

Köke, Sonja; Perino, Grischa

Conference Paper

The Ranking Measure of Life Satisfaction: A Constructive Critique

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Life satisfaction and mental health, No. F19-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Köke, Sonja; Perino, Grischa (2015) : The Ranking Measure of Life Satisfaction: A Constructive Critique, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Life satisfaction and mental health, No. F19-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/113169>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Ranking Measure of Life Satisfaction: A Constructive Critique

March 1, 2015

Abstract

Life satisfaction (LS) is used as a welfare-measure in many different ways to inform about what makes people better off and even to quantify tradeoffs between different things like putting a monetary value on environmental goods. All these welfare analyzes rely on the ability of the measurement method to reliably identify preference (or satisfaction) rankings. We demonstrate a potential and systematic identification failure of the currently used measurement method (*level measure*), which asks people to state their LS level on a bounded and discrete scale. We then propose a new measurement method (the *ranking measure*), which directly asks the individual to give a preference ranking of the current situation against a past situation. This ranking measure will allow us to test the ability of the level measure to capture preference rankings of two consecutively experienced situations. First results from a similar measure in the German Socio-Economic Panel from the years 1984-87 suggest that there is a systematic bias between the two measures. Our ranking measure will be included in the GESIS Panel in February this year. It will include not only general LS but also satisfaction with specific areas of life. With this paper we contribute to a better understanding of life satisfaction measures and the nature of their measurement biases by formally connecting the measurement method with a theoretical welfare concept and by then testing the derived hypotheses empirically with survey data.

Keywords: Life Satisfaction, Valuation, Identification Problem, Measurement, Subjective wellbeing, Ranking-Measure

JEL codes: I30, I31

1 Introduction

Life satisfaction is now a well established measurement of well-being in social sciences¹ and is increasingly used to inform governments about well-being of their citizens.² The data have been analyzed at the aggregate and individual level, to determine the correlates³, and it has also been used to infer (marginal) willingness to pay for non-market goods in the life satisfaction approach (e.g. Frey et al. (2004)). This approach has especially been used for environmental goods like air pollution, water pollution, noise nuisance, climate parameters and others (see Welsch & Kühling (2009) and Frey et al. (2004) for reviews, a more recent publication is Kopmann & Rehdanz (2013)).

Given that those studies have or might have strong policy implications, we want to take a step back and analyze what information the currently used measurement method (level measure) can provide. The level measure asks individuals to state their level of LS on a bounded and discrete scale, consisting of 3 to 11 categories in different surveys. Most policy

¹See for example Frey & Stutzer (2002); Di Tella et al. (2003); Di Tella & MacCulloch (2006); Luttmer (2005); Frey et al. (2004); Ferrer-i Carbonell (2005); Dolan et al. (2011); Easterlin et al. (2010); Easterlin (1974); Oswald (1997); Stevenson & Wolfers (2008); Dolan et al. (2008).

²Great Britain, Germany and France have funded research on alternative welfare measures, including subjective ones, or stated interest in it. In France, President Sarkozy appointed the "Stiglitz-Sen-Fitoussi Commission on Economic Performance and Societal Progress". Recommendation 10 of their report Stiglitz et al. (2009) states "Measures of both objective and subjective wellbeing provide key information about people's quality of life. Statistical offices should incorporate questions to capture people's life evaluations, hedonic experiences and priorities in their own survey". In Germany, the "Enquete-Kommission Wachstum, Wohlstand, Lebensqualität" of the German Parliament suggested a dashboard of welfare measures on which to found policy decision, however, in the end they excluded subjective measures Deutscher Bundestag (2013). In the UK, the ONS (Office of National Statistics) has started the "Measuring National wellbeing Programme" which also includes collecting subjective wellbeing data on a large scale to explicitly inform policy makers (see also Dolan et al., 2011) and there is a dedicated wellbeing team in the Cabinet Office. The European Commission has a Beyond GDP Initiative to provide an overview of a set of clear and appealing indicators of wellbeing, including subjective wellbeing indicators. And also a growing number of government independent measures can be observed, like the Happy Planet Index, the World Happiness Index and the OECD Better Life Index.

³First studies looked at averages of life satisfaction over time and found that in industrialized countries average life satisfaction does not increase with GDP ((Easterlin, 1974; Easterlin et al., 2010)). There is evidence that in the short run, GDP has a positive impact on life satisfaction (Di Tella et al., 2003). Later on, data were also analyzed at the individual level to isolate the impacts of personal income (Luttmer, 2005; Ferrer-i Carbonell, 2005) and other socio-economic variables on individual life satisfaction, which were often found to be significant. For different events, adaptation over time has been observed, i.e. in the short run there is a reaction of LS to a positive or negative event but over time, people tend to go back to their initial level of LS (see e.g. Diener et al. (2006) or Tella et al. (2010)).

implications that have been derived from LS studies rely on the assumption that the level measure of life satisfaction is able to inform about people' preference rankings (satisfaction rankings) over the situations they have experienced⁴ ⁵. We show that the level measure will in some cases make it impossible for individuals to give answers consistent with their preference rankings. We thereby show that for obtaining accurate information about the size of the tradeoffs people are willing to make, the tests for validity and reliability (see e.g. Diener et al. (2013)) that have been conducted are not sufficient. We therefore suggest a different measurement method (ranking measure), which directly asks individuals to rank their current situation against the situation of the previous period. It asks individuals whether they think that their life has improved (much or a little) or deteriorated (much or a little) or stayed constant over the past year.

Given that life satisfaction is an ordinal concept (it does not have any absolute unit of measurement but it is possible to say that someone is more or less satisfied)⁶, respondents have to "invent" a transformation function to map their ranking into a scale from 0 to 10. They might forget the function they used the previous year and "invent" a new one, such that the numbers are not comparable over time. If this re-invention is systematically affected by socio-economic characteristics, this will distort the estimation results and hence the policy implications drawn. This might for instance be the case if a participant experiences (or expects) a long period of continuously increasing LS. Repeatedly giving higher values on the bounded scale is not possible - at some point the upper limit is reached. In such a situation, respondents are forced to change their transformation function over time and the levels do no longer capture the underlying preference rankings. The ranking measure, instead, has the advantage that it does not need a transformation function and therefore is able to identify over time more accurately.

⁴This is especially the case when the question is aiming at a cognitive evaluation of the situation, which we want to aim at throughout the paper. It is less true for questions that ask for "happiness" rather than "life satisfaction" since happiness as an emotional state might refer to some objective level of brain activity.

⁵We will use the term preference ranking and satisfaction ranking interchangeably, both referring to some ranking of situations the individual makes when cognitively comparing two situations. So we do not restrict the term preferences to the ranking that can be inferred with the concept of revealed preferences in the standard economic theory, which is to infer preferences from observing the choice an individual makes given a certain choice set.

⁶Again: this argument does not hold in the same way for emotional states.

Of course, also this ranking measure has some shortcomings. It assumes that situations at different points in time can be ranked. It also assumes that people remember the situations they were in the previous year. However, these requirements seem less demanding than those needed to make the current measurement method consistent. Using the new ranking measure alongside with the level measure can help to better understand determinants of subjective life satisfaction and to improve the information provided to policy makers on what makes people better off. It will shed new light on the debate about the relative importance of absolute versus relative income, on the Easterlin Paradox and on the role of many other socio-economic variables. Moreover, this paper contributes to closing the gap between the measurement of subjective well-being and social choice theory.

First estimates from a similar question in the SOEP-Data from 1984-87 suggest that there is a systematic difference between the results from the level and the ranking measure. Our ranking measure will be implemented in the GESIS survey in February this year and allow for more data analysis.

In this paper we first introduce definitions and formal notation of what we want to measure (section 2). We then analyze the level measure to show why the identification problem arises (section 3). We then introduce the ranking measure in section 4 before discussing first results that we obtained from similar question in the SOEP in 1984-87 in section 5. Section 6 concludes.

2 Life Satisfaction

In this section we provide a formal definition of life satisfaction, the construct that is to be measured. The aim of any life satisfaction analysis is to evaluate concrete situations using statements by individuals about how they value the situation they are in. Let $x_{i,t}$ be the situation of individual $i \in \{1, 2, \dots, N\}$ in period $t \in \{1, 2, \dots, T\}$ ⁷. $x_{i,t}$ can be any element of $S = \{A, B, C, \dots\}$. Each element of S is a vector of concrete values of the variables describing the situation of some individual i in some point in time t . It contains concrete

⁷Time is of course continuous, however data for a panel is usually collected in some regular intervals, which is why it makes sense to have discrete time.

values of e.g. income, housing, education, profession, marital status, number of children, but also relative terms like the own rank in the income distribution of a peer group, or macro variables like inflation, interest rate, and GDP. S therefore is the set of all possible situations any individual can be in. For simplicity we will use $x_{i,t} = A$ equivalently to $A_{i,t}$.

Let the valuation of situation x by individual i at time t be $v_{i,t}(x)$, which can be thought of as a utility function. It represents the preference ranking of individual i in period t in the sense that it assigns higher values to more preferred situations, that give more satisfaction, than to less preferred situations, which give less satisfaction. The argument of the valuation function is x , where we dropped the subscripts to make clear that the valuation is about situations, independent of the time of their realization. Note that in addition to most utility functions, we allow the preferences to change over time. But in line with traditional utility functions, preferences do not refer to the order in which the situations occur. The only other difference is that the term “utility” has traditionally been used in the context of *revealed preferences* when eliciting preference relations from observed *choices*, while we here are interested in *life satisfaction* which elicits the underlying preference relations from *statements*. While one may argue that life satisfaction data are not meant to represent preference relations, they have nonetheless been evaluated as if they were. Higher values of life satisfaction are usually interpreted as being more desirable than lower values.

The valuation function therefore has the following definition:

$$A \succ_{i,t} B \iff v_{i,t}(A) > v_{i,t}(B), \tag{1}$$

where $\succ_{i,t}$ represents individual i 's preference relation (in terms of life satisfaction) in period t . It assigns higher values to preferred situations. The valuation function represents the preferences of individuals over different situations⁸.

Given the structure of observable situations and the concept of life satisfaction, it is only possible to get individual rankings over situations in different points in time. Preferences are

⁸An interesting alternative would be to allow for preferences over sequences of events, representing different "lives", but this is not the content of this paper. Also the usual analysis of LS data has been to interpret LS as referring to situations in a certain point in time, not to sequences of events.

not defined between individuals, such that comparing $x_{i,t}$ with $x_{j,t}$ for $i \neq j$ is not possible. Also rankings of two situations for one individual at the same point in time would require asking hypothetical questions, which is not the idea of life satisfaction, which is interested in actually experienced situations. Therefore, the only ranking that can be elicited with the life satisfaction approach is $x_{i,t}$ versus $x_{i,s}$ for $s \neq t$.

In the following we are interested in measurement methods of life satisfaction and their ability to identify individuals' valuation of situations. We denote by Z a specific measurement method for life satisfaction, consisting of a *survey question* and an *answer scale*.

3 The Level Measure

We now analyze the current measurement method for life satisfaction, the level measure (LM) to see how well it is able to identify preference relations. Level Measures typically ask questions like the following:

"All things considered, how satisfied are you with your life as a whole these days? Using this card on which 1 means you are "completely dissatisfied" and 10 means you are "completely satisfied" where would you put your satisfaction with your life as a whole?"

(World Value Survey 2005-2006 Wave, Core Version, V22)

While not all surveys use the exact same wording, they all have the following in common⁹:

- They ask participants to report the current level of their life satisfaction.
- The answer scale provides a fixed number of categories represented by labels and numbers (the number of categories ranging from 3 to 11).
- When analyzing the data, the categories are number coded and often used to compute means or apply estimators that require a cardinal structure.¹⁰

⁹This is a list of surveys which use a Level Measure of life satisfaction: World Value Survey, German SOEP, Eurobarometer, Latinobarometer, US General Social Survey, Happy Planet Index, World Happiness Index and the OECD Better Life Index.

¹⁰There are several papers using the appropriate statistical tools, e.g. Di Tella et al. (2003); Blanchflower & Oswald (2004); Luttmer (2005).

- The results are used to identify situations in which people are better off than in others.

Due to these similarities the LM scores are generally considered to be comparable across different surveys and the empirical results often are similar.¹¹

In the notation we introduced in the previous section, the measurement method is the level measure ($Z = LM$), the corresponding survey question has the features described above and the answer scale is $\{LM^{min}, \dots, LM^{max}\}$. We represent the answer individual i gives to a life satisfaction question at time t when using the measurement method LM by

$$LM_{i,t}(x_{i,t}) = f_i(v_{i,t}(x_{i,t})) \in \{LM^{min}, \dots, LM^{max}\}. \quad (2)$$

The individual transforms the preference ranking it has over all elements of S , represented by the valuation function $v_{i,t}(\cdot)$ into an answer on the answer scale by using his individual transformation function $f_i(\cdot)$. An observed change in $LM_{i,t}(x_{i,t})$ from one period to the other can therefore be driven by three different things. Either the situation to be evaluated ($x_{i,t}$), or the preferences ($v_{i,t}(\cdot)$), or the transformation function ($f_i(\cdot)$) or any combination of them might have changed. We will now argue that it is not possible to isolate the source of such a change with the LM. While a change in $LM_{i,t}$ due to a change in x or v are welfare relevant, a change in f should not be interpreted as welfare relevant and needs to be filtered out or at least it needs to be random when analyzing a large sample.

We now argue why and in which cases it seems likely that f changes systematically and therefore biases the information the LM can give on individuals' preferences.

Let us for now consider the case when the preferences of individuals, i.e. the v -functions, do not change over time, such that $v_{i,t}(\cdot) = v_{i,s}(\cdot)$ for all i , t and s . While the researcher can compare two consecutive values, i doesn't actually compare the two situations directly, when answering the survey question, but rather takes the following two separate steps: In $t - 1$, i places $x_{i,t-1}$ on a bounded scale (e.g. from 1 to 10) and in t , i places $x_{i,t}$ on the

¹¹Several studies find that the structure of the estimated equations that relate stated valuation from a LM to microdata are almost identical (Di Tella et al., 2003; Blanchflower & Oswald, 2004) and some have gone as far as pooling them (Easterlin et al., 2010).

same bounded scale. Only afterwards does the researcher compare the two statements to see which one is higher and to conclude which situation is preferred. From the data, we do not have any information about what people really chose as a transformation function when answering the question in a certain period. We cannot exclude that they adapt the transformation function f to the new situation. If this adaptation of the transformation function happens randomly, there is no problem for average interpretation of the data collected. However, we now show under which conditions a systematic adaptation of the transformation function seems likely.

One example where one could suspect such an identification error is the Easterlin Paradox. There, a typical observation is that in an earlier period $t - 1$ where GDP is lower (situation $x_{i,t-1} = B$) the average answer is $LM_{i,t-1}(x_{i,t-1}) = 7$ and that in a later period t , the average answer decreases $v_{i,t}(x_{i,t}) = 6$ while GDP has increased (situation $x_{i,t} = A$). However, it is not clear whether in a direct comparison people would nevertheless have a preference for the higher income: $v_{i,t}(x_{i,t}) - v_{i,t-1}(x_{i,t-1}) > 0$, such that $A \succ_{i,t} B$. Remember that we consider a case where preferences do not change over time, such that the valuation function is the same in t and $t - 1$. So even under this simplified preference structure, there is a potential identification problem with the LM.

A plausible reason to expect an adaptation of the valuation function is that the answer scale is bounded (and discrete) whereas the number of situations that a person can imagine is if not infinite then very large (S has a very large number of elements). This means that each individual has to come up with an interpretation of how to define the minimum, the maximum and each step of the scale, in order to translate his life satisfaction onto that scale. He has to anchor the ordinal concept of life satisfaction on the scale by arbitrarily defining what situation corresponds to, e.g., a 5. He then has to define when to go up one step of the scale, how much improvement is needed to answer with a 6. At the same time that he wants to signal improvements, he has to keep in mind that the scale is bounded, that if the answer is increasing too quickly, the upper limit will soon be reached. So there is a tradeoff between making changes visible and keeping room for further changes. This is no problem, if there is no trend in life satisfaction. However, if there is constant improvement or constant deterioration, this cannot adequately be captured by the scale. If

we then observe that average life satisfaction stayed constant and GDP has increased (at the aggregate level) or if we observe adaptation at the individual level, we cannot identify from the data if this means that LS stayed at or returned to its original level because of the preferences (signaling indifference) or because there was an adaptation in the answer scale. This shows that we need to separate out adaptation which is driven by preferences and adaptation that is simply driven by the properties of the answer scale.

What we can take from this example is that we cannot make sure that the transformation function $f_i(.)$ is the same in t and in $t - 1$. Instead, it is likely that the functional form depends on the situation, i.e. the situation is not only an argument of the function but also changes the functional form.¹²

For this argument we kept the valuation function $v_{i,t}(.)$ constant. Of course it is also interesting to know how preferences change over time. We will take this into account in the next section.

From this analysis it also becomes clear that the tests for validity and reliability that have been conducted so far are an indicator that something welfare-relevant is being measured by the LM, but that this is not enough to reliably capture preference rankings.

4 The Ranking Measure

After having described in the previous section how the level measure potentially biases the representation of the underlying preferences, we now suggest an alternative measurement method, the ranking measure (RM). To prevent an uncontrollable change in transformation function from one period to the other, we suggest to instead ask individuals directly to rank the current situation against the situation from the previous period. A question which

¹²Note: some may say that a feeling might be measured objectively by some brain activity and that with some objective unit of this brain activity, a cardinality of the corresponding feeling might be reached which makes the concept to be measured different from a purely ordinal preference relation. However, here the same change in the transformation function might take place in the brain for the exact same reason (see Rayo and Becker for an evolutionary, biological argument of adaptation in life satisfaction or happiness), keeping the preference relation constant within one period but adapting the interpretation of the limited scale to the new level.

captures this reads as follows:

“In your opinion, has your life improved or deteriorated over the last year? (Please evaluate any changes from today’s point of view.)”

And as the answer scale we propose

“considerably improved / somewhat improved / stayed the same / somewhat deteriorated / considerably deteriorated / don’t know”.

While we argued before that LS is an ordinal (and not cardinal) concept, we also believe that some qualitative statement about the magnitude of a LS change is possible. So the categories on the scale could be increased or reduced symmetrically around the "stayed the same" value.

The data that we can collect with this RM can be represented by

$$RM_{i,t}(x_{i,t}, x_{i,t-1}) = f_i(v_{i,t}(x_{i,t} - x_{i,t-1})),$$

denoting the fact that the individual evaluates the experienced change from $x_{i,t-1}$ to $x_{i,t}$. This term is positive if $x_{i,t} \succ_{i,t} x_{i,t-1}$ and negative if $x_{i,t-1} \succ_{i,t} x_{i,t}$, i.e. it captures today’s preferences. In contrast to the LM, the two situations are compared using the same transformation and valuation functions (by individual i at time t), which could not be guaranteed for the LM and which caused the potential identification failure. Also, the scale is no more bounded, since it is each year possible to observe improvements or to each year observe deteriorations.

Of course, this RM requires individuals to remember the past situation (without bias in the memory). However, the LM requires them to remember the transformation function used, in order to make the comparison possible. We believe that the former is less problematic. Another difference to the LM is that it is no longer possible to analyze and plot levels of LS. Rather the object of analysis will be changes in LS which can be attributed to changes in situations.

5 Data Evaluation

So far, our ranking measure has been accepted to be included in the GESIS Panel in February/March this year. So we don't have the data yet.

However, the German Socio-Economic Panel (SOEP) already included three questions that are similar enough to allow shedding some light on our research question. They indicate that the time series of current levels of life satisfaction does not reliably reflect how participants rank (in retrospect) situations they experienced at different points in time - and that the difference between the level measure and the ranking is correlated with key socio-economic characteristics suggesting that these variables also affect the transformation function used to translate experienced life satisfaction into a point on the answer scale and hence that the coefficients from regressions using the level measure as a dependent variable are biased.

The first variable from the SOEP that we want to analyze here is a retrospective LM implemented in the years 1984-87. In addition to the current LM of life satisfaction [plh0182], which we denote by $CurrentLM_{i,t} = LM_{i,t}(x_{i,t})$ there was a question about retrospective levels of life satisfaction of one year ago [plh0150], which we denote by $RetroLM_{i,t} = LM_{i,t}(x_{i,t-1})$. Both questions were answered on a LM scale from $LM^{min} = 0$ to $LM^{max} = 10$. While those two variables seem more likely to capture the ranking of two situations using the same valuation and transformation function¹³, they do not resolve the problem of the bounded scale, which means that for repeated improvements or repeated deteriorations, there still is the tradeoff between making changes visible and giving consistent answers. So we think that the data from our ranking measure would be more accurately measuring preference rankings if there is a trend (or even an expected trend) in LS. Having those similarities and differences to our measure in mind, we now conduct the analyzes we will conduct with our RM by computing what would be the equivalent to the RM as $ChangeRetro = CurrentLM_{i,t} - RetroLM_{i,t}$ and comparing this to the change in the LM over the corresponding period $ChangeCurrent = CurrentLM_{i,t} - CurrentLM_{i,t-1}$. We expect that the differences we find between the two measures (ChangeRetro and ChangeCurrent) are conservative estimates of the difference we will find between our RM and

¹³It is not exactly clear, if individuals really answer the retrospective question using $v_{i,t}$ instead of $v_{i,t-1}$, i.e. the old valuation function they used in t-1, such that even this property is not certain.

ChangeCurrent because by construction ChangeRetro is more similar to ChangeCurrent than RM is to ChangeCurrent, as explained above.

Comparing the change in the current measure (ChangeCurrent) with the retrospective change (ChangeRetro) shows that the former is negative (-0.144) and the latter positive (0.073) on average, both significantly different from zero ($p=0.0000$). Consequently, the changes are also statistically significantly different from each other ($p = 0.0000$). Hence, while the current measure of life satisfaction decreased, retrospectively participants felt that their lives have improved year-by-year. The correlation of the two variables is 0.35 which indicates that they do not strongly move together. To get a better understanding of the directions of movement, we look at the frequencies of sign combinations of the two measures. From the 30,172 observations which have non-missing values for both measures, 2,169 have opposite signs for the two measures, i.e. the two measures contradict each other in terms of ranking the present against the past. For 6,686 observations both variables are equal to zero and in 5,553 cases the sign is the same (either positive or negative), i.e. in those cases the two measures give the same ranking. In the remaining 15,765 cases one of the variables is zero and the other one either positive or negative, i.e. they neither really agree nor disagree. Those findings are summarized in Table 1. A Wilcoxon signed-rank test rejects the hypothesis that the two measures are the same ($p=0.0000$).

ChangeCurrent	ChangeRetro		
pos	neg	679	2,169
neg	pos	1,490	
0	0	6,686	6,686
pos	pos	2,627	5,553
neg	neg	2,926	
0	neg	745	15,765
0	pos	1,378	
neg	0	7,261	
pos	0	6,381	
Sum		30,173	

Table 1: Frequencies of sign combinations in ChangeCurrent and ChangeRetro for 30,173 observations.

Next, we are interested in what drives the observed differences. A first observation is that personal income and unemployment don't but gender and household income do explain some of the difference between the two measures, suggesting that women more strongly adapt the level scale and more household income also leads to more adaptation of the LM scale.

We now report the results of some ordered logistic regressions of ChangeCurrent and ChangeRetro on levels of unemployment, personal income, household income, being a homeowner, being chronically ill, having spent time at the hospital lately, and political interest, we find that only the level of being chronically ill has a significant negative impact on ChangeCurrent while for ChangeRetro also being a home owner has a negative and higher household income a positive effect.

To be done: a regression about adaptation, i.e. the question how people adapt to changes from two periods ago.

The second piece of evidence is from the question in 2009 that asked for the change of life satisfaction since the fall of the Berlin Wall [plh0169]. We compare the retrospective statements about whether a participant's life satisfaction has increased, decreased or stayed the same (a RM with three answer options, which we denote by $RM_{i,2009}(x_{i,2009}, x_{i,1989})$) with the change in the reported current level of life satisfaction from 1989 to 2009 ($LM_{i,2009} - LM_{i,1989}$). A Stuart-Maxwell test of marginal homogeneity rejects the null hypothesis that the changes go in the same direction ($p = 0.0000$) with the retrospective measure reporting a much lower share of participants that think themselves being worse off in 2009 than they were in 1989 (4,479 vs. 15,653). The full set of frequencies can be found in Table 2. While this question is similar to ours in that it asks for changes in life satisfaction (as a slight difference we want to ask for satisfaction with changes in life), it always has the same reference point of 1989, which is a very special year in Germany. It seems worthwhile to explore this with a more neutral and sliding point of reference.

To be done: include regression of RM and ChangeCurrent on changes in other variables. (not possible to control for regime shift with those data.)

The third indicator we have from the SOEP is based on the profile of life satisfaction over the

Change in the LM 1989 to 2009	RM comparing 2009 with 1989			
	increased	decreased	no change	total
increased	162	232	419	813
decreased	4,645	4,075	6,933	15,653
no change	138	172	397	707
total	4,945	4,479	7,749	17,173

Table 2: Frequencies of direction of change in the LM and RM from 1989 to 2009.

past ten years that has been elicited in 2000 and 2013 [plh0168]. Participants were asked to choose one of eight (nine in 2013) patterns which best represent the development of their LS over the past ten years. For 2000, the direction of the net changes implied by the graphical representation and the direction of change in the measure of current life satisfaction over the relevant period do not always coincide. Again the retrospective measure paints a more positive view of the change that has occurred over the previous ten years.

Taken together this clearly indicates that the level measure of current life satisfaction might not accurately reflect how participants retrospectively rank life situations experienced at different points in time and that these differences are not random.

6 Discussion and Conclusion

From the first suggestive empirical evidence we presented in the previous section, we now discuss the implications this would have on different results in the literature.

First, since overall LS seems to have increased for the RM but decreased for the LM, this suggests that the Easterlin Paradox is less severe than was assumed so far. It seems that the change in LS have been underestimated. So an increase in GDP and improvements in many other areas of life need not be canceled out by negative developments or by hedonic adaptation. So the changes that happened over time might have been valued more than suggested by the Easterlin Paradox (Easterlin, 1974).

Second, it seems like individual income has a significant effect on the RM but not on the LM. We cannot conclude from this alone, but would expect from theory, that an underestimation of a variable's effect increases with its (expected) growth rate. In the life satisfaction approach marginal willingness to pay for an environmental good is derived from the relative coefficients of income and the good from a regression like the following:

$$LM = \dots + \beta_E E + \beta_Y Y + \dots$$

The level measure of life satisfaction is regressed on an environmental variable E and on income Y, as well as on many other control variables, not specified here. The coefficients are then used to derive the marginal willingness to pay for the environmental good:

$$MWTP = dY/dE = \frac{dLM/dE}{dLM/dY} = \frac{\beta_E}{\beta_Y}$$

This states that the marginal willingness to pay (MWTP) for an environmental good is equal to the change in income (dY) per change in E (dE), keeping LM constant. This is equal to the relative coefficients from the regression of LM on E, Y and other control variables (see e.g. Kopmann & Rehdanz (2013)).

Assuming that Y has a larger growth rate than E, β_Y is more strongly underestimated than β_E , such that the MWTP would be smaller than currently estimated with the LM. This might not be the favorite result for environmental economists, since it would result in a lower likelihood of providing this good.

For further results we need to wait for our RM data.

References

- Blanchflower, D. G. & Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, 88, 1359–1386.

- Deutscher Bundestag (2013). *Gesamtbericht der Enquete-Kommission "Wachstum, Wohlstand Lebensqualität" - Wege zu nachhaltigem Wirtschaften und gesellschaftlichem Fortschritt in der Sozialen Marktwirtschaft*. Bundestrucksache 17/13300, Deutscher Bundestag, Berlin.
- Di Tella, R. & MacCulloch, R. (2006). Some use of happiness data in economics. *Journal of Economic Perspectives*, 20(1), 25–46.
- Di Tella, R., MacCulloch, R. J., & Oswald, A. J. (2003). The macroeconomics of happiness. *Review of Economics and Statistics*, 85(4), 809–827.
- Diener, E., Inglehart, R., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research*, 112(3), 497–527.
- Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: Revising the adaptation theory of well-being. *American Psychologist*, 61(4), 305–314.
- Dolan, P., Layard, R., & Metcalfe, R. (2011). *Measuring Subjective Well-being for Public Policy*. Report, Office for National Statistics, Richmond.
- Dolan, P., Peasgood, T., & White, M. (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(1), 94 – 122.
- Easterlin, R. (1974). Does economic growth improve the human lot? some empirical evidence. In P. David & M. Reder (Eds.), *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz* (pp. 98–125). New York and London: Academic Press.
- Easterlin, R. A., Angelescu McVey, L., Switek, M., Sawngfa, O., & Smith Zeig, J. (2010). The happiness-income paradox revisited. *Proceedings of the National Academy of Sciences*, 107(52), 22463–22468.
- Ferrer-i Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics*, 89, 997–1019.

- Frey, B. S., Luechinger, S., & Stutzer, A. (2004). Valuing public goods: The life satisfaction approach. *CESifo Working Paper*, (No. 1158).
- Frey, B. S. & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, 40(2), 402–435.
- Kopmann, A. & Rehdanz, K. (2013). A human well-being approach for assessing the value of natural land areas. *Ecological Economics*, 93(0), 20 – 33.
- Luttmer, E. F. P. (2005). Neighbors as negatives: relative earnings and well-being. *Quarterly Journal of Economics*, 120(3), 963–1002.
- Oswald, A. J. (1997). Happiness and economic performance. *The Economic Journal*, 107, 1815–1831.
- Stevenson, B. & Wolfers, J. (2008). Economic Growth and Subjective Well-Being: Re-assessing the Easterlin Paradox. *Brookings Papers on Economic Activity*, 39(1 (Spring)), 1–102.
- Stiglitz, J., Sen, A., & Fitoussi, J.-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Technical report, Paris.
- Tella, R. D., New, J. H.-D., & MacCulloch, R. (2010). Happiness adaptation to income and to status in an individual panel. *Journal of Economic Behavior and Organization*, 76(3), 834 – 852.
- Welsch, H. & Kühling, J. (2009). Using happiness data for environmental valuation: Issues and applications. *Journal of Economic Surveys*, 23(2), 385–406.