology*, 78, 928–942.

Wilson, Timothy D, Jay Meyers, and Daniel T Gilbert (2003), "How happy was i, anyway? a retrospective impact bias." *Social Cognition*, 21, 421–446.

Appendix: Further results

Table 7: Determinants of life satisfaction

	(1) OLS	(2) Fixed effects
Household income (in 1,000 Euros) (β_1)	0.500***	0.291***
	(0.023)	(0.015)
(Household income)^2 (β_2)	-0.025***	-0.011***
	(0.005)	(0.002)
Income gain (γ_1)	-0.084	0.023
	(0.079)	(0.070)
(Income gain)^2 (γ_2)	-0.177	-0.203
	(0.197)	(0.173)
Income loss (δ_1)	0.251***	0.240***
	(0.087)	(0.076)
(Income loss)^2 (δ_2)	0.581***	0.512***
	(0.222)	(0.193)
Health:very good (α_5)	4.207***	2.783***
	(0.033)	(0.046)
Health:good (α_4)	3.485***	2.413***
	(0.031)	(0.043)
Health:satisfactory (α_3)	2.632***	1.938***
	(0.031)	(0.042)
Health:not so good (α_2)	1.718***	1.276***
	(0.032)	(0.041)
Employed (ρ_E)	-0.089***	0.044***
	(0.009)	(0.015)
Unemployed (ρ_U)	-0.810***	-0.486***
	(0.021)	(0.026)
Male	-0.091***	
	(0.005)	
Partner	0.317***	0.249***
	(0.007)	(0.017)
Children	0.230***	0.078***
	(0.008)	(0.012)
Years of education	-0.015***	-0.018***
	(0.001)	(0.005)
Observations	320,053	320,053
Number of persons		51,079
R-squared	0.272	0.101

Robust standard errors in parentheses. Omitted categories are Health:bad, and not in labour force. Regressions include a full set of year fixed effects, dummies for five age categories, and transitions between all health and labor force states. The R-squared for the fixed effects model is the squared correlation between the de-meaned life satisfaction and predicted de-meaned life satisfaction. Stata reports this measure as R-squared within. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8: Summary of hypotheses to be tested - Fixed effects ordered logit model

Code	Alternative hypothesis H_a	H_a supports:	p-value
Panel A: Level effects			
Y1	LS is increasing in income levels.	EUT	0.000
H1	LS is increasing in good health.	EUT	0.000
L1	LS is increasing in employment.	EUT	0.000
Panel B: Reference point			
RP1	LS is increasing in income changes in the gains domain.	PT	0.359
	LS is increasing in health gains in the gains domain.	PT	$>$0.802
	LS is increasing when one becomes employed.	PT	0.983
RP2	LS is increasing in income changes in the loss domain.	PT	0.005
	LS is increasing in health gains in the loss domain.	PT	$>$0.999
	LS is decreasing when one becomes unemployed.	PT	0.009
Panel C: Diminishing marginal utility			
DMU1	LS is concave in income levels.	EUT	0.001
DMU2	LS is concave in income changes in the domain of gains.	PT	0.119
DMU3	LS is convex in income changes in the domain of losses.	PT	0.007
Panel D: Loss aversion			
LA	LS exhibits loss aversion in income.	PT	0.090
	LS exhibits loss aversion in health.	PT	n.a.
	LS exhibits loss aversion in (un)employment.	PT	0.008

*In the case of RP for health, we tested each possible transition between health states separately. The reported *p*–value is the lowest among the 10 tests.